Functional Characterization of Human Terminal Deoxynucleotidyl Transferase Polymorphisms

Anna Troshchynsky

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

GRADUATE PROGRAM IN BIOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

SEPTEMBER 2013

Abstract

Terminal deoxynucleotidyl transferase (TdT) contributes to antigen receptor diversity of B and T lymphocytes of the adaptive immune system. TdT is a DNA polymerase catalyzing template independent addition of nucleotides during the process of V(D)J recombination. This thesis investigates whether the polymerase activities of naturally occurring single nucleotide polymorphic (SNP) forms of human TdT vary and the possible effects on the antigen receptor repertoire. Two SNPs (A445T/L397S) demonstrated significantly lower *in vitro* polymerization activities compared to the wild type enzyme, while one (R431C) was completely inactive. *In vivo* extrachromosomal recombination assay demonstrated over 3 fold difference in the proportion of N-containing joints and significantly lower numbers of N-nucleotides per joint for both L397S and R431C variants compared to WT hTdTS. Variability in enzymatic activities of the SNPs may affect the diversity of antigen receptors, thus potentially impact the downstream adaptive immune response in healthy and autoimmune-prone individuals possessing these SNPs.

Acknowledgements

I would like to extend my sincere gratitude to my supervisors, Dr. Gillian E. Wu and Dr. Yi Sheng, for their support and guidance throughout my research. Dr. Wu has given me a great opportunity to be part of her laboratory and also went above and beyond as a mentor to encourage and provide a great learning environment.

Also, a special thanks goes to past and present members of Wu, Sheng and Saridakis labs, especially Lina Chen and Bhargavi Duvvuri.

Lastly, I would like to thank the Canadian Institute for Health Research for providing me with generous funding throughout my work. It truly helped me focus on what matters most- science.

Table of Contents

Abstract....ii
Acknowledgments....iii
Table of Contents....iv
List of Tables....vii
List of Figures....viii

Chapter 1: Introduction....1

Section 1.1 Biochemical properties of TdT....2
1.1.1 TdT Expression and Regulation....2
1.1.2 Isoforms of TdT....3
1.1.3 TdT protein domain structure....4
1.1.4 Mechanism of polymerization....7

Section 1.2 Generation of antigen receptor diversity....7
1.2.1 Structure of B and T cell receptors....7
1.2.2 Mechanism of V(D)J recombination....9

Section 1.3: TdT and its role in autoimmunity....15

Section 1.4 Single nucleotide polymorphisms....16

Section 1.5 Thesis objectives....18

Chapter 2: Materials and Methods....19

Section 2.1 Selecting hTdT SNPs for study....19

Section 2.2 Plasmid hTdTS DNA construction....19
2.2.1 Bacterial expression vector construction via site-directed mutagenesis....19
2.2.2 Mammalian expression vector construction via restriction site infusion based cloning....23

Section 2.3 hTdT His-tag protein purification....28
2.3.1 Heat-shock transformation of competent BL21 cells....28
2.3.2 Induction of hTdT protein expression....28
2.3.3 hTdT protein purification via nickel affinity chromatography....29
2.3.4 hTdT protein dialysis into storage buffer....30
2.3.5 Analytical SDS-PAGE electrophoresis of hTdT protein purification....30

Section 2.4 *In vitro* hTdT functional activity assays....31
2.4.1 General polymerase activity assay setup....31

2.4.2 dNTP substrate preference polymerase activity assay....32
2.4.3 DNA substrate preference polymerase activity assay....33
2.4.4 DNA substrate sequence preference polymerase activity assay....33
2.4.5 Cofactor preference polymerase activity assay....34

Section 2.5 *In vivo* V(D)J recombination assay....35
2.5.1 Assay overview....35
2.5.2 Reverse transcriptase PCR....40
2.5.3 Western blotting procedures....43
2.5.4 Mammalian cell transfection for the V(D)J recombination assay....45
2.5.5 Harvesting plasmid DNA via alkaline lysis....46
2.5.6 DpnI digest of harvested plasmid DNA....47
2.5.7 Transformation of harvested plasmid DNA via electroporation....47
2.5.8 Colony PCR....48
2.5.9 Recombination data analysis....49

Chapter 3: Results....51

Section 3.1 hTdT SNPs....51

Section 3.2 Plasmid hTdTS DNA construction....59

Section 3.3 hTdT His-tag protein purification....65

Section 3.4 *In vitro* hTdT functional activity assays....67
3.4.1 Activity of commercial recombinant TdT versus purified wild type hTdTS....67
3.4.2 Polymerase time course activity assay....69
3.4.3 dNTP substrate preference polymerase activity assay....73
3.4.4 DNA substrate preference polymerase activity assays....76
3.4.5 DNA substrate sequence preference polymerase assays....80
3.4.6 Cofactor preference polymerase activity assays....82

Section 3.3 *In vivo* V(D)J recombination assays....84
3.3.1 HEK293T cell line testing....84
3.3.2 Recombination frequencies of the substrate pGG51 plasmid....88
3.3.3 Analysis of recombined substrate pGG51 sequences....90

Chapter 4: Discussion....100

4.1 Selection of hTdTS SNPs....100

4.2 hTdTS SNPs demonstrate varied polymerization activities *in vitro*....106

4.3 hTdTS activity differs according to dCTPs and dGTPs usage.....108

4.4 Usage of double stranded 3' overhang DNA substrate primer is preferred by hTdTS.....109

4.5 Cobalt versus magnesium cofactor use.....111

4.6 Examining R-values in V(D)J recombination assay.....112

4.7 Analysis of recombined joints.....113

4.8 Implications of hTdTS SNPs on adaptive immune diversity.....117

Conclusion.....119
References.....120
Appendices.....128
Appendix A: Complete list of hTdT variant 1 SNPs.....128
Appendix B: Coomassie stained 12% SDS-PAGE gels containing His-tag purification samples of hTdT protein variants.....131
Appendix C: Raw data of *in vivo* V(D)J extrachromosomal recombination assay.....133
Appendix D: Analysis of all obtained recombined substrate pGG51 plasmid joint sequences from *in vivo* V(D)J recombination assay.....140

List of Tables

Table number	Title	Page
Table 2.1	Site-directed mutagenesis primers used to generate point mutations in wild-type hTdTS cDNA, corresponding to selected hTdTS SNPs	21
Table 2.2	Site-directed mutagenesis PCR program parameters	22
Table 2.3	PCR infusion primers used to amplify hTdTS cDNA	25
Table 2.4	PCR program parameters for amplification of hTdTS cDNA based on restriction site based cloning	26
Table 2.5	DNA sequences of DNA substrates used for *in vitro* DNA substrate preference polymerase assay	33
Table 2.6	PCR program parameters for amplification of hTdTS cDNA following cDNA synthesis via Reverse transcriptase	43
Table 2.7	PCR program parameters for colony PCR of recombination pGG51 substrate plasmid	49
Table 3.1	Human TdT variant 1 SNP candidates for functional activity assays	56
Table 3.2	Recombined substrate pGG51 plasmid nucleotide deletions	91
Table 3.3	Recombined substrate pGG51 plasmid P-nucleotide additions	92
Table 3.4	Recombined substrate pGG51 plasmid N-nucleotide additions, their frequency and AT versus GC base content	96

List of Figures

Figure number	Title	Page
Figure 1.1	Domain structure of TdT protein	5
Figure 1.2	Crystal structure of murine TdT (PDB code 1KEJ)	6
Figure 1.3	V(D)J recombination overview of immunoglobulin heavy chain locus	13
Figure 1.4	Schematic overview of a single V-J recombination event and associated protein factors, resulting in the generation of coding joint diversity	14
Figure 2.1	p15TV-L vector map (GenBank accession EF456736)	20
Figure 2.2	Features of pcDNA^{TM}4/myc-His A, B, C 5.1 kb vectors	24
Figure 2.3	Sequences of Cy3-labelled oligo 1 and 2 used for *in vitro* DNA substrate sequence preference polymerase activity assay	34
Figure 2.4	Overview of the *in vivo* extrachromosomal recombination assay	37
Figure 2.5	General design of the recombination substrate plasmid for *in vivo* assay	38
Figure 2.6	Substrate recombination plasmid pGG51 sequence features	39
Figure 3.1	Multi-sequence alignment of TdT amino acid sequences from 18 different species available at NCBI database	53
Figure 3.2	Crystal structure of murine TdT (PDB ID: 1KEJ) and the relative positions of hTdTS SNPs	57
Figure 3.3	Homo sapiens deoxynucleotidyl transferase, terminal (DNTT), transcript variant 1, mRNA	60
Figure 3.4	Protein multi-sequence alignment of DNA nucleotidylexotransferase isoform 1 [Homo sapiens] (NCBI Reference Sequence: NP_004079.3) with six chosen hTdTS SNPs	62

Figure 3.5 Restriction digest of hTdTS wild type, A445T, L397S, R431C cDNA p15TV-L vectors and pcDNA^{TM}4/myc-His A vector.....64

Figure 3.6 DNA multi-sequence alignment of cloned wild-type, L397S, R431C and A445T hTdTS cDNA in the pcDNA^{TM}4/myc-His A expression vector.....65

Figure 3.7 Coomassie stained 12% SDS-PAGE gel containing His-tag purification samples of wild type hTdT.....66

Figure 3.8 His-tag purified hTdT protein variants resolved on 10% denaturing SDS-PAGE gel stained with Coomassie blue dye.....66

Figure 3.9 *In vitro* polymerase activity of commercial recombinant TdT versus purified wild type hTdTS.....68

Figure 3.10 *In vitro* polymerase time course activity of purified hTdTS SNP variants.....70

Figure 3.11 Change in maximum size of hTdTS oligo product with respect to time of *in vitro* reaction.....72

Figure 3.12 *In vitro* dNTPs preference polymerase activity of purified hTdTS SNP variants.....74

Figure 3.13 *In vitro* polymerase activity of purified wild type hTdTS supplemented with dGTPs.....76

Figure 3.14 *In vitro* DNA substrate preference polymerase activity of purified hTdTS SNP variants.....77

Figure 3.15 *In vitro* DNA substrate sequence preference polymerase activity assay.....80

Figure 3.16 *In vitro* cofactor polymerase activity assay.....83

Figure 3.17 Reverse transcriptase PCR products using hTdTS gene-specific primers, separated on 0.7% agarose gel.....86

Figure 3.18 Western blot image of total HEK293T cell lysates probed with rabbit polyclonal TdT specific antibody and mouse monoclonal IgG GAPDH specific antibody.....87

Figure 3.19 Recombination frequencies of *in vivo* V(D)J recombination assay.....89

Figure 3.20 P-values obtained from standard T-test or chi-square test.....97

Figure 3.21 Distribution of N-nucleotide additions per single recombined joint.....98

Figure 3.22 Mean length of arbitrary complementary-determining region.....99

Figure 4.1 Crystal structure of murine TdT and location of A445 and T450 residues within incoming dNTP-binding site.....102

Figure 4.2 Crystal structure of murine TdT and salt bridge interactions of R431 with catalytic aspartate residues.....103

Figure 4.3 Crystal structure of murine TdT in complex with ddATP is superimposed onto the murine TdT structure in complex with primer single stranded DNA.....104

Chapter 1: Introduction

The human adaptive immune system is equipped with mechanisms to recognize and defend against numerous pathogens through the generation of B and T lymphocytes. A pathogen-specific immune response is mounted through antigen recognition via T and B lymphocyte receptors. The diversity of B and T antigen receptors is encoded by genes that are modified by recombination during the development of B and T cells. The recombination results in well over a million different receptors with different specificities and affinities for essentially any antigen. The receptors contain a variable domain region that constitutes the antigen-binding site through gene rearrangements and recombination of variable (V), diversity (D) and joining (J) gene segments. This process is known as V(D)J recombination and contributes to diverse repertoire of lymphocyte antigen receptors that are crucial for proper adaptive immune response.

My thesis work focuses on one of the enzymes critical for V(D)J recombination and T cell and B cell receptor diversity, terminal deoxynucleotidyl transferase (hTdT). I undertook a functional characterization of a number of human naturally occurring polymorphic forms of hTdT. Specifically, single nucleotide polymorphisms (SNPs) that result in a protein amino acid change were considered. TdT plays an important role in antigen receptor diversification. This chapter will introduce hTdT properties, its function in V(D)J recombination and its role in immune-related disorders.

Section 1.1 Biochemical properties of TdT

Terminal deoxynucleotidyl transferase (TdT) is a 58kDa nucleus localized X-family DNA polymerase that is able to add non-templated (N) nucleotides to a free 3' hydroxyl group of single stranded DNA during V(D)J recombination. It contributes to the diversity at the coding joints between V, D and J gene segments. TdT is the only member of X-family of DNA polymerases that is template-independent. Other family members including DNA polymerase β, DNA polymerase λ and DNA polymerase μ require a double-stranded DNA substrate to catalyze the phosphoryl transfer reaction, (1,2).

1.1.1 TdT Expression and Regulation

TdT was the first DNA polymerase to be identified and was originally purified from calf thymus cell lysates (3,4). It is highly conserved in vertebrates, including humans, mice, zebrafish and rats (5). TdT expression is confined to primary lymphoid tissues, namely thymus and bone marrow, and its expression is tightly controlled at specific stages of lymphocyte development (6,7). TdT activity is tightly regulated at the transcriptional level, post-translational modifications and interaction with other proteins. At the transcriptional level, TdT expression was shown to correlate to the expression of the Recombination Activating Genes (RAG) that initiate the process of V(D)J recombination at B and T lymphocyte receptor loci (8). In addition, activation protein-1 (AP-1), a transcription factor, was found to down-regulate TdT mRNA expression via protein kinase C activation (9). Post-translational modifications of TdT were also suggested to modulate its expression. For instance, *in vivo* labeling studies demonstrated TdT phosphorylation in human lymphoblastoid cells however, the downstream effect of

this phosphorylation on TdT activity was not established (10, 1). TdT has been shown to interact with other proteins, such as the Ku proteins that participate in non-homologous end joining (NHEJ) processes during V(D)J recombination. Sandor *et.al* have shown that random nucleotide additions resulting from TdT activity are significantly longer (over 10 nucleotides long) in Ku80 knockout Chinese hamster ovary cells compared to Ku80+/+ cells, thus demonstrating Ku80 ability to modulate TdT catalytic activity *in vivo* (11).

1.1.2 Isoforms of TdT

Genomic organization of human TdT gene is complex. The TdT gene is composed of thirteen exons. A total of three human mRNA TdT splice variants have been identified. These translate into three mature protein isoforms: short (TdTS, 11 exons, 509 amino acids) and two long (TdTL1, TdTL2). TdTL1 isoform contains a 9 amino acid insert, while L2 isoform contains a 17 amino acid insert. Both long isoforms are found to possess 3' to 5' *in vitro* exonuclease activity that allows for nucleotide removal (as well as addition), while the short isoform is associated primarily with nucleotide additions (12). Normal B and T lymphocytes express exclusively hTdTS and hTdTL2, whereas hTdTL1 expression appears to be restricted to transformed lymphoid cell lines (2). The significance of this finding has not been explored further. Thus, the length of N-nucleotide regions during V(D)J recombination is thought to be regulated through the combined activity of its polymerase and exonuclease activities.

1.1.3 TdT protein domain structure

Two functionally independent human TdT regions have been identified: breast cancer susceptibility protein BRCA1 C-terminal (BRCT) domain at the N-terminus and the polymerase β-like domain at the C-terminus (Figure 1.1). BRCT domain is involved in protein-protein and protein-DNA interactions during DNA repair and cell cycle check-point pathways (13). The polβ-like domain is the catalytic core of the enzyme and contains the active site of the phosphoryl transfer reaction (1,2). In addition, a nuclear localization signal (NLS) motif is found at the N-terminus. To date no protein structure of hTdT has been published. However, the crystal structure of the catalytic core of murine TdT (mTdT) that shares 78% amino acid sequence homology with the hTdT has been solved by Delarue *et.al* (5). I used this established murine TdT structure as a model of human TdT structure in my thesis.

According to the solved crystal structure of murine TdT at 2.35Å resolution, TdT tertiary structure is largely α helical and contains a central antiparallel β sheet, where the active site resides, along with two unstructured regions (Loop1 and 2) (5). TdT catalytic core is further subdivided into four domains named: 8 kDa, finger, palm, and thumb domains. The four domains assume a three dimensional right-handed shape (Figure 1.2). The four subdomains resemble the ones first described in polymerase β, and thus are termed polβ-like domains (15). The phosphoryl transfer reaction takes place at the palm subdomain. The enzyme's active site is composed of three highly conserved aspartic acid residues and a divalent cation binding pocket. The finger subdomain aligns the incoming nucleotide and the DNA primer strand in proper orientation for phosphoryl transfer

reaction. The role of the thumb subdomain is to position the DNA primer strand to accept an incoming nucleotide prior to catalysis and to translocate TdT to next base position after catalysis. The 8 kDa subdomain contacts the thumb subdomain to create TdT's ring-like shape that allows DNA primer strand and incoming nucleotide to reach the active site (1-2, 5, 14).

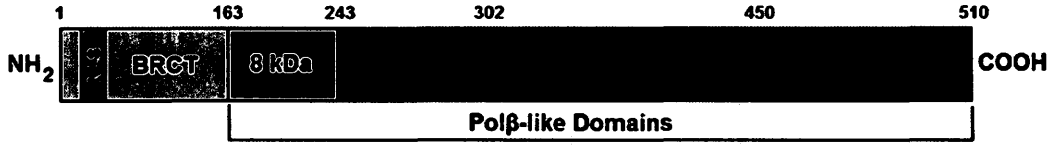


Figure 1.1: Domain structure of TdT protein. TdT is composed of four subdomains: 8 kDa (residues 163-243), finger (residues 243-302), palm (residues 302-450) and thumb (residues 450-510). Amino acid numbering is shown on top of the figure (coinciding with murine TdT). Nuclrear localization signal (NLS) located at the N-terminus.

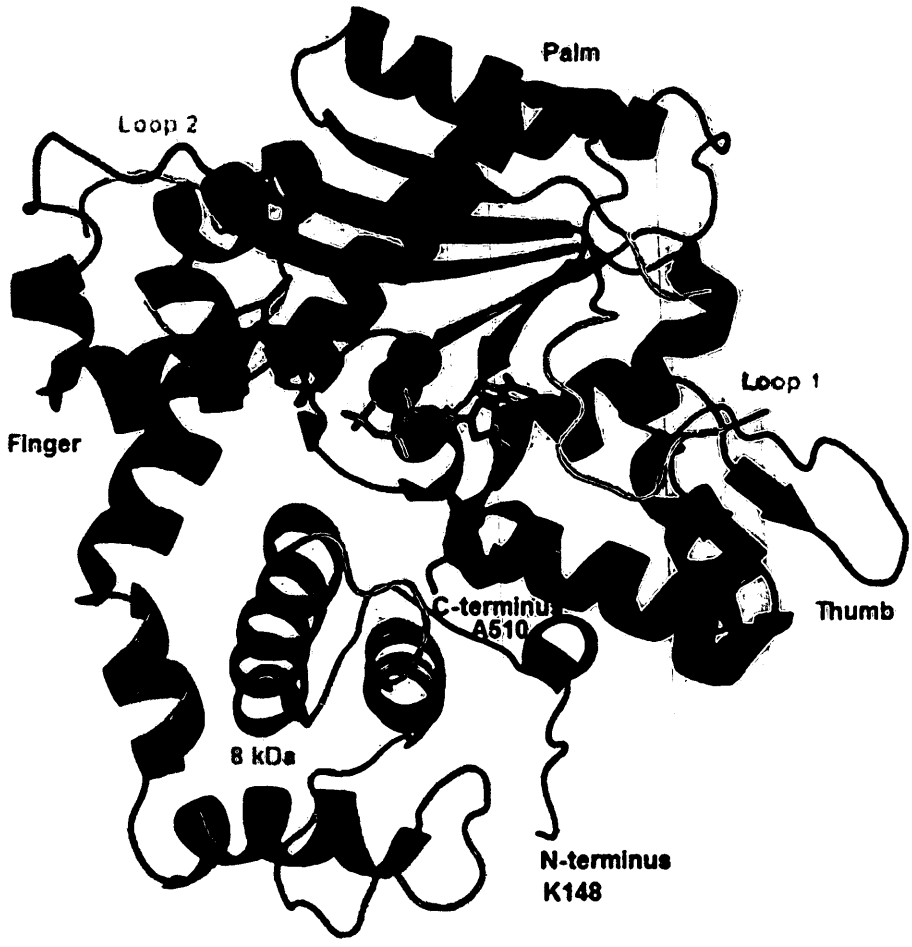


Figure 1.2: Crystal structure of murine TdT (PDB code 1KEJ). TdT is composed of four subdomains: 8 kD, finger, palm and thumb. The crystal structure excludes the BRCT domain at the N-terminus, thus only residues 148-510 are shown. Two unstructured loop regions have been determined (Loop1 and Loop2) within the palm subdomain, shown in yellow. The incoming nucleotide substrate, ddATP, is shown in orange. Two cobalt ions are colored cyan. Figures generated using *PyMol* software.

1.1.4 Mechanism of polymerization

The active site of X-family DNA polymerases is highly conserved and is found at the palm subdomain. It consists of catalytic aspartates that coordinate the two metals involved in the nucleotidyl transferase reaction (14). A detailed kinetic mechanism of template-independent polymerization by TdT has not been reported to date. Nevertheless, according to other established mechanisms of nucleotidyl transfer reaction of X-family DNA polymerases an aspartic acid residue abstracts a proton from the 3' hydroxyl group of DNA primer strand resulting in a reactive nucleophile. The nucleophile attacks the α-phosphate of the incoming nucleotide, leading to the formation of a phosphodiester bond. The positively charged metal ions stabilize the reaction transition state as well as facilitating the release of the negatively charged pyrophosphate group (15, 16). A key structural difference between the template dependent DNA polymerases, such as polymerase β, and TdT is the presence of a 16 amino acid 'lariat-like' loop that does not allow TdT to interact with double stranded DNA. The lariat-like loop is believed to physically block the diffusion of duplex DNA primer into the active site of TdT (5, 17).

Section 1.2: Generation of antigen receptor diversity

1.2.1 Structure of B and T cell receptors

The B cell receptor is a glycoprotein consisting of four polypeptide chains: two identical heavy chains and two identical light chains, linked to each other via disulfide bonds and hydrophobic interactions (6) (Figure 1.3). The gene loci of the BCR encode for a heavy chain and two light chains. In humans, the B cell immunoglobulin genes are found at three loci: the heavy chain locus (IgH) on chromosome 14, the κ light chain

(Igκ) locus on chromosome 2 and the λ light chain locus (Igλ) on chromosome 22. Each chain loci contains a constant region with limited diversity and a variable region containing three hypervariable regions, also called complementary-determining regions (CDRs) (6, 18). The IgH variable region is further subdivided into gene segments: 51 V_H segments, 27 D_H segments and 6 J_H segments (counting human functional genes). The Igκ and Igλ variable regions only contain V and J gene segments. Human Igκ has approximately 40 V_κ and 5 J_κ functional gene segments, while Igλ has about 31 V_λ and 4 J_λ segments (6, 20). The paired variable regions of heavy and light chain are responsible for the specificity and diversity of the antigen-binding site. The presence of multiple V, D and J gene segments and their combinatorial joining, as well as nucleotide deletions and additions at the joint junctions, form a distinctive DNA sequence at the variable region. An appropriate antigen-receptor is selected based on its recognition and binding to an epitope of an encountered pathogen, followed by proliferation of the antigen-activated lymphocytes and their differentiation into effector cells that function to eliminate the pathogen (19).

The human T cell receptor (TCR) consists of the following gene loci: TCRα (chromosome 14), TCRβ (chromosome 7), TCRγ (chromosome 7) and TCRδ (chromosome 14) chains. The TCR is a trans-membrane glycoprotein composed of either an α and a β chain (TCRαβ) or a γ and a δ chain (TCRγδ) (6). Each polypeptide chain contains a variable and a constant region, similarly to the BCR. For instance, the variable region of the α chain is divided into approximately 70 V_α and 61 J_α gene segments while

the β chain consist of 52 V_β, 2 D_β and 13 J_β gene segments (6, 20). The TCR loci undergo similar processes at the variable regions to reconstitute a distinctive antigen-binding site.

1.2.2 Mechanism of V(D)J recombination

The diversity of B and T cell antigen-receptors is achieved through gene rearrangements of V, D and J segments. This process contributes significantly to the generation of a diverse immune repertoire. V(D)J recombination involves four consecutive phases: the recognition of recombination signal sequences (RSS), followed by single-strand DNA nicking at these sites, coding-end modifications and re-joining of modified ends via DNA repair pathway (20, 21) (Figure 1.4). Nuclear proteins called Recombination Activating Gene products, RAG1 and RAG2, recognize and align evolutionary-conserved RSS flanking V, D and J gene segments. RAG proteins are exclusively expressed in lymphoid tissues undergoing V(D)J recombination (23, 24). There are two types of RSSs, one of each flanking gene segments that can be joined together. Each of the two types has a heptamer consensus sequence 5′-CACAGTG, a non-conserved spacer, and a nanomer consensus sequence 5′-ACAAAAACC. The spacer lengths differ in the two types of RSS. It is either 12±1 base pairs or a 23±1 base pairs. In order for two gene segments to be joined, the RSS of the gene segments must be of different spacer lengths. These sequences function as recognition sites for the RAG complex and ensure that the joining order of gene segments occurs between appropriate gene clusters (V-D-J or V-J, but not V-V, D-D or J-J). This is known as the '12/23 rule', meaning that recombination can only occur between a 12 base pair spacer RSS and a 23 base pair spacer RSS (6, 25). The heptamer and nanomer sequences may vary from the

consensus sequence, however the non-consensus RSS are used less frequently than the consensus RSS in the recombination reaction (26-28).

The RAG1/2 complex recognizes a pair of RSS and cleaves at the border between a gene segment and a specific RSS, thus producing a DNA nick at the phosphodiester bond of the 5' end of RSS. The RAG complex was shown to interact with DNA histones to render the gene segments accessible at RSS regions, however the specific mechanism remains unclear (29, 30). The DNA strand break exposes a 3' hydroxyl group at the coding segment end that serves as a nucleophile. The nick is converted to a covalent hairpin structure by a direct trans-esterification reaction, in which the 3' hydroxyl group attacks the 5' end of the complementary strand to form a new phosphodiester bond (20, 31, 32). The result is the formation of two DNA hairpins at the end of gene segments encoding the variable segments to be joined and two blunt signal ends which are the RSS and downstream DNA. When the variable gene segments to be joined are in the same transcription orientation, the DNA region containing the RSS is thus excised, the RSS are ligated forming what has been termed a "diamond structure" (Figure 2.5) forming a circular DNA, known as the signal joint. (These signal joint circles are generally lost during subsequent cellular divisions.) The hairpin structures at the coding ends of the variable domains remain associated with the postcleavage complex. The postcleavage complex which includes RAG1 and RAG2 proteins, is thought to keep the double stranded DNA breaks in close proximity to allow further coding end processing (33, 34).

The hairpin structure is opened and processed at the coding joint via non-homologous end joining proteins (NHEJ) that are associated with cellular (genetic) DNA

damage responses (35, 36). NHEJ pathway joins double stranded DNA breaks in a homology-independent manner, however short homologous sequences of few nucleotides long may be used to align DNA ends prior to ligation. The importance of NHEJ process is evident in animal models with severe combined immunodeficiency (SCID). The SCID defect as a result of abnormal NHEJ components including the Ku and DNA-PK^{cs} proteins (37, 38). The processing of the coding joint is mediated through the recruitment of Ku heterodimer (named Ku70 and Ku86) to the double stranded DNA breaks (39). Ku proteins recruit the DNA-dependent protein kinase catalytic subunit (DNA-PK^{cs}) that then activates the nuclease Artemis, that has been implicated in hairpin opening reaction, via phosphorylation (40). *In vitro* studies have also implicated RAG proteins and Mre11 complex as possible nucleases in the hairpin opening process (41, 42). Hairpin opening may occur off-center, generating 3' or 5' overhang, or at the tip, generating blunt ends. The final coding joint usually has evidence of processing of the open hairpins before or during the process of the final repair and ligation of these ends. Palindromic (P) nucleotides, which can be found (*in vivo* occur at about 5-10% of the coding joints) in the final coding joint, occur when the hairpin is opened off-center, and joined (Figure 1.4). The P nucleotides are also called P-additions, and they are typically 1-2 nucleotides long (36). Nucleotide deletions also may, and usually do, occur at the coding ends, however the nuclease(s) responsible remains undetermined.

The coding joints are usually processed by other enzymes including TdT which adds 1-15 random nucleotides (preferentially C and G *in vivo*) in a template-independent manner to free 3' hydroxyl group ends of DNA (1, 43-45). As mentioned earlier, TdT is

expressed in primary lymphoid tissues, namely thymus and bone marrow (6,7). In B cells TdT is expressed in pro-B stage when the heavy chain genes begin to rearrange. Its expression ceases before the rearrangement of the light chain genes during small pre-B stage, thus the variable region do not contain N additions (although exceptions were noted) (1, 46, 47). TdT is expressed in T cells at two stages as α and β chain rearrangements occur. The majority of BCR and TCR contain N additions in their variable domains. Moreover, N additions have been reported in developing human fetus however the number of N nucleotides is significantly lower as compared to adults (48). Fetal mouse B and T cell repertoire contains little to no N additions and TdT has been shown to be expressed only after birth (49).

The final joining of the modified coding ends is somewhat unclear. It is believed that the modified ends (with one or more P, N, deletions modifications) align through microhomology-mediated NHEJ that involves annealing of a complementary stretch of nucleotides present at processed coding end overhangs. Unpaired nucleotides are believed to be excised and template-dependent DNA polymerase (Polymerase μ) adds templated nucleotides to the gaps at the coding joints (50-52). The final ligation step is catalyzed by DNA ligase IV in complex with XRCC4 (X-ray-repair cross-complementing protein 4) to form the complete recombined coding joint (53, 54). Combinatorial and junctional diversity achieved through the process of V(D)J recombination generates a large antigen receptor repertoire, in the order of 10^{10}-10^{11} unique receptors (36), that is required for the broad adaptive immunity present in all vertebrates (other than the jawless fish).

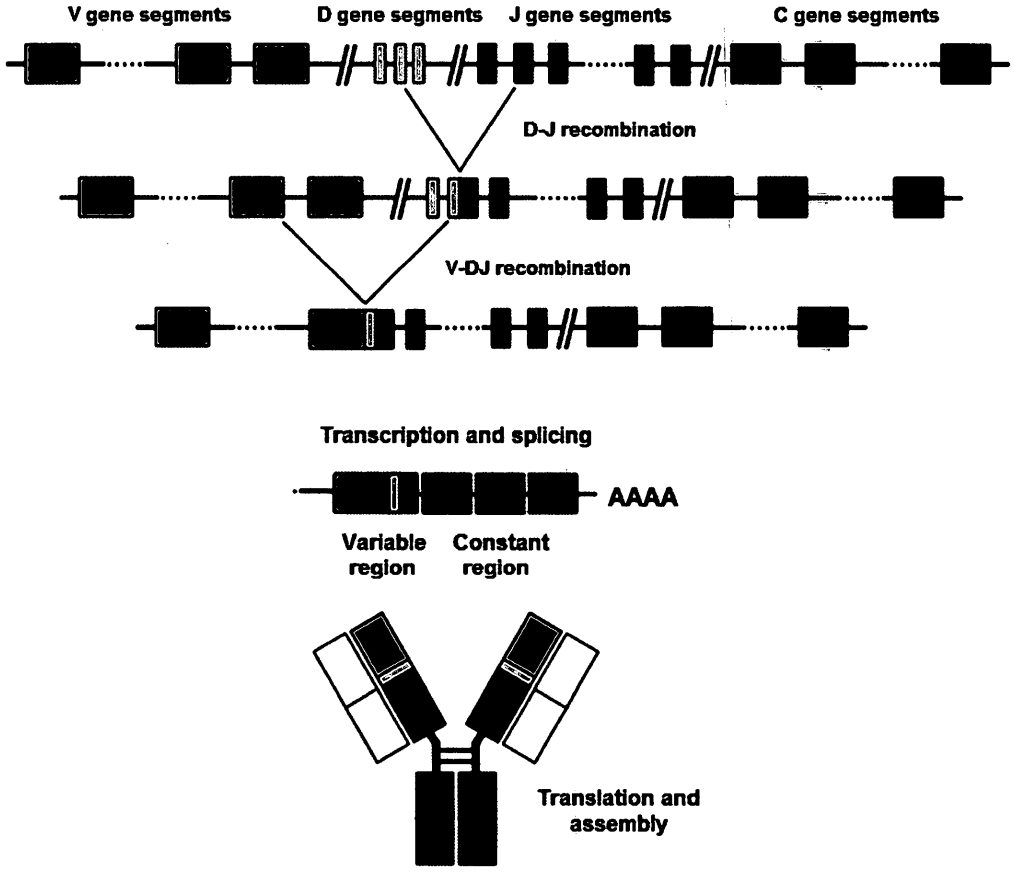


Figure 1.3: V(D)J recombination overview of immunoglobulin heavy chain locus.
The germline configuration of heavy chain variable region of B cell receptor consists of multiple V, D and J gene segments. The heavy chain locus undergoes D-J recombination followed by V-DJ recombination. The final product is the B cell receptor composed of two identical heavy chains and two identical light chains. The variable region of heavy and light chain is the antigen-binding site.

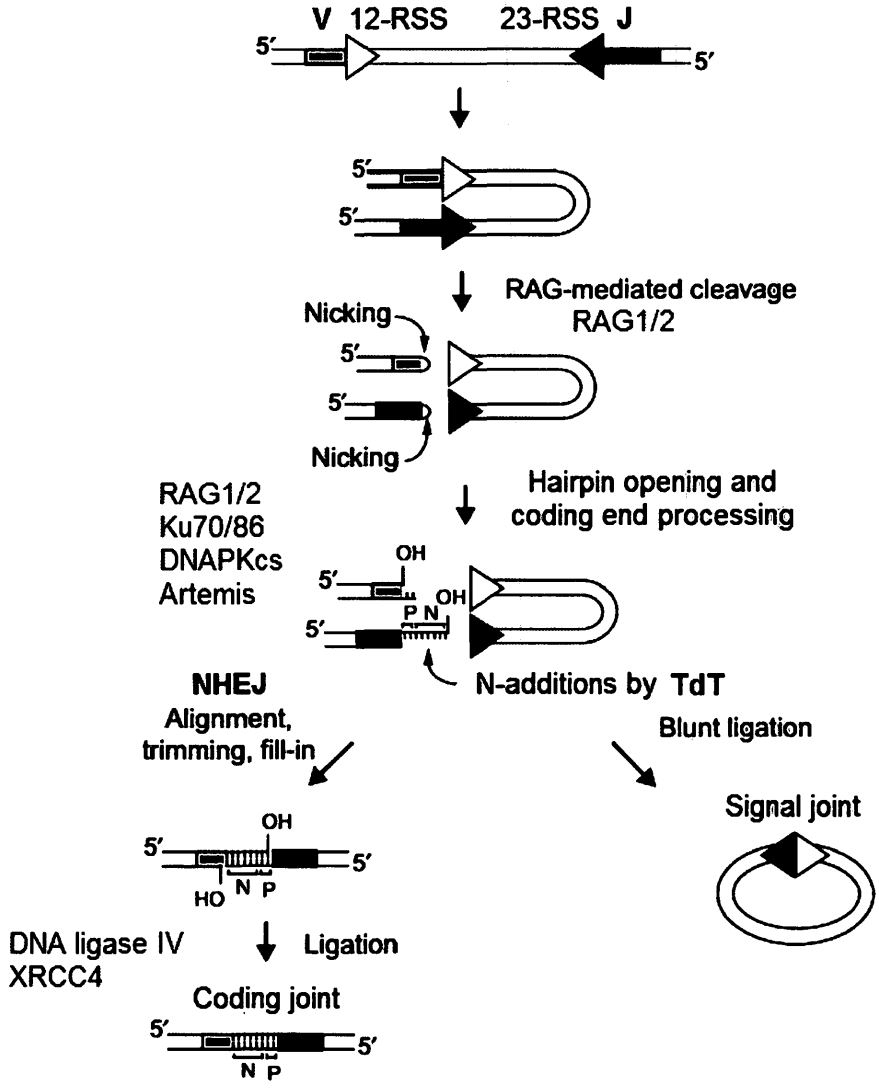


Figure 1.4: Schematic overview of a single V-J recombination event and associated protein factors, resulting in the generation of coding joint diversity.
RAG complex recognize recombination signal sequences (12-RSS and 23-RSS, represented by triangles) and catalyze DNA nicking at those sites. Non-homologoes end joining components (Ku70/86, DNAPKcs, Artemis) are recruited to the site of DNA damage. The nick is converted to a hairpin structure that is processed. Palindromic (P) nucleotide additions result from off-center hairpin opening. TdT catalyzes random template independent N-additions at the coding end. NHEJ proceeds to align, trim and fill in gaps at the coding end. Ligation via DNA ligase IV in complex with XRCC4 proceeds to generate a unique coding joint. Circular signal joint is generated following blunt ligation of DNA found in between V and J gene segments. Figure adapted from (22).

Section 1.3: TdT and its role in autoimmunity

TdT's connection to immune disorders has been widely investigated in mice animal models. Delayed onset of TdT expression during ontogenesis is evident due to the difference in N-region diversity in B and T cell receptor repertoire among newborn versus adult mice (45, 55). Gavin and Bevan were among the first to demonstrate that TdT-knockout mice lacking N-additions resulted in significant loss of BCR specificity towards peptide antigens compared to TdT positive mice. Moreover, TdT positive animals were able to recognize a broader spectrum of antigen mimics than TdT-knockout mice (56). Numerous studies have been conducted in autoimmune mouse models. B and T cell reactivity towards self antigens results in autoimmunity. In this case, B and T cell receptors are activated and initiate an adaptive immune response against self molecules in the absence of a true pathogen. TdT deficient lupus-prone mice demonstrated prolonged lifespan along with decreased severity of disease symptoms, specifically reduction of acute liver inflammation and occurrence of skin lesions associated with lupus, suggesting that N addition resulted in specificities in their BCRs that in these mice, rendered them more susceptible to early onset lupus (57, 58). Similarly, research with TdT-deficient nonobese diabetic mice showed a decrease in type I diabetes incidence when compared to littermates with active TdT and hence N addition in their antigen receptors (59). Molano *et.al.* argue that TdT presence correlates with higher autoantibody production causing disease (60). Thus, there is strong evidence that TdT-deficient mice experience a milder course or a lower incidence of autoimmune diseases.

TdT expression was also shown to be linked to cancer and response to chemotherapy. TdT overexpression is noted in B and T cell acute lymphocytic leukemias (ALL) and acute myeloid leukemia (AML). More than 90% of ALL leukemic cells show high TdT activity. In fact, remission rates of TdT-positive patients are two-fold lower as compared to TdT-negative patients (2, 61, 62). The exact role of TdT in these cancers has not been determined, however research is being conducted to develop anti-cancer drugs in the form of TdT-specific inhibitors. For instance, synthetic non-nucleoside diketo hexenoic acid (DKHA) derivative was shown to allosterically inhibit TdT activity and suppress cell proliferation of TdT-positive leukemic cells (62). Therefore, nontoxic selective TdT inhibitors have the potential to be employed in anti-leukemic chemotherapy.

Section 1.4 Single nucleotide polymorphisms

Genetic variations in the form of single nucleotide polymorphisms (SNPs) are common in the human population. These variations can be found in an appreciable (>1%) number of individuals. In fact, SNPs have been estimated to occur in 1 out of every 1,000 bases in the human genome (63, 64). SNPs can occur within gene's introns or exons. Those that are located within an exon may affect the protein-coding region. SNPs are divided into two general types: synonymous and non-synonymous. Synonymous SNPs do not affect the amino acid sequence of the translated protein, while non-synonymous SNPs result in a change in amino acid. Non-synonymous SNPs located at the protein-coding region have the potential to alter the structure and/or function of the encoded protein. SNPs have been link to disease susceptibility. One of the notable examples is sickle cell

anemia that results from a non-synonymous SNP in β-globin gene. A glutamine to valine amino acid change (A to T base change) affects haemoglobin polymerization with itself to form protein aggregates under low oxygen conditions. This in turn distorts red blood cell shape and decreases their elasticity, leading to ischemia (65, 66).

To date hTdT SNPs have not been studied. However, research has been conducted on RAG-1 single nucleotide changes. As mentioned earlier, RAG proteins are the key components responsible for initiation of V(D)J recombination. One type of Omenn Syndrome, resulting in severe immunodeficiency characterized by activated and self-reactive T-cells along with low levels of lymphocytes in the blood, has been shown to be associated with RAG protein SNPs. For instance, Villa *et. al.* identified specific non-synonymous RAG-1 SNPs in Omenn syndrome patients causing decreased DNA binding activity to recombination specific signals, which lower the overall efficiency of V(D)J recombination (67, 68).

Due to advances in genome sequencing technologies, SNP databases are continuously updated to include recent genotyping data. There are a number of SNP databases available, such as the National Center for Biotechnology Information (NCBI) SNP database and the international HapMap Project database. Genetic variation in the form of SNPs is thought to allow researchers to identify genes that may predispose individuals to disease, thus may be potentially used as markers in personalized medicine (63).

Section 1.5 Thesis objectives

My thesis work aims to characterize the enzymatic activities of naturally occurring human TdT polymorphisms that result in single amino acid change within TdT short isoform. Junctional diversity generated by TdT during the process of V(D)J recombination is crucial in obtaining widely diverse antigen receptor repertoire of the adaptive immunity. It is now evident that genetic variations in the form of SNPs are able to contribute to individual differences in a population. Functional activity differences in various polymorphic forms of hTdT may affect the adaptive immune response of individual human carriers to foreign pathogens and susceptibility to immune-related disorders. Therefore, the goal of my thesis is to explore the possible effects of hTdT SNPs on V(D)J recombination and subsequent effects on the level of diversity within the adaptive immune system. My thesis includes identification of hTdT SNPs that may potentially alter the enzyme's activity and characterization of their *in vitro* polymerase function. *In vivo* V(D)J recombination assays were conducted to investigate the ability of hTdT variants to generate a diverse coding joint population as compared to the wild type polymerase and the potential effects of hTdT SNPs on the antigen receptor diversity.

Chapter 2: Materials and Methods

Section 2.1 Selecting hTdT SNPs for study

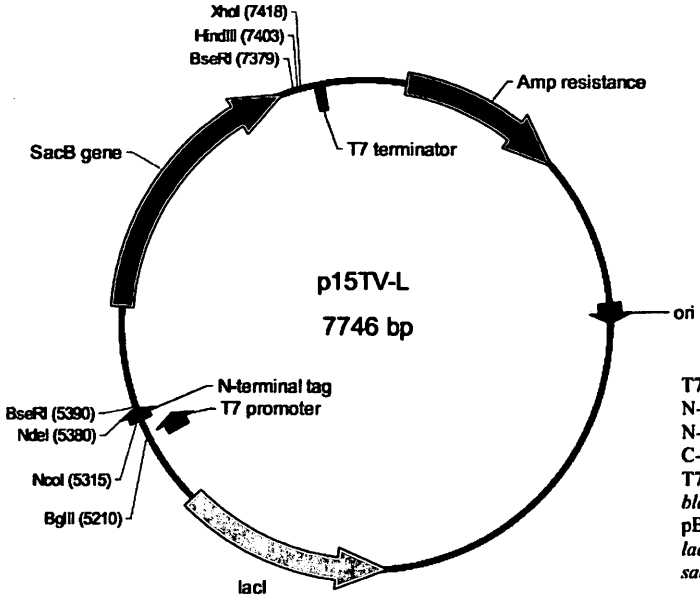
The National Center for Biotechnology Information (NCBI) SNP database was used in the selection of potential hTdT short isoform SNP candidates for the study within the coding region of the protein (NCBI reference NP_004079.3). First, all known hTdT SNPs in the human population that resulted in a single amino acid change of the protein were identified. The identified SNPs were classified based on a number of criteria to select for those candidates hypothesized to affect hTdT function, namely chemical nature of the amino acid change and modeling of the relative proximity of the SNP to the TdT active site using *PyMol* software. In addition, TdT multi-sequence alignment of other species was completed. Based on the above criteria, the best hTdTS SNP candidates were chosen for this study. The results of SNP analysis are detailed in Chapter 3, Table 3.1.

Section 2.2 Plasmid hTdTS DNA construction

2.2.1 Bacterial expression vector construction via site-directed mutagenesis

Bacterial p15TV-L expression vector containing wild-type hTdTS cDNA in place of SacB gene (GenBank accession EF456736) was provided by Irakli Dzneladze, a previous laboratory member. The expression vector contains an N-terminal fusion tag of six Histidine residues (6xHis) and has T7 promoter driven expression of recombinant protein, in this case wild-type hTdTS (Figure 2.1). The expression of recombinant protein is under the regulatory response to lactose. In the absence of lactose, the Lac repressor protein encoded by the *lacI* gene prevents the transcription of downstream gene.

However, when lactose or its metabolite Isopropyl β-D-1-thiogalactopyranoside (IPTG) is present it binds to the Lac repressor and in turn triggers transcription of the downstream gene and subsequent translation to its protein product.



T7 promoter	5230-5246
N-terminal tag	5317-5379
N-terminal cloning site	5365-5379
C-terminal cloning site	7395-7409
T7 terminator	7488-7534
bla coding sequence	209-1066
pBR322 origin	1827
lacI coding sequence	3764-4843
sacB coding sequence	5901-7319

Figure 2.1: p15TV-L vector map (GenBank accession EF456736).
The vector contains T7 promoter region, N-terminal tag, two cloning sites, T7 terminator, ampicillin resistance gene, pBR322 origin of replication and *lacI* gene that codes for lac repressor. Wild type hTdTS cDNA was inserted in place of SacB gene, downstream of T7 promoter region.

Site directed mutagenesis (SDM) protocol (QuickChange® by Agilent Technologies) was utilized to produce point mutations in wild-type hTdTS p15TV-L expression vector corresponding to the chosen six hTdT SNP candidates.

Table 2.1: Site-directed mutagenesis primers used to generate point mutations in wild-type hTdTS cDNA, corresponding to selected hTdTS SNPs.

Primer Name	Primer Sequence	Melting Temp.
A445T_F	5′ CTACGAGCGTCGTGCCTTTACCCTGTTGGGATGG	70.8°C
A445T_R	5′ CCATCCCAACAGGGTAAAGGCACGACGCTCGTAG	70.8°C
L397S_F	5′ CCTAGCAGGAAGGTTGATGCTTCGGATCATTTTCAAAAGTGC	68.8°C
L397S_R	5′ GCACTTTTGAAAATGATCCGAAGCATCAACCTTCCTGCTAGG	68.8°C
R431C_F	5′ GGAAGGCCATCTGTGTGGATTTAGTTCTGTGCCCC	69.5°C
R431C_R	5′ GCACAGAACTAAATCCACACAGATGGCCTTCCAGG	68.8°C
R460Q_F	5′ GCAGTTTGAGAGAGACCTCCAGCGCTATGCCAC	69.8°C
R460Q_R	5′ GTGGCATAGCGCTGGAGGTCTCTCTCAAACTGC	69.8°C
T450S_F	5′ GCCCTGTTGGGATGGTCTGGCTCCCGGCAG	72.9°C
T450S_R	5′ CTGCCGGGAGCCAGACCATCCCAACAGGGC	72.9°C
D280H_F	5′ CTCTGAGTAAAGTAAGGTCGCACAAAAGCCTG	73.1°C
D280H_R	5′ CGACCTTACTTTACTCAGAGAAGACTTTGG	67.5°C

Each SDM PCR reaction was set up as follows; 1x reaction buffer (10mM KCl, 10mM $(NH_4)_2SO_4$, 20mM Tris-HCl (pH 8.8), 2mM $MgSO_4$, 1% Triton® X-100, 0.1 mg/ml BSA), 0.2mM dNTP mix, 100ng WT hTdTS p15TV-L vector, 150ng specific primer DNA, 0.4µl *PfuTurbo* DNA polymerase (2.5U/µl) and dH_2O up to 20µl. The reactions were cycled according to the PCR program in Table 2.2. Each reaction was allowed to proceed through 6 PCR cycles, followed by combination of Forward and Reverse primer SDM reactions into one reaction. An extra 0.5µl *PfuTurbo* DNA polymerase (2.5U/µl) was added to the combined reaction and total volume was adjusted

to 50µl using dH_2O. The combined PCR reaction was allowed to proceed for the remaining 20 PCR cycles.

Table 2.2: Site-directed mutagenesis PCR program parameters.

Step		**Temperature**	**Duration**
Denaturation		95°C	5 minutes
26 cycles	Denaturation	95°C	45 seconds
	Annealing	55°C	1 minute
	Extension	68°C	20 minutes
Final extension		68°C	30 minutes

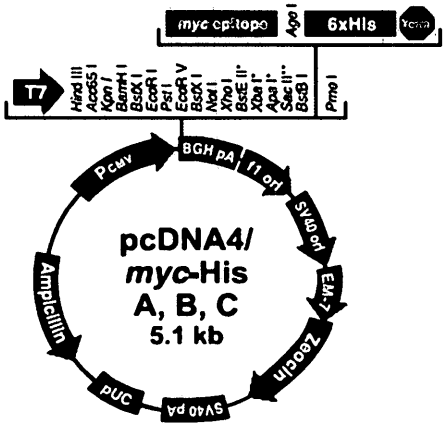
Completed SDM PCR reactions were then digested with 20U of DpnI restriction enzyme (New England BioLabs) at 37°C overnight to degrade non-mutated, methylated parental DNA template in the reactions. The digested DNA samples were ran on 0.7% agarose gel followed by DNA gel extraction protocol (Qiagen's QIAEXII gel extraction). The DNA was eluted into 20µl dH_2O. Next, bacterial transformation of 25µl of *E.Coli* DH5α electrocompetent cells with 2.5µl of the DpnI-digested DNA was performed at 1700V (Biorad's Gene Pulser electoporator). The transformation mix was plated on Luria Bertani (LB) agar plates (10g/l NaCl, 5g/l yeast extract, 10g/l tryptone and 15g/l agar in dH_2O) containing 100µg/ml ampicillin. Bacterial plates were incubated at 37°C overnight. Bacterial colonies were picked from each plate and inoculated in 5ml LB media containing 100µg/ml ampicillin at 37°C overnight, shaking at 250rpm. The DNA from the p15TV-L expression vectors carrying SNPs of hTdTS cDNA were purified from the overnight culture using Qiagen's QIAprep Spin Miniprep kit. The purified vector DNA was ran on 0.7% agarose gel and SDM was verified through sequencing of the recovered hTdTS cDNA p15TV-L vectors using hTdTS-specific primer SQ-TdT-974

(5′CCTTGTCAGCTGTGTGACCA) capable of binding to nucleotide position 974 of hTdTS cDNA (note; the selected hTdTS SNPs are positioned downstream of nucleotide position 974). The generated p15TV-L hTdTS SNPs expression vectors were later used in hTdTS protein purification.

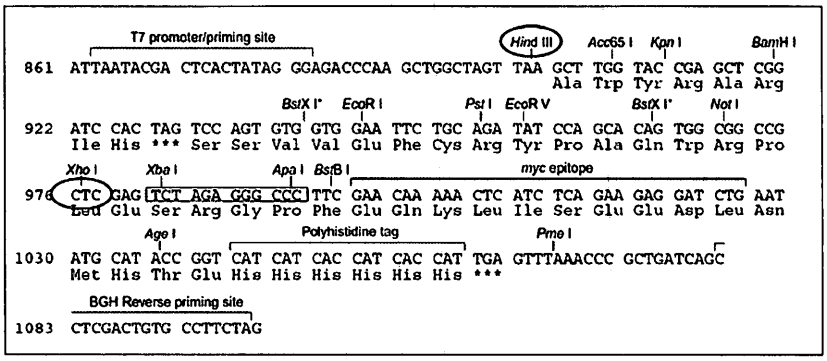
2.2.2 Mammalian expression vector construction via restriction site infusion based cloning

In vivo functional TdT activity studies required mammalian hTdTS expression vector. Cloning of hTdTS cDNA into pcDNA^{TM}4/myc-His A vector was performed via restriction site based infusion. This pcDNA vector is 5075 base pairs long and contains single HindIII and XhoI restriction cut sites within its multiple cloning region (Figure 2.2). Restriction endonuclease digest reactions of hTdTS wild type, A445T, L397S, R431C cDNA p15TV-L vectors and pcDNA^{TM}4/myc-His A vector were set up as follows. Uncut p15TV-L: 1µg WT hTdTS p15TV-L and 1xNEB Buffer 2 adjusted to 50µl with dH_2O. HindIII cut p15TV-L; 1µg hTdTS p15TV-L, 1xNEB Buffer 2, 10 units HindIII (NEB) adjusted to 50µl with dH_2O. XhoI cut p15TV-L; 1µg hTdTS p15TV-L, 1xNEB Buffer 4, 1xBSA, 10 units XhoI (NEB) adjusted to 50µl with dH_2O. Uncut pcDNA; 1µg pcDNA and 1xNEB Buffer 2 adjusted to 50µl with dH_2O. HindIII cut pcDNA; 1µg pcDNA, 1xNEB Buffer 2, 10 units HindIII (NEB) adjusted to 50µl with dH_2O. XhoI cut pcDNA; 1µg pcDNA, 1xNEB Buffer 4, 1xBSA, 10 units XhoI (NEB) adjusted to 50µl with dH_2O. HindIII and XhoI double digested pcDNA; 1µg pcDNA, 1xNEB Buffer 2, 1xBSA, 10 units HindIII (NEB), 10 units XhoI adjusted to 50µl with dH_2O. All restriction digest reactions were incubated at 37°C overnight.

A



B



Human cytomegalovirus immediate-early (CMV) promoter 209-863	High-level expression of recombinant protein in mammalian cells
T7 promoter 863-882	Proper orientation and sequencing of the insert
Multiple cloning site 902-999	Insertion of the gene and in frame cloning with *myc* epitope and 6XHis tag
myc epitope 997-1026	Detection of recombinant protein with anti-*myc* antibody
Polyhistidine tag (6XHis) 1042-1059	Purification of recombinant protein on metal chelating resin
Bovine growth hormone (BGH) polyadenylation signal 1085-1312	Efficient transcription termination and polyadenylation of mRNA
f1 origin 1358-1786	Rescue of single stranded DNA
SV40 early promoter and origin	Efficient, high level expression of Zeocin™ resistance gene
EM-7promoter + Zeocin™ resistance gene 2244-2618	Expression of Zeocin™ resistance gene for selections of transformants in *E.coli*
SV40 polyadenylation signal 2748-2878	Efficient transcription termination and polyadenylation of mRNA
pUC origin 3261-3934	High-copy number replication and growth in *E.coli*
Ampicillin resistance gene 4079-4939	Selection of transformants in *E.coli*

Figure 2.2: Features of pcDNA™4/myc-His A, B, C 5.1 kb vectors (Adapted from Invitrogen pcDNA™4/myc-His A, B, C User Manual, catalog no.V863-20).

A. Map of pcDNA™4/myc-His A, B, C vectors and the common functional features of the vectors (shown in the table).
B. Multiple cloning site of pcDNA™4/myc-His A. Hind III and Xho I cleavage sites used to clone hTdTS cDNA into the mammalian vector are circled in red. hTdTS cDNA containing pcDNA vector is about 6.5 Kb.

Bacterial p15TV-L expression vectors containing wild type, A445T, L397S and R431C hTdTS cDNA were used to amplify the cDNA region using PCR extension primers designed to contain HindIII and XhoI restriction endonuclease cut sites for later insertion into pcDNA backbone.

Table 2.3: PCR infusion primers used to amplify hTdTS cDNA. The forward primer contains HindIII restriction cut site (5′ AAGCTT), while the reverse primer contains XhoI restriction cut site (5′ CTCGAG), shown in bolded font below.

Primer name	Primer Sequence	Melting Temp.
pcDNA_Infusion_F	5′ GTTTAAACTT**AAGCTT**GCCATGGATCCACCACGAGCG	68.2°C
pcDNA_Infusion_R	5′ GCCCTCTAGA**CTCGAG**CTAGGCATTTCTTTCCCACGG	71.5°C

The PCR reaction was set up as follows; 0.1µg hTdTS cDNA p15TV-L vector, 1x *Pfu* buffer supplemented with $MgSO_4$ (20mM Tris-HCl (pH 8.8 at 25°C), 10mM $(NH_4)_2SO_4$, 10mM KCl, 0.1mg/ml BSA, 1% (v/v) Triton X-100, 2mM $MgSO_4$), 0.2mM dNTP mix, 50pmol pcDNA_Infusion_F, 50pmol pcDNA_Infusion_R, 0.5µl *Pfu* DNA polymerase (2.5U/µl) and dH_2O up to 50µl. The reactions were cycled according to the PCR program in Table 2.4.

Table 2.4: PCR program parameters for amplification of hTdTS cDNA based on restriction site based cloning.

Step		Temperature	Duration
Denaturation		95°C	5 minutes
30 cycles	Denaturation	95°C	45 seconds
	Annealing	75°C	1 minute
	Extension	70°C	5 minutes
Final extension		70°C	20 minutes

The generated hTdTS cDNA PCR products were then double digested with HindIII and XhoI restriction endonucleases. The restriction digest reaction was set up as follows; 50µl of generated hTdTS cDNA PCR product, 1xNEB buffer 2, 1xBSA, 20 units HindIII, 20 units XhoI and dH_2O up to 60µl. In addition, pcDNA vector was also set up for double digestion; pcDNA 2µg, 1xNEB buffer 2, 1xBSA, 20 units HindIII, 20 units XhoI and dH_2O up to 50µl. Reactions were incubated at 37°C overnight. The digested reactions were purified using Qiagen's QIAquick PCR purification protocol and the DNA was eluted in 15µl dH_2O.

The purified hTdTS cDNA PCR inserts digested with HindIII and XhoI restriction endonucleases were ligated into HindIII and XhoI digested pcDNA backbone at 1:3 ratio of vector to insert. The ligation reaction was set up as follows; 0.2µg HindIII/XhoI digested pcDNA, 0.6µg HindIII/XhoI digested hTdTS cDNA PCR insert, 1x T4 DNA Ligase reaction buffer (50mM Tris-HCl, 10mM $MgCl_2$, 1mM ATP, 10mM DTT, pH 7.5 at 25°C), 400 units T4 DNA Ligase (NEB) and dH_2O up to 20µl. The ligation reactions were incubated at 16°C overnight.

Bacterial transformation of 5µl of the ligation reaction was transformed into 50µl of DH5α electrocompetent cells at 1700V. The transformation mixture was plated on LB

agar plates containing 100μg/ml ampicillin. Bacterial plates were incubated at 37°C overnight. Bacterial colonies were picked from the plates and colony PCR was performed to identify the presence of the hTdTS cDNA insert in the pcDNA backbone. A single colony was picked and resuspended in 20μl dH_2O. PCR reaction was set up as follows; 2μl template bacterial colony, 10μl Fermentas 2xPCR Master mix (1x reaction buffer, 0.05U/μl *Taq* DNA polymerase, 4mM $MgCl_2$, 0.4mM dNTP mix), 50pmol pcDNA_ Infusion_F primer, 50pmol pcDNA_ Infusion_R primer and dH_2O up to 20μl. Colony PCR reactions were cycled according to parameters in Table 2.4. The generated PCR products were ran on 0.7% agarose gel to check for the purity and size of the DNA product. Verified bacterial colonies were inoculated in 5ml LB containing 100μg/ml ampicillin at 37°C overnight, shaking at 250rpm. Mammalian pcDNA vectors containing wild type, A445T, L397S, R431C hTdTS cDNA were purified from the overnight culture using Qiagen's QIAprep Spin Miniprep kit. The purified vector DNA was ran on 0.7% agarose gel and its size was verified though restriction digest with either HindIII or XhoI restriction endonucleases, as previously described. Single point mutations in hTdTS cDNA were verified through sequencing using SQ-TdT-974 primer. The correct hTdTS pcDNA clones were then amplified as according to Qiagen plasmid Maxi purification protocol.

Section 2.3 hTdT His-tag protein purification

Bacterial p15TV-L hTdTS SNPs expression vectors were transformed into *E.coli* BL21 cells to express the different protein variants. hTdT protein variants containing N-terminal His-tag were purified through nickel affinity chromatography and used for *in vitro* functional studies.

2.3.1 Heat-shock transformation of competent BL21 cells

BL21 competent cells were thawed on ice. hTdTs p15TV-L vector (50ng) was mixed with 25µl BL21 cells and incubated for 45 seconds at 42°C water bath. The transformation mix was then immediately placed on ice for 10 minutes followed by the addition of 200µl LB media to the transformation mix. The reaction was then incubated for 1 hour at 37°C, shaking at 250rpm. The reaction content was then transferred to bacterial 15ml conical tube containing 5ml LB with 100µg/ml ampicillin and incubated at 37°C overnight, shaking at 250rpm. Glycerol stock of overnight BL21 cultures were prepared by mixing 500µl bacterial culture with 500µl 50% (v/v) glycerol and stored at -80°C.

2.3.2 Induction of hTdT protein expression

Transformed BL21 overnight cultures (5ml culture) were transferred into 1 litre LB medium containing 100µg/ml ampicillin and grown at 37°C, shaking at 200rpm, for about 4 hours until the absorbance reading at 600nm (OD_{600}) reached 1.0. Protein expression was induced with the addition of 1ml of 1M IPTG to the 1 litre bacterial culture. The culture was then incubated at 16°C overnight, shaking at 200 rpm.

2.3.3 hTdT protein purification via nickel affinity chromatography

Note: the following purification protocol was performed at 4°C and all buffers were kept on ice.

IPTG-induced overnight BL21 culture was centrifuged at 1,500xg for 30 minutes at 4°C. The bacterial pellet was resuspended in freshly prepared Lysis buffer (50mM Tris pH 7.5, 500mM NaCl, 10mM imidazole, 1mM benzamidine, 0.5mM phenylmethylsulfonyl fluoride (PMSF), 1mM β-mercaptoethanol (BME)) (about 10ml Lysis buffer per 1g bacterial pellet). The bacterial cells were further lysed by sonication for 8 minutes (0.5 seconds pulse on, 2 seconds pulse off at 30% amplitude) on ice. A sample of the crude lysate (50µl) was taken for analysis using sodium dodecyl sulfate polyacrylamide (SDS-PAGE) gel electrophoresis. The lysed cells were centrifuged at 25,000xg for 30 minutes at 4°C. Nickel-nitrilotriacetic acid (Ni-NTA) metal-affinity chromatography resin was equilibrated with freshly prepared Wash buffer (50mM Tris pH 7.5, 500mM NaCl, 20mM imidazole, 1mM benzamidine, 0.5mM PMSF, 1mM BME). This was done by adding 50ml Wash buffer to the resin, centrifuging at 500xg for 5 minutes at 4°C and discarding the Wash buffer. The supernatant was then incubated with equilibrated Ni-NTA resin (3.5ml resin per 5g bacterial pellet) in a chromatography column for 1 hour at 4°C, slowly rocking, to allow for binding of the resin with His-tagged hTdT protein. The flow-through was decanted from the column (50µl of flow-through sample was taken for analysis on SDS-PAGE gel). The column was washed 3 times with 50ml of Wash buffer. The wash fractions were checked for protein absence by mixing 10µl of each wash fraction with 100µl Bradford reagent (Biorad). The wash

fraction samples were also collected for SDS-PAGE analysis. Elution of His-tagged hTdT protein was done by adding 500µl aliquots of freshly prepared Elution buffer (50mM Tris pH 7.5, 500mM NaCl, 500mM imidazole, 1mM benzamidine, 0.5mM PMSF, 1mM BME) to the column at and checking for protein presence using Biorad's Bradford reagent. The elution fractions that showed deep blue color using Bradford protein assay were collected in a15ml falcon tube on ice. The final elution sample was collected for analysis on SDS-PAGE gel.

2.3.4 hTdT protein dialysis into storage buffer

Eluted hTdT protein collected in Elution buffer was transferred into a dialysis bag and submerged into 500ml of TdT storage buffer (50mM KPO_4, 100mM NaCl, 1.43mM BME, 50% glycerol, 0.1% Triton X-100 pH 7.3 at 25°C). The storage buffer recipe was prepared as according to commercial recombinant TdT protein (NEB #M0315S) storage buffer. Dialysis proceeded at 4°C overnight with gentle stirring. The dialysis buffer was changed the next day followed by additional 3 hour dialysis. The dialyzed hTdT protein was aliquot on ice and stored as 100µl aliquots at -80°C. *In vitro* functional activity assays were carried out using freshly thawed hTdT protein aliquots.

2.3.5 Analytical SDS-PAGE electrophoresis of hTdT protein purification

Following protein purification, SDS-PAGE gel electrophoresis was performed as follows. The collected 50µl samples of crude lysate, flow-through, wash and elution were mixed with 10µl of 6x SDS-PAGE loading dye (0.03% bromophenol blue, 0.03% xylene cyanol FF, 60% glycerol, 1% SDS, 100mM EDTA (pH 7.6)). The samples were boiled at

95°C for 5 minutes and span down. A total of 10µl of each sample was loaded onto 12% SDS-PAGE gel along with 6µl of protein standard ladder (PageRuler Unstained Broad Range- Thermo Scientific #26630). The gel was ran for 1 hour at 180V followed by gel staining using Coomassie staining solution (50% methanol, 10% glacial acetic acid, 3g/l Coomassie Brilliant Blue dye in dH_2O) for 1 hour. The gel was destained overnight in Destain solution (40% methanol, 10% acetic acid in dH_2O).

Section 2.4 *In vitro* hTdT functional activity assays

2.4.1 General polymerase activity assay setup

The *in vitro* polymerization activity of purified wild type hTdTS was tested and compared to commercially available recombinant TdT from calf thymus (NEB #M0315S). TdT protein was incubated at 37°C with a single stranded oligonucleotide that is fluorescently labeled at the 5' DNA end with a Cy3 label (Cy3 oligo 5′ Cy3-CTACTGGTACTTCGATCTCTGGGGCCGTGACGC) along with dNTP mix in 1xTdT reaction buffer (50mM potassium acetate, 20mM Tris-acetate, 10mM magnesium acetate at pH 7.9 at 25°C), supplemented with 0.25mM $CoCl_2$. A typical TdT polymerization reaction included; 0.3nM Cy3 oligo, 0.15mM dNTP mix, 0.25mM $CoCl_2$, 1xTdT reaction buffer, 3.4µM purified wild type hTdT protein for a total reaction volume of 30µl with dH_2O. Reaction with recombinant calf TdT was set up in the same manner using 20 units of polymerase. Reaction master mix was incubated at 37°C to allow TdT to add nucleotides to substrate oligo and 30µl aliquots were taken at various time points. The reactions were stopped by the addition of 30µl TdT stopping buffer (95% formamide, 50mM EDTA, 0.05% bromophenol blue), thus inactivating the polymerase

via chelating its metal ion. All reactions were carried out using standardized amount of protein using Biorad's Bradford protein assay. The completed TdT reactions were then boiled for 5 minutes and placed on ice to denature any duplex DNA. The samples were then loaded onto 10% denaturing TBE-urea gel (40% acrylamide, 7M urea, 1xTBE (89mM Tris, 89mM boric acid, 2mM EDTA pH 8.0), 10% ammonium persulfate, Tetramethylethylenediamine). A total of 10µl of each TdT reaction was loaded on a gel along with 6µl of DNA ladder (GeneRuler^{TM} DNA Ladder, Ultra Low Range, Fermentas #SM1213). The denaturing TBE-urea gel was ran at 180V for 1.5 hours at 4°C. In order to visualize the DNA ladder, the gel was stained for 15 minutes in 1:10,000 dilution of SYBR® Green I Nucleic Acid Gel Stain (Invitrogen #S7567). The gel was then scanned by the Typhoon Imager to visualize Cy3-labelled DNA (scanning done at excitation wavelength of 532nm, emission wavelength of 580nm).

Note; preliminary purified hTdTS titration and dNTPs titration experiments were performed to determine the optimal amount of polymerase for *in vitro* functional assays.

2.4.2 dNTP substrate preference polymerase activity assay

The utilization of deoxyribonucleotide triphosphates (dNTPs) by hTdTS variants was tested. The assay was set-up as described in section 2.4.1, however instead of dNTP mix the reactions were supplemented with either of the single dNTPs; namely dATPs, dTTPs, dCTPs or dGTPs.

2.4.3 DNA substrate preference polymerase activity assay

The type of DNA substrate was varied under the same reaction condition (section 2.4.1). Purified hTdTS variants were supplied with either 0.3nM of single stranded 33 nucleotide long Cy3 oligo, a double stranded 33 nucleotide long Cy3 oligo having a blunt ends or a double stranded 33 nucleotide long Cy3 oligo having a 3' end overhang of 6 nucleotides (Table 2.5).

The double stranded Cy3 oligo substrates were obtained by annealing the single stranded Cy3-labelled oligo to unlabelled sequence complementary DNA oligo. The annealing reactions were performed by first boiling 1:3 equimolar ratio of single stranded Cy3 labelled oligo to single stranded unlabelled complementary oligo for 10 minutes. This reaction mix was allowed to slow-cool for about an hour. The annealed DNA product was then ready for use.

Table 2.5: DNA sequences of DNA substrates used for *in vitro* DNA substrate preference polymerase assay.

Single stranded Cy3 oligo	5′ Cy3-CTACTGGTACTTCGATCTCTGGGGCCGTGACGC
Double stranded blunt end Cy3 oligo	5′ Cy3-CTACTGGTACTTCGATCTCTGGGGCCGTGACGC GATGACCATGAAGCTAGAGACCCCGGCACTGCG
Double stranded 3' overhang end Cy3 oligo	5′ Cy3-CTACTGGTACTTCGATCTCTGGGGCCGTGACGC GATGACCATGAAGCTAGAGACCCCGGC

2.4.4 DNA substrate sequence preference polymerase activity assay

The polymerase activity of purified wild type hTdTS was investigated using two single stranded Cy3-labelled oligos with different base sequences. Oligo 1 sequence of 33 bases is found at the beginning of the human immunoglobulin heavy chain J2 region

(IGHJ2); whereas oligo 2 sequence of 33 bases is complementary to the oligo 1 sequence (Figure 2.3). The assay was set-up as described in section 2.4.1 supplemented with either 0.3nM Oligo 1 or Oligo 2.

Note; all of the *in vitro* polymerase activity assays were performed using single stranded Cy3-labelled oligo 1 sequence unless otherwise stated.

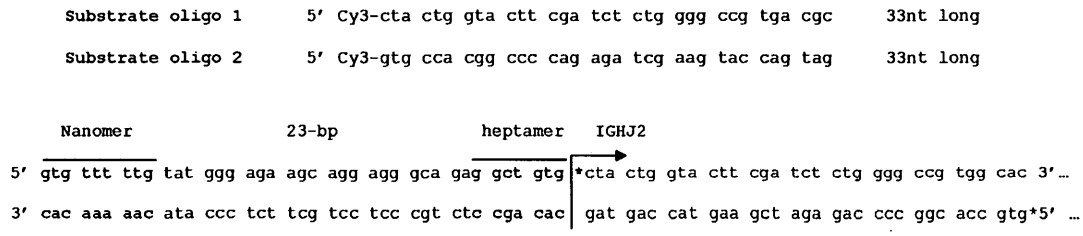


Figure 2.3: Sequences of Cy3-labelled oligo 1 and 2 used for *in vitro* DNA substrate sequence preference polymerase activity assay. Germline sequence of human immunoglobulin heavy chain J2 region flanked by 23-RSS region (adapted from NCBI database) shown at the bottom. Star (*) indicates Cy3 fluorescent label.

2.4.5 Cofactor preference polymerase activity assay

The type of divalent cofactor supplemented in the TdT reaction, specifically cobalt and magnesium was investigated. The assay was set-up as described in section 2.4.1, supplemented with either 0.25mM $CoCl_2$ or 0.25mM $MgCl_2$.

Section 2.5 *In vivo* V(D)J recombination assay

2.5.1 Assay overview

Analysis of *in vivo* functional activity differences between wild type hTdT and its polymorphic forms was carried out using an extrachromosomal V(D)J recombination substrate assay in human mammalian cells that lack endogenous TdT expression (Figure 2.4). The set up of the assay involved transfection of mammalian hTdTS expression vector (pcDNA^{TM}4/myc-His A), vectors expressing human RAG1 and RAG2 proteins (pEBG-hRAG1 and pEBG-hRAG2) along with recombination substrate plasmid (pGG51) into human mammalian cells not expressing TdT (human embryonic kidney cells; HEK293T). HEK293T cells were chosen for the assay due to their relative fast proliferation and transfection efficiency. The cells were tested for hTdT expression at the RNA and protein level (see below in section 2.5.2 and 2.5.3).

Human RAG1 and RAG2 expression vectors (pEBG-hRAG1 and hRAG2) were obtained from Dr. Patricia Q. Cortes, Mount Sinai Medical School, New York. These are mammalian expression vectors that contain human RAG1 core and C-terminus cDNA (amino acid 384-1040) and human RAG2 full length cDNA (amino acid 1-527). pEBG vector contains human elongation factor 1alpha (EF-1α) promoter. It is a glutathione-S-transferase (GST) fusion vector, thus hRAG1 and hRAG2 proteins contain a GST tag. The presence of the tag was previously shown not to affect recombination efficiency (69).

Once RAG proteins and hTdT are expressed in the mammalian cells following transfection, the recombination substrate plasmid is able to undergo recombination similarly to V(D)J recombination process in T and B lymphocytes (Figure 2.5). The

substrate plasmid pGG51 was generously supplied by Dr. Patricia Q. Cortes at the Mt Sinai School of Medicine, New York and Dr. Michael R. Lieber at the University of Southern California, USA (70, 71) (Figure 2.6). The substrate is 6284 base pairs long, containing two recombination signal sequences (RSSs); RSS with a 12 base pair spacer (12-RSS) and a second RSS with a 23 base pair spacer (23-RSS). A prokaryotic transcriptional termination sequence is found in between 12-RSS and 23-RSS regions which causes transcription elongation complex to pause and fall off the DNA template due to a stem-loop structure generated by its GC-rich sequence. The pGG51 plasmid contains a gene encoding for chloramphenicol acetyltransferase (*cat*) located downstream of the transcriptional stop, which confers chloramphenicol-resistance in bacteria. Expression of chloramphenicol acetyltransferase is driven by the prokaryotic *trp* promoter generally responsible for tryptophan biosynthesis. In addition, the plasmid contains a gene encoding for β-lactamase (*amp*) which confers ampicillin-resistance in bacteria, located upstream of the RSS region. Additional features of pGG51 substrate plasmid include simian virus 40 (SV40) eukaryotic replication origin and prokaryotic Col E1 replication origin that allow the plasmid to replicate in both mammalian and bacterial cells.

As RAG-mediated recombination occurs, RSS region along with the transcriptional stop signal are excised out and form a signal joint. The recombined substrate plasmid forms a coding joint, thus allowing for expression of chloramphenicol-resistance gene. The newly recombined substrate plasmid confers both ampicillin and chloramphenicol resistance in bacteria. Its coding joint sequence may then be analyzed.

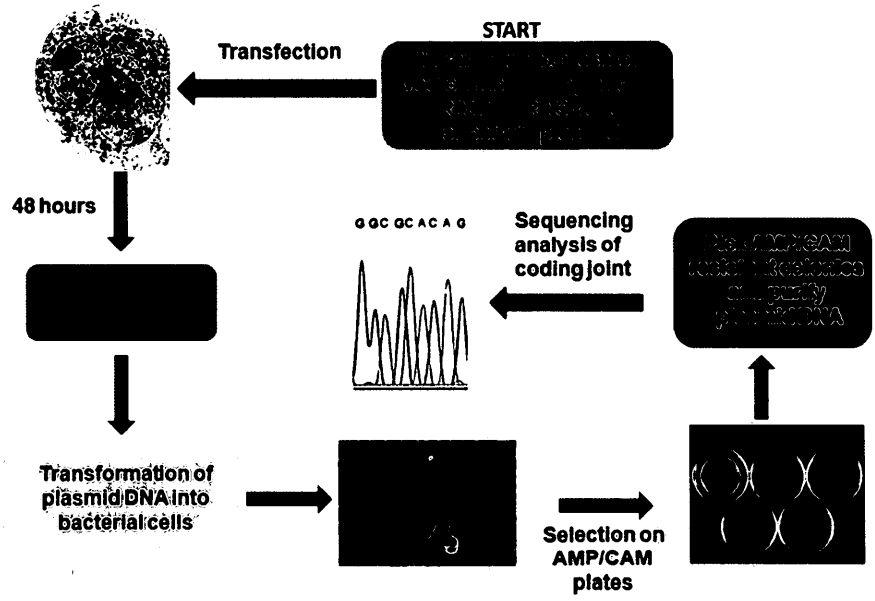


Figure 2.4: Overview of the *in vivo* extrachromosomal recombination assay. Human mammalian cells are first transfected with three mammalian expression vectors (hTdTS, hRAG1 and hRAG2) along with substrate plasmid (pGG51). After a 48 hour incubation, the DNA is harvested and transformed into bacterial cells for later selection on ampicillin and chloramphenicol containing bacterial plates for recombined substrate plasmid. In the last stage, ampicillin and chloramphenicol resistant colonies are picked and the recombined substrate plasmid is purified. The coding joint of the recombined substrate plasmid is analyzed through sequencing.

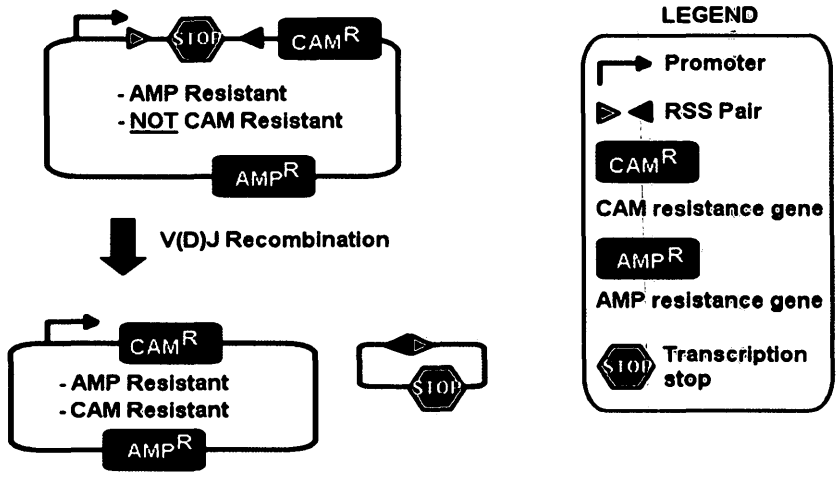


Figure 2.5: General design of the recombination substrate plasmid for *in vivo* assay. The unrecombined plasmid shown in top left contains a transcriptional terminator sequence flanked by a pair of recombination signal sequences (RSS). If the substrate does not undergo recombination, transcription of chloramphenicol-resistant gene is prevented by the transcriptional terminator. However, if recombination occurs, the RSS region along with the transcriptional terminator is excised out and a coding joint forms in its place. The transcription of the chloramphenicol-resistance gene is allowed to proceed. Thus, the recombined substrate plasmid, shown in bottom left, confers both ampicillin and chloramphenicol resistance.

Figure 2.6: Substrate recombination plasmid pGG51 sequence features. pGG51 is a recombination substrate plasmid (6284 bp). It contains a *trp* promoter region consisting of -35 and -10 box (highlighted in green). It also contains a 12-RSS region consisting of a heptamer (7 bp, highlighted in yellow), followed by a 12 bp spacer and a nanomer (9 bp, highlighted in yellow). A 23-RSS region also consist of a nanomer (highlighted in yellow), followed by a 23 bp spacer and a heptamer (highlighted in yellow). Prokaryotic transcriptional stop sequence (highlighted in red) is found between the RSS pair region. Chloramphenicol-resistance gene (highlighted in pink) is downstream of the 23-RSS region and its expression is driven by the *trp* promoter. pGG51 forward and reverse primers (undersigned in blue) are used to identify recombined plasmid.

2.5.2 Reverse transcriptase PCR

Reverse transcriptase PCR (RT-PCR) was performed to test for endogenous TdT production in HEK293T cells. The cells were passaged for at least a week prior transfection in DMEM/high glucose media (Sigma Aldrich's Dulbecco's Modified Eagle's Medium - high glucose with 4500 mg/l glucose, L-glutamine, and sodium bicarbonate) supplemented with 10% fetal bovine serum (FBS) at 37°C and 5% CO_2. Cells were seeded for transfection in 60mm tissue culture dishes at ~80-90% confluency in a total volume of 2.5ml culture media. One hour prior to transfection, cells were washed with warm 1xPBS (137mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 2mM KH_2PO_4) and fresh culture media was added for a total volume of 2.5ml per 60mm dish. HEK293T cells were transfected with either substrate recombination plasmid pGG51, empty pcDNA^{TM}4/myc-His A vector, wild type hTdTS pcDNA^{TM}4/myc-His A expression vector or co-transfected with pGG51 plasmid, wild type hTdTS pcDNA, pEBG-hRAG1 and hRAG2 expression vectors.

Per transfection, 2.5µg DNA was gently resuspended in 100µl serum-free culture media in an eppendorf tube. In a different eppendorf tube, 7.5µl Poly-Jet^{TM} DNA transfection reagent (FroggaBio #SL100688.1) was resuspended in 100µl serum-free culture media. Diluted Poly-Jet^{TM} reagent was immediately added to diluted DNA and pipetted up and down few times to mix. The transfection mix was incubated for 15 minutes at room temperature to allow Poly-Jet^{TM}/DNA complexes to form. The transfection mix was then added drop-wise to the media in the culture dish and gently swirled to mix. The transfected cells were incubated at 37°C and 5% CO_2 for 6 hours. Six

hours post-transfection, cells were washed with warm 1xPBS and fresh culture media was added for a total volume of 2.5ml per 60mm dish.

HEK293T cells were then harvested 48 hours post-transfection and total RNA was extracted using TRIzol® Reagent (Invitrogen by Life Technologies #15596-026). Appropriate precautions were taken to avoid RNase contamination. As control, RNA from non-transfected HEK293T cells was also harvested. First, culture media was removed from the culture dish and 1ml of TRIzol® Reagent was added to each 60mm dish. The cells were lysed through pipetting up and down few times to homogenize the samples and transferred to a clean tube. The homogenized samples were then incubated for 5 minutes at room temperature. To each sample 0.2ml of chloroform was added followed by vigorous hand shaking for 15 seconds. The samples were incubated for 2-3 minutes at room temperature and centrifuged at 12,000xg for 15 minutes at 4°C. The aqueous upper layer of the mixture was removed after the spin and placed in a clean tube. Total RNA was precipitated after the addition of 0.5ml 100% isopropanol to the aqueous layer, followed by incubation for 10 minutes at room temperature. The samples were then centrifuged at 12,000xg for 10 minutes at 4°C. The supernatant was removed and the RNA pellet was washed with 1ml of 75% ethanol (in DEPC-treated water), vortexed briefly and centrifuged at 7,500xg for 5 minutes at 4°C. The supernatant was discarded and the centrifugation step was repeated to remove any residual ethanol. The RNA pellet was then air dried for 5-10 minutes and resuspended in 20µl of RNase-free water. The resuspended RNA was then incubated on a heat block set at 55-60°C for 10-15 minutes. The recovered RNA sample was immediately used in a downstream RT-PCR protocol.

Reverse transcription reaction was set up using ThermoScript™ RT-PCR system (Invitrogen by Life Technologies #11146-024). First cDNA synthesis reaction was set up in a PCR tube containing 150ng of RNA sample, 50μM oligo $(dT)_{20}$ primer, 1.5mM dNTP mix and DEPC-treated water for a total volume of 12μl. The reaction was incubated at 65°C for 5 minutes in a thermocycler to denature the sample and then placed on ice. Next, cDNA synthesis mix was added to the reaction (4μl of 5xcDNA synthesis buffer, 1μl of 0.1M DTT, 40 units of RNaseOUT™, 15 units of ThermoScript™ RT, DEPC-treated water up for a total volume of 8μl). The 20μl cDNA reaction was incubated in a thermocycler for 50 minutes at 50°C, followed by reaction termination for 5 minutes at 85°C.

PCR amplification reaction of target cDNA was set up using hTdTS gene-specific primers that were designed to span an intron-exon boundry (DNTT_F and DNTT_R 5′CACCAGCTTGTTGTGAGAAGAGAC, 5′CTCTCTCAAACTGCCGGGAGCCAGT, respectively). A single PCR reaction included 2μl cDNA, 1xreaction buffer (10mM Tris-HCl (pH 9.0), 50 mM KCl, 1% Triton X-100, 1.5mM $MgCl_2$), 50pmol DNTT_F and DNTT_R primers, 0.2mM dNTP mix, 2.5 units EconoTaq® DNA Polymerase (Lucigen #30031-1) and dH_2O for a total reaction volume of 50μl. Negative PCR control reaction did not include cDNA template, while the positive PCR control reaction included 0.25μg wild type hTdTS pcDNA expression vector as template DNA. The PCR reactions were cycled according to parameters in Table 2.6. Following PCR amplification, the reaction products were ran on 0.7% agarose gel stained with ethidium bromide and visualized under UV light.

Table 2.6: PCR program parameters for amplification of hTdTS cDNA following cDNA synthesis via Reverse transcriptase.

Step		Temperature	Duration
Denaturation		95°C	2 minutes
30 cycles	Denaturation	95°C	45 seconds
	Annealing	75°C	1 minute
	Extension	70°C	3 minutes
Final extension		70°C	15 minutes

2.5.3 Western blotting procedures

Endogenous hTdT expression was also tested at the protein level by Western blot using rabbit polyclonal TdT specific antibody (Novus Biologicals #NBP1-58254) raised against N-terminal peptide sequence of TdT gene (FQDLVVFILEKKMGTTRRAFLMEL ARRKGFRVENELSDSVTHIVAENNSG). HEK293T cells were transfected with either wild type hTdTS pcDNA^{TM}4/myc-His A expression vector or co-transfected with substrate recombination pGG51 plasmid, wild type hTdTS pcDNA, pEBG-hRAG1 and hRAG2 expression vectors. The transfections were performed with Poly-Jet^{TM} DNA transfection reagent in 60mm culture dishes as described in section 2.5.2. Note, culture media was supplemented with caffeine for a final concentration of 1mM 6 hours post-transfection.

HEK293T cells were harvested 48 hours post-transfection. As experimental control, non-transfected HEK293T cells were also harvested. First, cells were washed with warm 1xPBS. Cells were then resuspended in 2.5ml 1xPBS and transferred to a clean 15ml falcon tube. The cells were spun at 500xg for 5 minutes at room temperature and the supernatant was removed. The cell pellet was resuspended in 75µl RIPA Lysis buffer (150mM NaCl, 50mM Tris pH 8.0, 0.5% NP-40, 1x protease inhibitor) and

transferred to a clean eppendorf tube. The samples were incubated for 20 minutes at 4°C on a rotator. The cells were then further lysed by sonication (0.5 second pulse on, 2 seconds pulse off, at 10% amplitude for a total pulse time of 3 seconds). The lysed samples were centrifuged at 22,500xg for 5 minutes at 4°C. The supernatant was transferred to a clean eppendorf tube and total protein concentration was measured using Bradford reagent. A total of 85μg of protein per sample was measured and boiled for 5 minutes prior to loading onto freshly prepared 12% SDS-PAGE gel.

The protein samples were loaded onto 12% SDS-PAGE gel along with pre-stained protein marker (Thermo Scientific PageRuler #26616). The gel was ran for 150V until the dye front reached the bottom of the gel. The proteins were then set up for a transfer onto a nitrocellulose membrane that was activated in 100% methanol. The transfer was conducted overnight at 30V at 4°C. After the transfer, the membrane was blocked in 1xPBST (137mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 2mM KH_2PO_4, 0.1% Tween-20 pH 7.4) containing 5% milk (w/v) for 2 hours, on a rocker, at room temperate. The membrane was then cut in half and the top portion was probed with 1:1,000 dilution of rabbit polyclonal TdT specific antibody (Novus Biologicals #NBP1-58254) in milk, while the bottom portion was probed with 1:5,000 dilution of mouse monoclonal IgG GAPDH specific antibody (Santa Cruz Biotech #SC-47724) in milk. The membrane was incubated in primary antibody for 1 hour, on a rocker, at room temperature. The membrane was then washed for 5 minutes with 1xPBST. The wash was repeated twice more. The top portion of the membrane was probed with 1:10,000 dilution of anti-rabbit horseradish peroxidase (HRP)- conjugated antibody (Promega #W4011) in milk. The

bottom portion of the membrane was probed with 1:10,000 dilution of anti-mouse IgG HRP-conjugated antibody (Promega #W4021) in milk. The secondary antibody incubation proceeded for 1 hour, on a rocker, at room temperature. The membrane was then washed for 20 minutes with 1xPBST. The wash was repeated twice more. The membrane was then washed with Clarity™ Western ECL substrate (Biorad #170-5060) and transferred into a western blot cassette to be kept in the dark. The membrane was then developed in the dark by exposing western blotting film directly onto a membrane, followed by film development.

2.5.4 Mammalian cell transfection for the V(D)J recombination assay

HEK293T cells were transfected with Poly-Jet™ DNA transfection reagent in 60mm culture dishes as described in section 2.5.2. Each transfection was composed of 0.62μg substrate recombination pGG51 plasmid, 0.62μg pEBG-hRAG1 expression vector, 0.62μg pEBG-hRAG2 expression vector and 0.62μg wild type hTdTS (or other hTdTS variant) pcDNA™4/myc-His A expression vector for a total of 2.5μg DNA per transfection. Note, culture media was supplemented with caffeine for a final concentration of 1mM 6 hours post-transfection. A total of 8 independent HEK293T cell transfection sets were performed. Culture media supplemented with 1mM caffeine post transfection was shown to increase recombination frequency, at least five fold, by increasing expression levels of RAG proteins (72, 73).

2.5.5 Harvesting plasmid DNA via alkaline lysis

Substrate plasmid DNA was harvested 48 hours post-transfection via alkaline lysis methods. The cells were washed with warm 1xPBS, resuspended in 2.5ml 1xPBS and transferred to a clean 15ml falcon tube. The cells were spun at 500xg for 5 minutes at room temperature and the supernatant was removed. The pellet was then washed with cold 1xPBS and cells were centrifuged at 500xg for 5 minutes at room temperature. The supernatant was removed and pellet was gently resuspended in 400µl of Lysis buffer (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA in dH_2O) and transferred into a clean tube. Freshly prepared 800µl of alkaline solution (0.2 M NaOH, 0.6% SDS in dH_2O) was added to the cell lysate. The solution was gently inverted 6 times and left on ice for 4 minutes. To the solution, 600µl of 7.5M ammonium acetate was added and the tube were gently inverted to mix and left on ice for 10 minutes to precipitate protein and chromosomal DNA. Next, the solution was centrifuged at 20,000xg for 15 minutes at 4°C. The supernatant was transferred to a tube containing 100% ethanol at about twice the volume of supernatant and incubated at -80°C for 45 minutes to precipitate plasmid DNA. Next, the solution was centrifuged at 20,000xg for 25 minutes at 4°C. The precipitated DNA pellet was washed twice with 1ml of ice cold 70% ethanol and centrifuged at 20,000xg for 5 minutes at 4°C. The supernatant was removed and the pellet was air dried for 10 minutes and resuspended in 44µl of water.

2.5.6 DpnI digest of harvested plasmid DNA

DpnI digest was carried out in order to remove any unreplicated DNA plasmid from the harvested DNA samples. DpnI restriction enzyme cleaves methylated adenines in 5′GATC base pair sequence, thus digesting unreplicated DNA plasmid in transfected HEK293T cells. To each of the resuspended plasmid DNA samples, 20 units of DpnI restriction enzyme (New England BioLabs #R0176S) and 1xNEBuffer 4 (20mM Tris-acetate, 50mM potassium acetate, 10mM magnesium acetate, and 1mM DTT, pH 7.9 at 25°C) were added for a total volume of 50µl. The restriction digest reaction was incubated at 37°C for two hours. To each sample 200µl of 100% ethanol and 40µl of 7.5M ammonium acetate were added. The samples were incubated at -80°C for 30 minutes followed by centrifugation at 20,000xg for 15 minutes at 4°C. The DNA pellet was washed with 500µl of 70% ethanol and centrifuged at 20,000xg for 5 minutes at 4°C. The wash was repeated twice more and the DNA pellet was allowed to air dry for 10 minutes. The plasmid DNA was resuspended in 6µl of dH_2O and used in subsequent bacterial transformation.

2.5.7 Transformation of harvested plasmid DNA via electroporation

The harvested pGG51 substrate plasmid from each transfection was transformed into *E.coli* ElectroMAX™ DH10B™ Cells (Invitrogen #18290-015). Competent cells were thawed on ice and 25µl of cells were aliquot to a pre-chilled eppendorf tube and mixed with 2.5µl of freshly harvested plasmid DNA for each transformation reaction. The mixture was transferred into a dry pre-chilled 0.1cm BioRad electroporation cuvette and pulsed at 1700V using Gene Pulser electoporator. Immediately, 1ml of pre-warmed

SOC medium (20g/l tryptone, 5g/l yeast extract, 0.5g/l NaCl, 2.5mM KCl, 10mM $MgCl_2$, and 20mM glucose) was added to the cells and mixed gently. The mixture from was transferred to a clean 15ml falcon tube and incubated 1 hour at 37°C, shaking at 200rpm. After incubation, each of the electroporation samples was used to plate onto pre-warmed LB agar plates containing either 100µg/ml ampicillin (AMP plates) or 33µg/ml chloramphenicol and 100µg/ml ampicillin (CAM/AMP plates). Per each transformation, 300µl of transformation reaction was plated per every CAM/AMP for a total of 3 plates. In addition, 100µl of 1:100 and 1:500 dilution of transformation reaction were plated on AMP plates. All plates were incubated at 37°C overnight or for one full day, as needed, to allow colony growth. The number of colonies on each plate was counted and recorded.

2.5.8 Colony PCR

Bacterial colonies from CAM/AMP plates were picked and colony PCR was performed in order to select for true recombined substrate pGG51 plasmids. A single colony was picked and resuspended in 20µl dH_2O. A PCR reaction was set up using pGG51 forward and reverse primers (shown in Figure 2.6) were used to identify recombined plasmid. Recombined substrate plasmid was expected to produce PCR product of about 100 base pairs, while the non-recombined plasmid produced PCR product of about 400 base pairs. A single PCR reaction included 2µl bacterial colony DNA template, 10µl PCR master mix (Thermo Scientific #K0171), 0.25µg of each of pGG51 forward and reverse primers and dH_2O for a total reaction volume of 20µl. The PCR reaction was cycled as per program in Table 2.7. The PCR products were ran on 2% agarose gel, stained with ethidium bromide and visualized under UV light. Only the

recombined bacterial colonies were inoculated overnight in 5ml LB containing 100μg/ml ampicillin. The recombined pGG51 substrate plasmid was purified from the overnight culture via QIAprep® Miniprep kit (Qiagen #27106).

Table 2.7: PCR program parameters for colony PCR of recombination pGG51 substrate plasmid.

Step		**Temperature**	**Duration**
Denaturation		95°C	5 minutes
30 cycles	Denaturation	95°C	45 seconds
	Annealing	53°C	1 minute
	Extension	72°C	2 minutes
Final extension		72°C	15 minutes

2.5.9 Recombination data analysis

The analysis of V(D)J recombination assay included calculating recombination frequency (also known as R-value) and sequencing recombined substrate plasmid to analyze the arbitrary complementary-determining region (CDR). The R-value represents the ratio of the total number of double-resistant (ampicillin and chloramphenicol resistance) colonies post transformation and the total number of ampicillin resistant colonies. The ampicillin resistant colonies represent the total number of bacterial cells that have been transformed with the substrate pGG51 plasmid, while the double-resistant colonies represent the total number of substrate plasmids that undergone RAG-mediated recombination. The equation for calculation of R-value is shown below.

$$\text{R-value (\%)} = \frac{\text{The total number of AMP + CAM resistant colonies in a specific volume of undiluted transformation sample}}{\text{The total number of AMP resistant colonies in the same volume of undiluted transformation sample}} \times 100\%$$

The harvested recombined pGG51 substrate plasmid was sequenced using pGG51 forward primer (5′ TTGTCGATGAATTCCCCTGT) (Shown in Figure 2.6) in order to examine the recombined coding joint. Analysis was performed on the following parameters: number of nucleotide deletions and additions (palindromic and random nucleotide additions), frequency of nucleotide additions, AT versus GC base content, the distribution of number of N-nucleotide additions per recombined joint and the length of arbitrary complementary determining region (CDR). The analysis was performed to determine if there was a statistical difference between different populations of wild-type and hTdTS variant generated joints. The difference between two groups was considered statistically significant if the p-value obtained was less than 0.05. The analysis was performed using SPSS statistical analysis and data management system. In this study, repeated recombined pGG51 substrate sequences within the same mammalian transfection were excluded from the analysis due to the possibility of plasmid replication in mammalian cells after recombination.

Chapter 3: Results

The objective of my research is to determine whether naturally occurring variants of a human polymerase, TdT, have differences in their activity, and hence function. TdT is a non-template dependent polymerase that is needed to generate the high level of diversity found in the receptors that recognize pathogens, specifically, the diversity generated in the CDR3 regions during the process of V(D)J recombination of the B cell and T cell receptor genes. I investigated the functions of the identified variant hTdTS using two different approaches, namely, biochemical kinetic analyses and *in vivo* functional V(D)J recombination assays. This chapter details my characterization of potential functional variants, including their cloning and expression, their function in biochemical assays, and finally their *in vivo* functions in generating diversity in the Immune System.

Section 3.1 hTdT SNPs

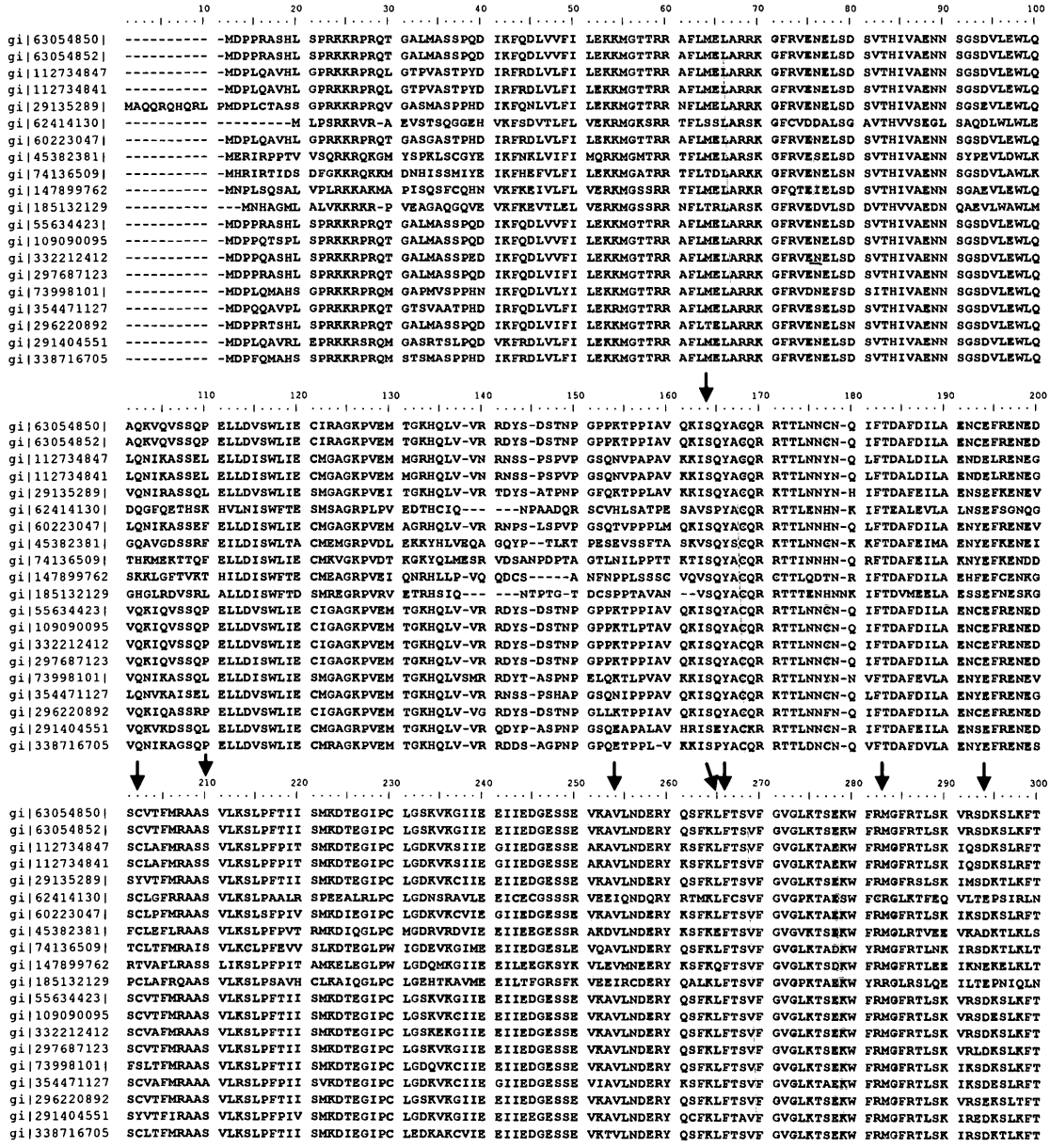
The NCBI SNP database was used to map a total of 62 SNPs of hTdTS isoform (1530 base pairs long, 509 amino acid residues) (Appendix A). Among these, 24 SNPs were not considered in the study since they all localized in the BRCA1-C terminus of the protein not associated with TdT polymerase activity (Appendix A, first 24 SNPs). Another 12 synonymous SNPs were discarded since they did not result in an amino acid change. Thus, a total of 26 missense SNPs were screened based on a number of criteria to identify the top hTdTS variants I would use for activity studies. First, the TdT protein sequence homology across species was considered. A total of 20 TdT protein sequences from 18 different species including mouse, zebrafish and rat were multi-aligned using

BioEdit software (Figure 3.1). The residues that were highly conserved at the SNP positions across all species were hypothesized to be significant for TdT structure and possibly its function. Secondly, the chemical nature of the amino acid change was considered.

SNPs resulting in significant changes in amino acid chemical property, such as negatively to positively charged amino acid residue were hypothesized to have a greater effect on the protein and its function as compared to less significant changes, such as substituting one polar residue for a different polar residue. The third screening process included modeling of the SNP residue change using the available crystal structure of murine TdT via *PyMol* Mutagenesis wizard software (note, the human TdT structure has not been determined to date). This provides structural information of the SNP residue and the proximity of the mutation to the polymerase active site, composed of three highly conserved aspartic acid residues, D343, D345 and D434, and two cation metal binding sites (Figure 3.2).

A total of six hTdTS SNPs were identified for further characterization of their effects on the polymerase activity (Table 3.1). R431C, R460Q and T450S SNPs are localized in close proximity to the active site of the polymerase (Figure 3.2, Panel C). T450S SNP is part of the polymerase dNTP-binding site. In addition, A445T SNP is within 8Å of the incoming ddATP substrate, is also located within the same dNTP-binding site (Figure 3.2, Panel C). L397S SNP is positioned within the loop region shown to associate with template independent polymerase activity (Figure 3.2, Panel B). D280H SNP was chosen as an experimental control due to its location at the periphery of the

protein, away from the active site, and thus was hypothesized not to affect TdT activity (Figure 3.2, Panel B). Next, *in vitro* and *in vivo* functional activity assays were carried out to compare activity of these SNPs to the wild-type TdT protein.



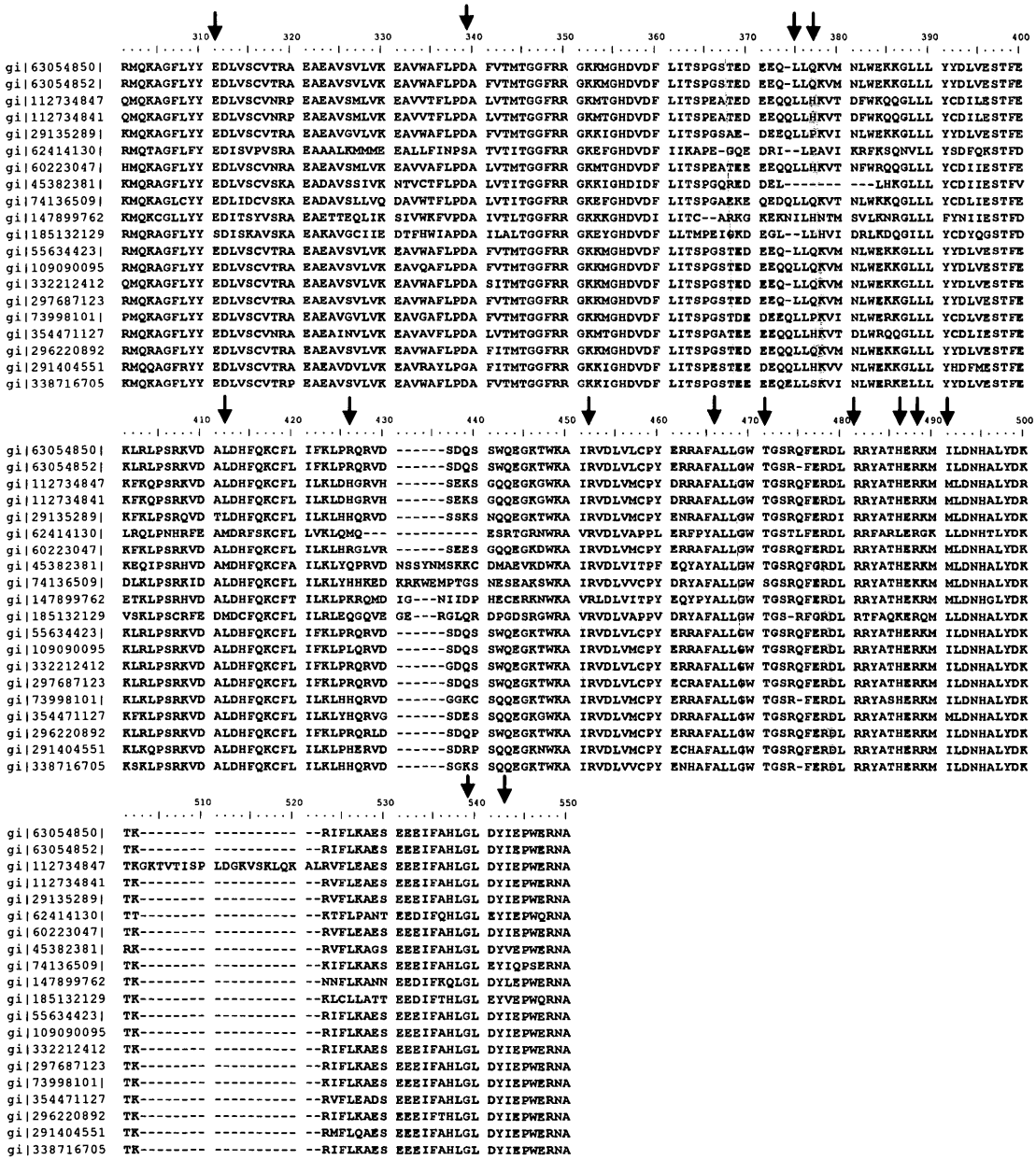


Figure 3.1: Multi-sequence alignment of TdT amino acid sequences from 18 different species available at NCBI database using *BioEdit* software. Black errors represent the location of amino acid change in hTdT isoform 1 caused by the identified SNPs. Red errors represent the top six candidates picked for the study.

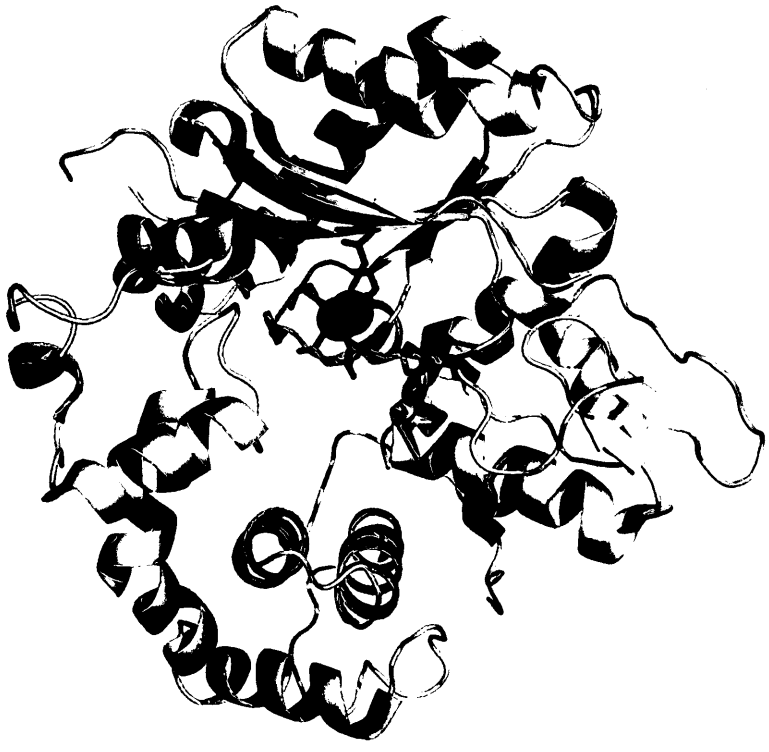
The species used in the multi-alignment, in order, are: gi|63054850| TdT isoform 1 [Homo sapiens], gi|63054852| TdT isoform 2 [Homo sapiens], gi|112734847| TdT isoform 1 [Mus musculus], gi|112734841| TdT isoform 2 [Mus musculus], gi|29135289| TdT [Bos taurus], gi|62414130| TdT [Danio rerio], gi|60223047| TdT [Rattus norvegicus], gi|45382381| TdT [Gallus gallus], gi|74136509| TdT [Monodelphis domestica], gi|147899762| TdT [Xenopus laevis], gi|185132129| TdT [Oncorhynchus mykiss], gi|55634423| TdT isoform 3 [Pan troglodytes], gi|109090095| TdT isoform 2 [Macaca mulatta], gi|332212412| TdT [Nomascus leucogenys], gi|297687123| TdT [Pongo abelii], gi|73998101| TdT isoform 2 [Canis lupus familiaris], gi|354471127| TdT [Cricetulus griseus], gi|296220892| TdT [Callithrix jacchus], gi|291404551| TdT [Oryctolagus cuniculus], gi|338716705| TdT [Equus caballus].

Table 3.1: Human TdT variant 1 SNP candidates for functional activity assays, characterized based on mRNA position, base change, amino acid position and change, the chemical nature of the amino acid change, amino acid homology across species and their respective significance.

SNP ID	mRNA position	Base change	Amino acid position	Amino acid change	Chemical nature of amino acid change	Amino acid homology across species*	Significance
rs41291616	1008	G to C	280	Asp to His	Negatively to positively charged	100%	Serves as control; localized at periphery of protein
rs142078021	1360	T to C	397	Leu to Ser	Non-polar to polar	100%	Localized close to loop region associated with template-dependent TdT activity
rs139202533	1461	C to T	431	Arg to Cys	Positively charged to polar	100%	Close proximity to active site
rs149947558	1503	G to A	445	Ala to Thr	Non-polar to polar	100%	Localized within incoming dNTP-binding site and in close proximity to active site
rs145011041	1518	A to T	450	Thr to Ser	No change; polar to polar	95%	
rs142141382	1549	G to A	460	Arg to Gln	Positively charged to polar	100%	Close proximity to active site

*Amino acid homology across species is based on TdT amino acid sequence multi-alignment (Figure 3.1), shown as percentage of amino acid identity across 20 TdT protein sequences analyzed.
**The genotype frequencies of above SNPs do not exceed 2% of assayed cohort populations (according to the NCBI).

A



B

C

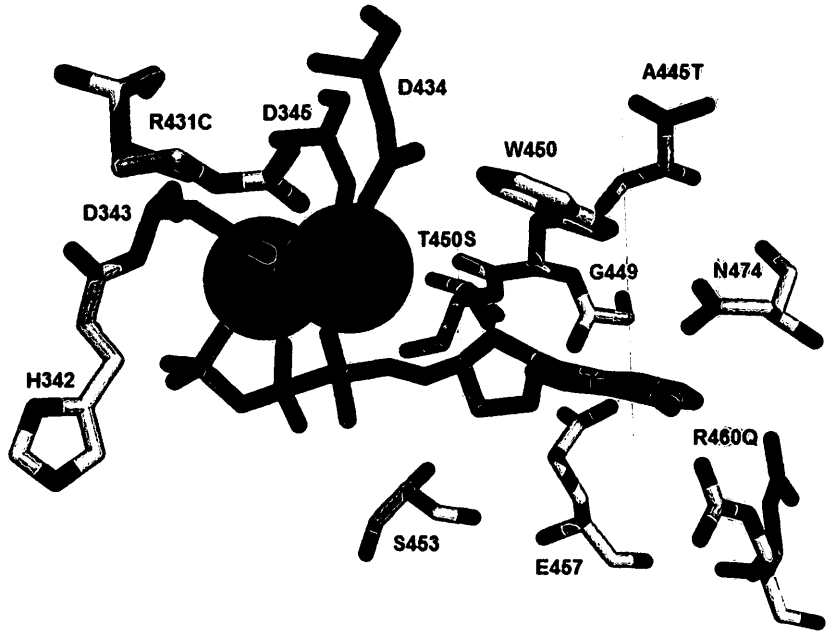


Figure 3.2: Crystal structure of murine TdT (PDB ID: 1KEJ) and the relative positions of hTdTS SNPs. Figures generated using *PyMol* software.

- **A.** TdT secondary structure elements (alpha helices and beta strands) are shown. Three conserved aspartic acid residues are highlighted in green (D343, D435 and D434). The incoming nucleotide substrate, ddATP, is shown in orange. Two cobalt ions are colored cyan.
- **B.** TdT structure showing the relative positions of D280H and L397S. The positions of SNPs are superimposed on the wild type amino acid residues, shown in magenta.
- **C.** Active site of TdT showing the relative positions of R431C, T450S, A445T and R460Q SNPs within 8Å of incoming nucleotide substrate, ddATP. The positions of SNPs are superimposed on the wild type amino acid residues, shown in magenta.

Section 3.2 Plasmid hTdTS DNA construction

The chosen six hTdTS SNPs candidates for the study were generated via site-directed mutagenesis hTdTS cDNA in bacterial p15TV-L expression vector. Figure 3.3 indicates the base change and position of the point mutations. Figure 3.4 shows the corresponding protein multi-sequence alignment of all six hTdTS variants. The generated p15TV-L hTdTS SNPs expression vectors were later used in hTdTS protein purification.

Mammalian hTdTS pcDNA^{TM}4/myc-His A expression vectors carrying specific point mutations were generated via restriction site infusion based cloning. Specifically, HindIII and XhoI restriction cut sites were used to insert hTdTS PCR product generated from hTdTS cDNA p15TV-L vector into multiple cloning site of pcDNA^{TM}4/myc-His A vector. Restriction digest DNA products shown in Figure 3.5 were used in mammalian expression vector construction. Bacterial p15TV-L vector contains a single HindIII and XhoI restriction cut site, thus upon digestion a single predominant band of about 7Kb is produced (note; the smaller size bands of the digestion represent, I believe, non-specific cutting by the restriction enzymes). Mammalian pcDNA vector also contains a single HindIII and XhoI cut site at its multiple cloning region separated by 74 base pairs from one another (Figure 2.2). Thus, upon HindIII/XhoI double digest a single predominant band is seen around 5Kb.

The final successfully cloned wild-type, L397S, R431C and A445T hTdTS cDNA pcDNA expression vector were verified through sequencing (Figure 3.6) and used in subsequent functional activity assays.

```
1       ACACTTTGGC AGGAAGCTGT TGCCAGGGCA GCACCTGTGA AGCCCTGGCC
51      TGGCTTCAGA GTCTGCTGGT GAGATGACAT CAAAACCCTT CGTGTAGGAG
101     GGTGGCAGTC TCCCTCCCTT CTGGAGACAC CACCAGATGG GCCAGCCAGA
151     GGCAGCAGCA GCCTCTTCCC ATGGATCCAC CACGAGCGTC CCACTTGAGC
201     CCTCGGAAGA AGAGACCCCG GCAGACGGGT GCCTTGATGG CCTCCTCTCC
251     TCAAGACATC AAATTTCAAG ATTTGGTCGT CTTCATTTTG GAGAAGAAAA
301     TGGGAACCAC CCGCAGAGCG TTCCTCATGG AGCTGGCCCG CAGGAAAGGG
351     TTCAGGGTTG AAAATGAGCT CAGTGATTCT GTCACCCACA TCGTAGCAGA
401     GAACAACTCG GGTTCGGATG TTCTGGAGTG GCTTCAAGCA CAGAAAGTAC
451     AAGTCAGCTC ACAACCAGAG CTCCTCGATG TCTCCTGGCT GATCGAATGC
501     ATAAGAGCAG GGAAACCGGT GGAAATGACA GGAAAACACC AGCTTGTTGT
551     GAGAAGAGAC TATTCAGATA GCACCAACCC AGGCCCCCCG AAGACTCCAC
601     CAATTGCTGT ACAAAAGATC TCCCAGTATG CGTGTCAGAG AAGAACCACT
651     TTAAACAACT GTAACCAGAT ATTCACGGAT GCCTTTGATA TACTGGCTGA
701     AAACTGTGAG TTTAGAGAAA ATGAAGACTC CTGTGTGACA TTTATGAGAG
751     CAGCTTCTGT ATTGAAATCT CTGCCATTCA CAATCATCAG TATGAAGGAC
801     ACAGAAGGAA TTCCCTGCCT GGGGTCCAAG GTGAAGGGTA TCATAGAGGA
851     GATTATTGAA GATGGAGAAA GTTCTGAAGT TAAAGCTGTG TTAAATGATG
901     AACGATATCA ATCCTTCAAA CTCTTTACTT CTGTATTTGG AGTGGGGCTG
951     AAGACTTCTG AGAAGTGGTT CAGGATGGGT TTCAGAACTC TGAGTAAAGT
1001    AAGGTCGGAC AAAAGCCTGA AATTTACACG AATGCAGAAA GCAGGATTTC
1051    TGTATTATGA AGACCTTGTC AGCTGTGTGA CCAGGGCAGA AGCAGAGGCC
1101    GTCAGTGTGC TGGTTAAAGA GGCTGTCTGG GCATTTCTTC CGGATGCTTT
1151    CGTCACCATG ACAGGAGGGT TCCGGAGGGG TAAGAAGATG GGGCATGATG
1201    TAGATTTTTT AATTACCAGC CCAGGATCAA CAGAGGATGA AGAGCAACTT
1251    TTACAGAAAG TGATGAACTT ATGGGAAAAG AAGGGATTAC TTTTATATTA
1301    TGACCTTGTG GAGTCAACAT TTGAAAAGCT CAGGTTGCCT AGCAGGAAGG
1351    TTGATGCTTT GGATCATTTT CAAAAGTGCT TTCTGATTTT CAAATTGCCT
1401    CGTCAAAGAG TGGACAGTGA CCAGTCCAGC TGGCAGGAAG GAAAGACCTG
1451    GAAGGCCATC CGTGTGGATT TAGTTCTGTG CCCCTACGAG CGTCGTGCCT
1501    TTGCCCTGTT GGGATGGACT GGCTCCCGGC AGTTTGAGAG AGACCTCCGG
1551    CGCTATGCCA CACATGAGCG GAAGATGATT CTGGATAACC ATGCTTTATA
1601    TGACAAGACC AAGAGGATAT TCCTCAAAGC AGAAAGTGAA GAAGAAATTT
1651    TTGCGCATCT GGGATTGGAT TATATTGAAC CGTGGGAAAG AAATGCCTAG
1701    GAAAGTGTTG TCAACATTTT TTTCCTATTC TTTTCAAGTT AAATAAATTA
1751    TGCTTCATAT TAGTAAAAGA TGCCATAGGA GAGTTTGGGG TTATTTAGGT
1801    CTTATTGAAA TGCAGATTGC TACTAGAAAT AAATAACTTT GGAAACATGG
1851    GAAGGTGCCA CTGGTAATGG GTAAGGTTCT AATAGGCCAT GTTTATGACT
1901    GTTGCATAGA ATTCACAATG CATTTTTCAA GAGAAATGAT GTTGTCACTG
1951    GTGGCTCATT CAGGGAAGCT CATCAAAGCC CACTTTGTTC GCAGTGTAGC
2001    TGAAATACTG TCTATCTCTA ATAAAAACAG GAGGAAACAA AAAAAAAAAA
2051    AAAAAAAAAA AAAAAAAAAA A
```

Figure 3.3: Homo sapiens deoxynucleotidyl transferase, terminal (DNTT), transcript variant 1, mRNA. NCBI Reference Sequence: NM_004088.3, GI:63054849. Adapted from the NCBI database.

hTdTS mRNA sequence is 2071 base pairs long. The coding region of hTdT is between 171-1700 base pairs. The start (ATG) and stop (TAG) codons are shown in bolded font. The chosen hTdTS SNPs positions are shown in bolded font;

- D280H; G to C point mutation at position 1008
- L397S; T to C point mutation at mRNA position 1360
- R431C; C to T point mutation at mRNA position 1461
- A445T; G to A point mutation at mRNA position 1503
- T450S; A to T point mutation at mRNA position 1518
- R460Q; G to A point mutation at mRNA position 1549

```
WT          MDPPRASHLSPRKKRPRQTGALMASSPQDIKFQDLVVFILEKKMGTTRRAFLMELARRKG 60
D280H       MDPPRASHLSPRKKRPRQTGALMASSPQDIKFQDLVVFILEKKMGTTRRAFLMELARRKG 60
L397S       MDPPRASHLSPRKKRPRQTGALMASSPQDIKFQDLVVFILEKKMGTTRRAFLMELARRKG 60
R431C       MDPPRASHLSPRKKRPRQTGALMASSPQDIKFQDLVVFILEKKMGTTRRAFLMELARRKG 60
A445T       MDPPRASHLSPRKKRPRQTGALMASSPQDIKFQDLVVFILEKKMGTTRRAFLMELARRKG 60
T450S       MDPPRASHLSPRKKRPRQTGALMASSPQDIKFQDLVVFILEKKMGTTRRAFLMELARRKG 60
R460Q       MDPPRASHLSPRKKRPRQTGALMASSPQDIKFQDLVVFILEKKMGTTRRAFLMELARRKG 60
            ************************************************************

WT          FRVENELSDSVTHIVAENNSGSDVLEWLQAQKVQVSSQPELLDVSWLIECIRAGKPVEMT 120
D280H       FRVENELSDSVTHIVAENNSGSDVLEWLQAQKVQVSSQPELLDVSWLIECIRAGKPVEMT 120
L397S       FRVENELSDSVTHIVAENNSGSDVLEWLQAQKVQVSSQPELLDVSWLIECIRAGKPVEMT 120
R431C       FRVENELSDSVTHIVAENNSGSDVLEWLQAQKVQVSSQPELLDVSWLIECIRAGKPVEMT 120
A445T       FRVENELSDSVTHIVAENNSGSDVLEWLQAQKVQVSSQPELLDVSWLIECIRAGKPVEMT 120
T450S       FRVENELSDSVTHIVAENNSGSDVLEWLQAQKVQVSSQPELLDVSWLIECIRAGKPVEMT 120
R460Q       FRVENELSDSVTHIVAENNSGSDVLEWLQAQKVQVSSQPELLDVSWLIECIRAGKPVEMT 120
            ************************************************************

WT          GKHQLVVRRDYSDSTNPGPPKTPPIAVQKISQYACQRRTTLNNCNQIFTDAFDILAENCE 180
D280H       GKHQLVVRRDYSDSTNPGPPKTPPIAVQKISQYACQRRTTLNNCNQIFTDAFDILAENCE 180
L397S       GKHQLVVRRDYSDSTNPGPPKTPPIAVQKISQYACQRRTTLNNCNQIFTDAFDILAENCE 180
R431C       GKHQLVVRRDYSDSTNPGPPKTPPIAVQKISQYACQRRTTLNNCNQIFTDAFDILAENCE 180
A445T       GKHQLVVRRDYSDSTNPGPPKTPPIAVQKISQYACQRRTTLNNCNQIFTDAFDILAENCE 180
T450S       GKHQLVVRRDYSDSTNPGPPKTPPIAVQKISQYACQRRTTLNNCNQIFTDAFDILAENCE 180
R460Q       GKHQLVVRRDYSDSTNPGPPKTPPIAVQKISQYACQRRTTLNNCNQIFTDAFDILAENCE 180
            ************************************************************

WT          FRENEDSCVTFMRAASVLKSLPFTIISMKDTEGIPCLGSKVKGIIEEIIEDGESSEVKAV 240
D280H       FRENEDSCVTFMRAASVLKSLPFTIISMKDTEGIPCLGSKVKGIIEEIIEDGESSEVKAV 240
L397S       FRENEDSCVTFMRAASVLKSLPFTIISMKDTEGIPCLGSKVKGIIEEIIEDGESSEVKAV 240
R431C       FRENEDSCVTFMRAASVLKSLPFTIISMKDTEGIPCLGSKVKGIIEEIIEDGESSEVKAV 240
A445T       FRENEDSCVTFMRAASVLKSLPFTIISMKDTEGIPCLGSKVKGIIEEIIEDGESSEVKAV 240
T450S       FRENEDSCVTFMRAASVLKSLPFTIISMKDTEGIPCLGSKVKGIIEEIIEDGESSEVKAV 240
R460Q       FRENEDSCVTFMRAASVLKSLPFTIISMKDTEGIPCLGSKVKGIIEEIIEDGESSEVKAV 240
            ************************************************************

WT          LNDERYQSFKLFTSVFGVGLKTSEKWFRMGFRTLSKVRSDKSLKFTRMQKAGFLYYEDLV 300
D280H       LNDERYQSFKLFTSVFGVGLKTSEKWFRMGFRTLSKVRSHKSLKFTRMQKAGFLYYEDLV 300
L397S       LNDERYQSFKLFTSVFGVGLKTSEKWFRMGFRTLSKVRSDKSLKFTRMQKAGFLYYEDLV 300
R431C       LNDERYQSFKLFTSVFGVGLKTSEKWFRMGFRTLSKVRSDKSLKFTRMQKAGFLYYEDLV 300
A445T       LNDERYQSFKLFTSVFGVGLKTSEKWFRMGFRTLSKVRSDKSLKFTRMQKAGFLYYEDLV 300
T450S       LNDERYQSFKLFTSVFGVGLKTSEKWFRMGFRTLSKVRSDKSLKFTRMQKAGFLYYEDLV 300
R460Q       LNDERYQSFKLFTSVFGVGLKTSEKWFRMGFRTLSKVRSDKSLKFTRMQKAGFLYYEDLV 300
            ************************************.***********************

WT          SCVTRAEAEAVSVLVKEAVWAFLPDAFVTMTGGFRRGKKMGHDVDFLITSPGSTEDEEQL 360
D280H       SCVTRAEAEAVSVLVKEAVWAFLPDAFVTMTGGFRRGKKMGHDVDFLITSPGSTEDEEQL 360
L397S       SCVTRAEAEAVSVLVKEAVWAFLPDAFVTMTGGFRRGKKMGHDVDFLITSPGSTEDEEQL 360
R431C       SCVTRAEAEAVSVLVKEAVWAFLPDAFVTMTGGFRRGKKMGHDVDFLITSPGSTEDEEQL 360
A445T       SCVTRAEAEAVSVLVKEAVWAFLPDAFVTMTGGFRRGKKMGHDVDFLITSPGSTEDEEQL 360
T450S       SCVTRAEAEAVSVLVKEAVWAFLPDAFVTMTGGFRRGKKMGHDVDFLITSPGSTEDEEQL 360
R460Q       SCVTRAEAEAVSVLVKEAVWAFLPDAFVTMTGGFRRGKKMGHDVDFLITSPGSTEDEEQL 360
            ************************************************************
```

```
WT          LQKVMNLWEKKGLLLYYDLVESTFEKLRLPSRKVDALDHFQKCFLIFKLPRQRVDSDQSS 420
D280H       LQKVMNLWEKKGLLLYYDLVESTFEKLRLPSRKVDALDHFQKCFLIFKLPRQRVDSDQSS 420
L397S       LQKVMNLWEKKGLLLYYDLVESTFEKLRLPSRKVDASDHFQKCFLIFKLPRQRVDSDQSS 420
R431C       LQKVMNLWEKKGLLLYYDLVESTFEKLRLPSRKVDALDHFQKCFLIFKLPRQRVDSDQSS 420
A445T       LQKVMNLWEKKGLLLYYDLVESTFEKLRLPSRKVDALDHFQKCFLIFKLPRQRVDSDQSS 420
T450S       LQKVMNLWEKKGLLLYYDLVESTFEKLRLPSRKVDALDHFQKCFLIFKLPRQRVDSDQSS 420
R460Q       LQKVMNLWEKKGLLLYYDLVESTFEKLRLPSRKVDALDHFQKCFLIFKLPRQRVDSDQSS 420
            ********************************** ********************

WT          WQEGKTWKAIRVDLVLCPYERRAFALLGWTGSRQFERDLRRYATHERKMILDNHALYDKT 480
D280H       WQEGKTWKAIRVDLVLCPYERRAFALLGWTGSRQFERDLRRYATHERKMILDNHALYDKT 480
L397S       WQEGKTWKAIRVDLVLCPYERRAFALLGWTGSRQFERDLRRYATHERKMILDNHALYDKT 480
R431C       WQEGKTWKAICVDLVLCPYERRAFALLGWTGSRQFERDLRRYATHERKMILDNHALYDKT 480
A445T       WQEGKTWKAIRVDLVLCPYERRAFTLLGWTGSRQFERDLRRYATHERKMILDNHALYDKT 480
T450S       WQEGKTWKAIRVDLVLCPYERRAFALLGWSGSRQFERDLRRYATHERKMILDNHALYDKT 480
R460Q       WQEGKTWKAIRVDLVLCPYERRAFALLGWTGSRQFERDLQRYATHERKMILDNHALYDKT 480
            ********** ************:****:*********:********************

WT          KRIFLKAESEEEIFAHLGLDYIEPWERNA 509
D280H       KRIFLKAESEEEIFAHLGLDYIEPWERNA 509
L397S       KRIFLKAESEEEIFAHLGLDYIEPWERNA 509
R431C       KRIFLKAESEEEIFAHLGLDYIEPWERNA 509
A445T       KRIFLKAESEEEIFAHLGLDYIEPWERNA 509
T450S       KRIFLKAESEEEIFAHLGLDYIEPWERNA 509
R460Q       KRIFLKAESEEEIFAHLGLDYIEPWERNA 509
            ****************************
```

Figure 3.4: Protein multi-sequence alignment of DNA nucleotidylexotransferase isoform 1 [Homo sapiens] (NCBI Reference Sequence: NP_004079.3) with six chosen hTdTS SNPs. The positions of the point mutations are shown in bold;

- D280H; Asp to His mutation at protein position 280
- L397S; Leu to Ser mutation at protein position 397
- R431C; Arg to Cys mutation at protein position 431
- A445T; Ala to Thr mutation at protein position 445
- T450S; Thr to Ser mutation at protein position 450
- R460Q; Arg to Gln mutation at protein position 460

Star (*) represents fully conserved residues.
Colon (:) represents strong conservation between residues.
Gap ( ) represents non-conserved residues in the sequence alignment.
Multi-sequence alignment generated using *ClustalW* software.

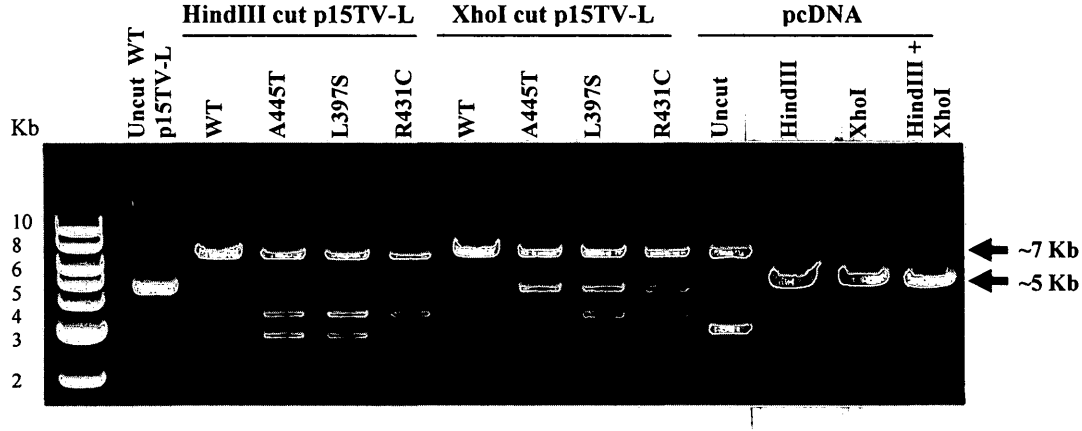


Figure 3.5: Restriction digest of hTdTS wild type, A445T, L397S, R431C cDNA p15TV-L vectors and pcDNA^{TM}4/myc-His A vector. A total of 10µl of each 50µl restriction digest reaction was ran on 0.7% agarose gel electrophoresis, stained with ethidium bromide and visualized under ultra-violet light.

- Uncut WT p15TV-L: 1µg WT hTdTS p15TV-L and 1xNEB Buffer 2 adjusted to 50µl with dH_2O.
- HindIII cut p15TV-L: 1µg hTdTS p15TV-L, 1xNEB Buffer 2, 10 units HindIII (NEB) adjusted to 50µl with dH_2O.
- XhoI cut p15TV-L: 1µg hTdTS p15TV-L, 1xNEB Buffer 4, 1xBSA, 10 units XhoI (NEB) adjusted to 50µl with dH_2O.
- Uncut pcDNA: 1µg pcDNA and 1xNEB Buffer 2 adjusted to 50µl with dH_2O.
- HindIII cut pcDNA: 1µg pcDNA, 1xNEB Buffer 2, 10 units HindIII (NEB) adjusted to 50µl with dH_2O.
- XhoI cut pcDNA: 1µg pcDNA, 1xNEB Buffer 4, 1xBSA, 10 units XhoI (NEB) adjusted to 50µl with dH_2O.
- HindIII+XhoI cut pcDNA: 1µg pcDNA, 1xNEB Buffer 2, 1xBSA, 10 units HindIII (NEB), 10 units XhoI adjusted to 50µl with dH_2O.

All restriction digest reactions were incubated at 37°C overnight.

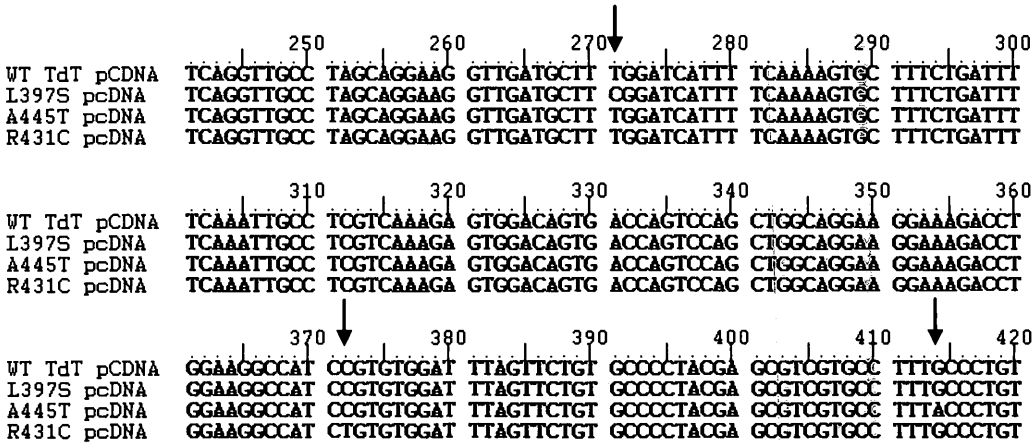


Figure 3.6: DNA multi-sequence alignment of cloned wild-type, L397S, R431C and A445T hTdTS cDNA in the pcDNA^{TM}4/myc-His A expression vector. Sequencing was carried out using SQ-TdT-974 primer. The positions of the single point mutations are represented by black arrows; L397S (T to C base change), A445T (G to A base change), R431C (C to T base change). Generated using *BioEdit* software.

Section 3.3 hTdT His-tag protein purification

Wild type hTdTS protein along with six selected hTdTS SNP candidate proteins were expressed in *E.Coli* BL21 and purified using nickel affinity chromatography. Figure 3.7 depicts His-tag wild type hTdTS purification. A 58kDa predominant protein band is seen in the elution sample, corresponding to purified wild type hTdTS. The remaining six purified hTdTS protein variant SDS-PAGE gels can be found in supplementary figures of Appendix B. All of the seven purified hTdTS protein variants were visualized on an SDS-PAGE gel in Figure 3.8.

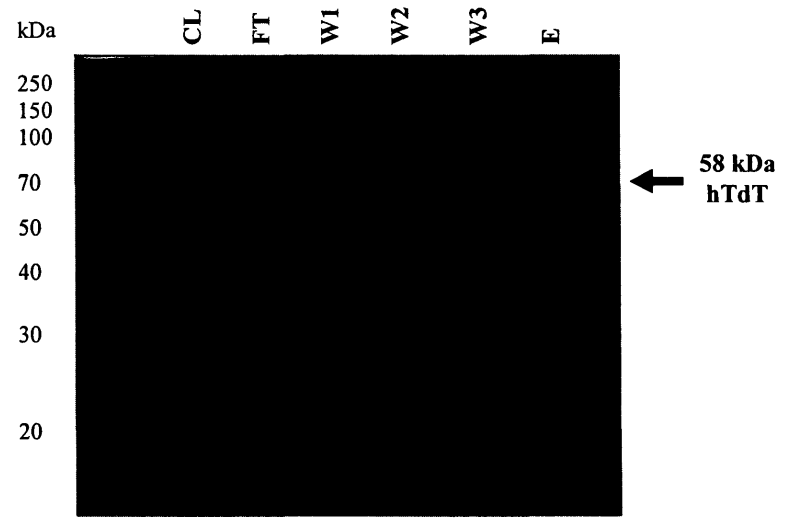


Figure 3.7: Coomassie stained 12% SDS-PAGE gel containing His-tag purification samples of wild type hTdTS. Crude lysate (CL) fraction was collected after cell lysis. Flow-through (FT) fraction was collected after incubation of crude lysate with the nickel resin. The wash (W) fractions were collected to test for removal of non-specific nickel resin binding. The elution (E) fraction depicts single protein band at around 58kDa, the expected size of hTdT.

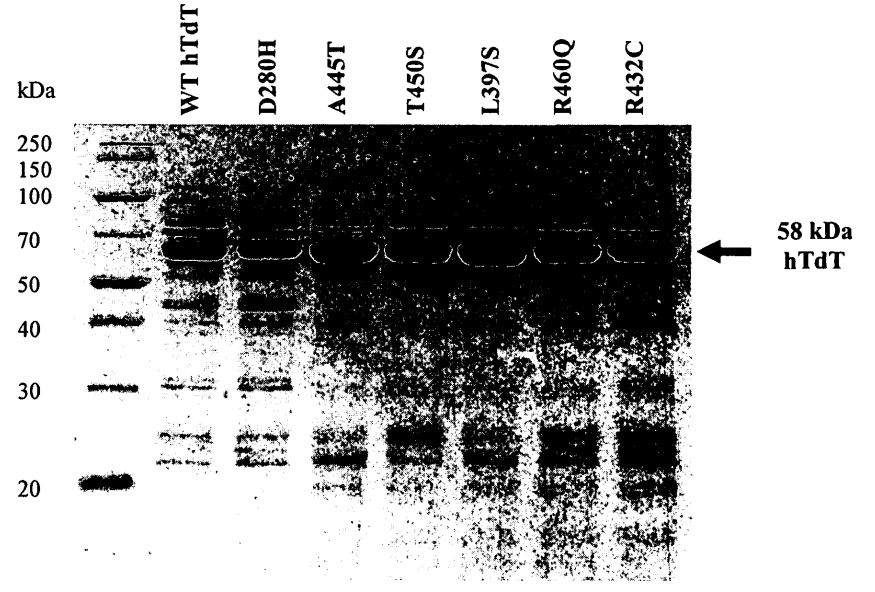


Figure 3.8: His-tag purified hTdTS protein variants resolved on 10% denaturing SDS-PAGE gel stained with Coomassie blue dye. A total of 8µg of protein was loaded in each lane.

Section 3.4 *In vitro* hTdT functional activity assays

3.4.1 Activity of commercial recombinant TdT versus purified wild type hTdTS

In order to determine whether the purified polymerase was functionally active, the polymerase activity of recombinant calf TdT was examined and compared to that of purified wild type hTdTS. The results are shown in Figure 3.9. In the absence of TdT, the Cy3-labelled oligo substrate was not extended even though the reaction contained available dNTPs. The same outcome was observed in the absence of free dNTPs, using either calf TdT or purified wild type hTdTS, as dNTPs are required by the polymerase reaction. When both TdT and dNTPs were present, the Cy3-labelled oligo substrate was extended by calf TdT or the purified wild type hTdTS. The activity was shown by the presence of a smear above the control band reaction as well as the disappearance of the Cy3-labelled oligo starting material in these reactions. As the TdT reaction was allowed to proceed for 1, 5 and 10 minutes the length of the reaction product was increased, indicative of increasing numbers of dNTPs polymerizing to the DNA substrate oligo with time. In addition, wild type hTdTS polymerase reactions resulted in longer polymerization products that appeared as darker smears as compared to recombinant calf TdT reactions.

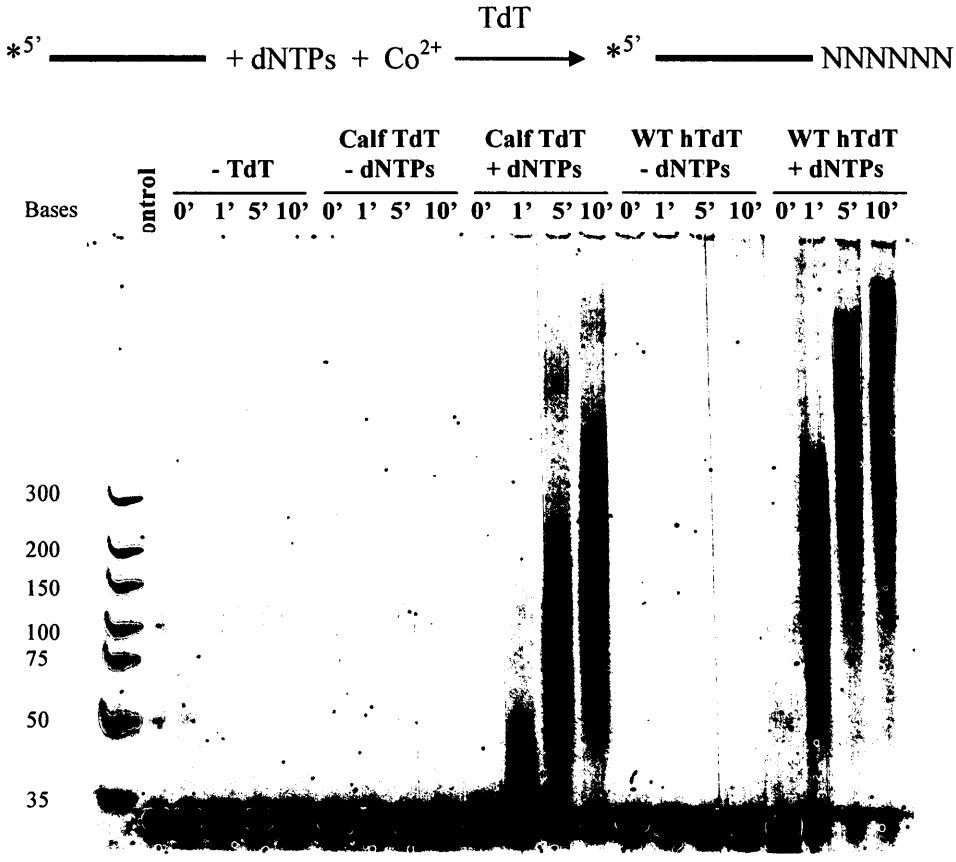


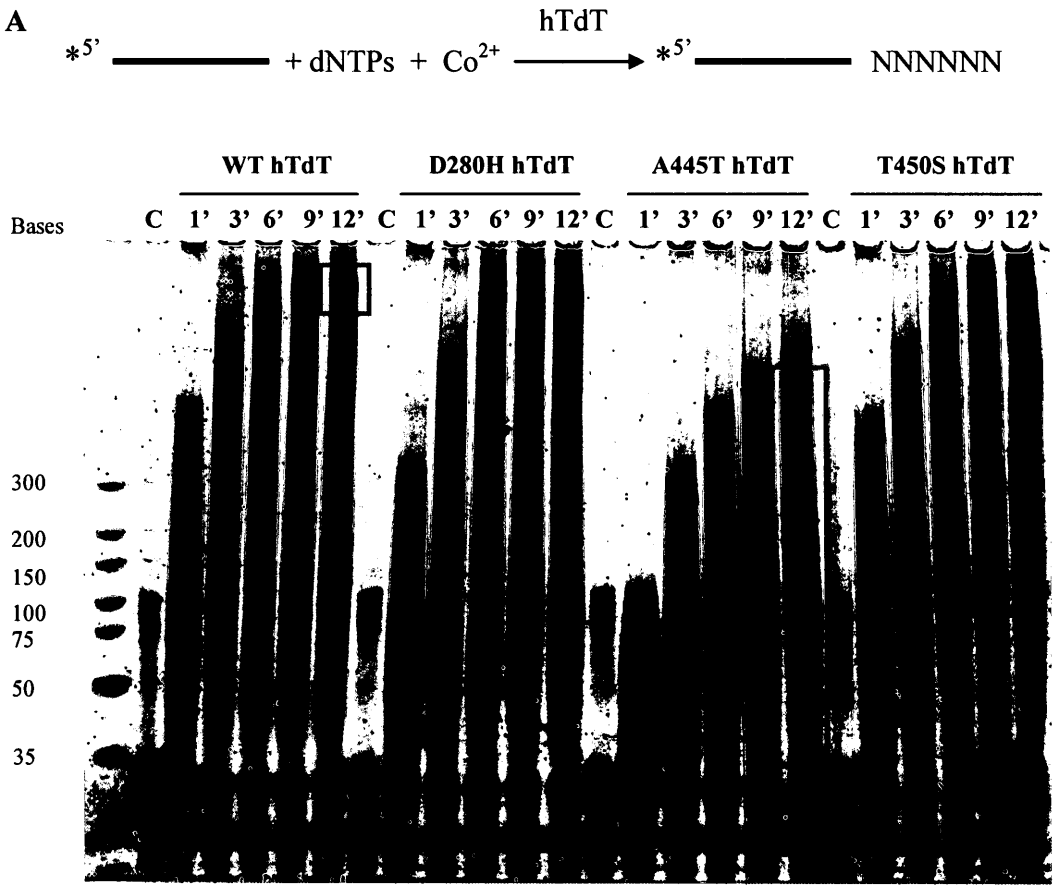
Figure 3.9: *In vitro* polymerase activity of commercial recombinant TdT versus purified wild type hTdTS. Reactions were set up as follows;
Control: 0.3nM Cy3 oligo, 1xTdT reaction buffer, 0.25mM $CoCl_2$, dH_2O up to 30µl.
-TdT: 0.3nM Cy3 oligo, 1xTdT reaction buffer, 0.25nM $CoCl_2$, 0.15mM dNTP mix, dH_2O up to 30µl. Reactions were incubated at 37°C for 0, 1, 5 and 10 minutes.
Calf TdT/-dNTPs: 0.3nM Cy3 oligo, 1xTdT reaction buffer, 0.25mM $CoCl_2$, 20 units recombinant calf TdT, dH_2O up to 30µl. Reactions were incubated at 37°C for 0, 1, 5 and 10 minutes.
Calf TdT/+dNTPs: 0.3nM Cy3 oligo, 1xTdT reaction buffer, 0.25mM $CoCl_2$, 0.15mM dNTP mix, 20 units recombinant calf TdT, dH_2O up to 30µl. Reactions were incubated at 37°C for 0, 1, 5 and 10 minutes.
WT hTdT/-dNTPs: 0.3nM Cy3 oligo, 1xTdT reaction buffer, 0.25mM $CoCl_2$, 3.4µM purified wild type hTdT protein, dH_2O up to 30µl. Reactions were incubated at 37°C for 0, 1, 5 and 10 minutes.
WT hTdT/+dNTPs: 0.3nM Cy3 oligo, 1xTdT reaction buffer, 0.25mM $CoCl_2$, 0.15mM dNTP mix, 3.4µM purified wild type hTdT protein, dH_2O up to 30µl. Reactions were incubated at 37°C for 0, 1, 5 and 10 minutes.
The reactions were stopped by the addition of 30µl TdT stopping buffer and resolved on denaturing 10% TBE-urea gel. Image obtained by Typhoon Imager scanning.

3.4.2 Polymerase time course activity assay

In order to assess the differences between wild type hTdTS and the six purified hTdTS variants, time course of polymerase activity was performed (Figure 3.10). All of the six assayed hTdTS variants demonstrated different polymerase activities under the assay conditions. Based on the length of TdT reaction products as well as the banding patterns observed. Five out of six assayed hTdTS variants were functionally active, while R431C hTdTS was unable to extend DNA oligo substrate (Figure 3.10, panel B). Thus, R431C variant was termed a 'dead' mutant in the biochemical assays.

The graph in Figure 3.11 plots the change in maximum size of hTdTS oligo product with respect to time. The initial polymerization rate within the first minute of reaction differs among hTdTS variants. Wild type, D280H and T450S hTdTS reactions all resulted in polymerization product of about 300 bases, while R460Q, A445T and L397S hTdTS reaction resulted in products of 160, 75 and 50 bases respectively.

Another key difference among the five hTdTS variants as compared to wild type hTdTS is the banding pattern of the polymerization products, specifically broadness versus tightness of the smears. For instance, wild type hTdTS reactions produce tight smears as compared to A445T variant which results in overall broader smears (highlighted by a red box in Figure 3.10, panel A).


A
hTdT
*5'
+ dNTPs + Co2+
*5'
NNNNNN
WT hTdT
D280H hTdT
A445T hTdT
T450S hTdT
Bases
C 1' 3' 6' 9' 12' C 1' 3' 6' 9' 12' C 1' 3' 6' 9' 12' C 1' 3' 6' 9' 12'
300
200
150
100
75
50
35

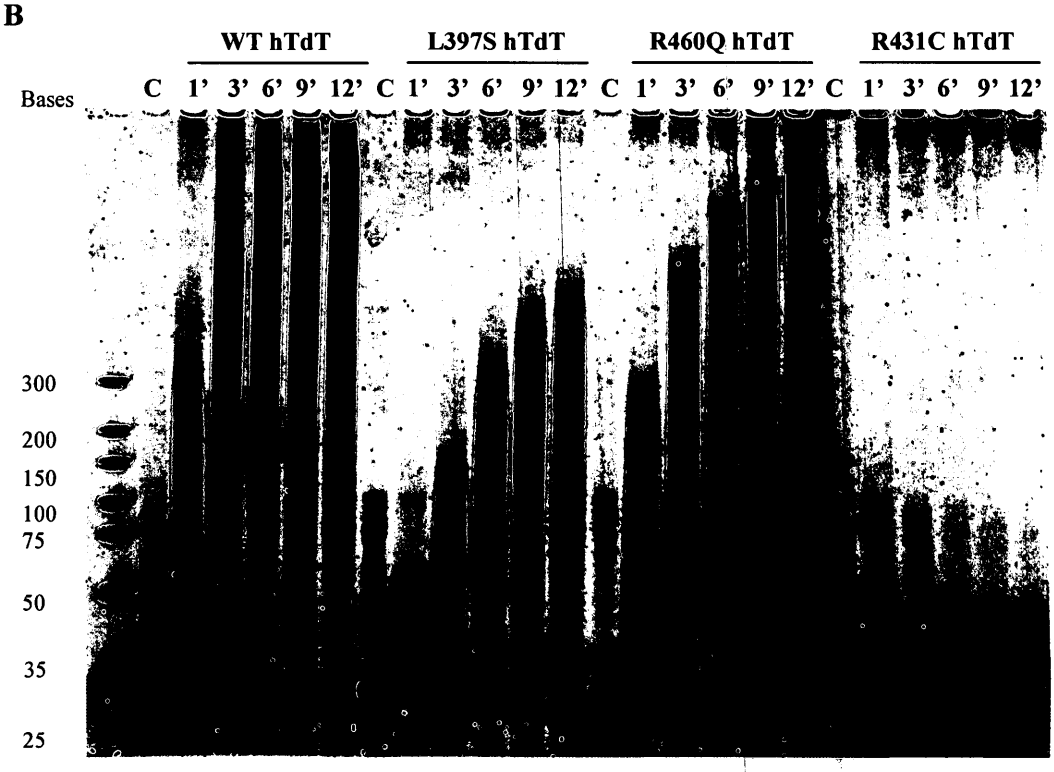


Figure 3.10: *In vitro* polymerase time course activity of purified hTdTS SNP variants. Panel A assays the activity of wild type, D280H, A445T and T450S hTdTS. Red boxes highlight the banding pattern difference between the polymerization products of wild type and A445T variants. Panel B assays the activity of wild type, L397S, R460Q and R431C hTdTS.
Control reaction contained 0.3nM Cy3 oligo, 1xTdT reaction buffer, 0.25mM $CoCl_2$, 0.15mM dNTP mix, dH_2O up to 30µl. Time course reactions were set up as a master mix. Each polymerase reaction contained 0.3nM Cy3 oligo, 1xTdT reaction buffer, 0.25mM $CoCl_2$, 0.15mM dNTP mix, 3.4µM purified hTdTS protein, dH_2O up to 30µl. Reactions were set up as a master mix and incubated at 37°C. Aliquots of 30µl were taken at 1, 3, 6, 9 and 12 minutes and added to 30µl TdT stopping buffer. The completed reactions were resolved on a denaturing 10% TBE-urea gel and the image obtained by Typhoon Imager scanning.

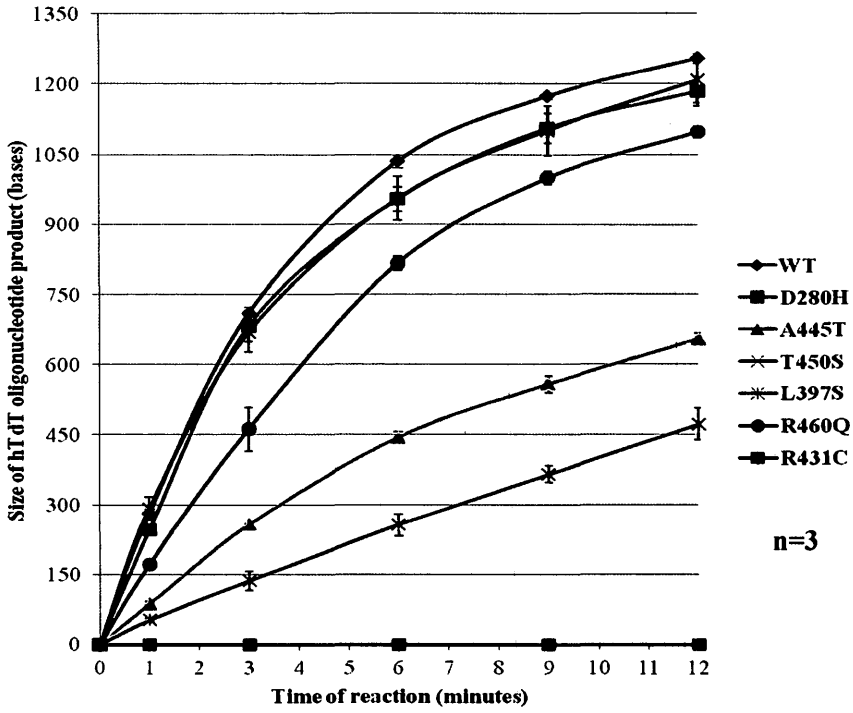


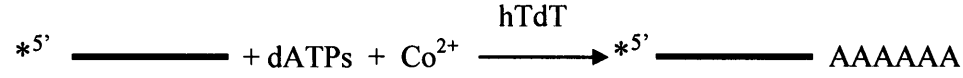
Figure 3.11: Change in maximum size of hTdTS oligo product with respect to time of *in vitro* reaction. The graph plotted based on time course polymerase activity assay of wild type hTdTS and all six purified hTdTS SNP variants. The size of hTdT oligonucleotide product was extrapolated based on known sizes of DNA ladder that were resolved along with TdT reactions. The assays were independently repeated 3 times and the standard error was calculated and plotted on the graph.

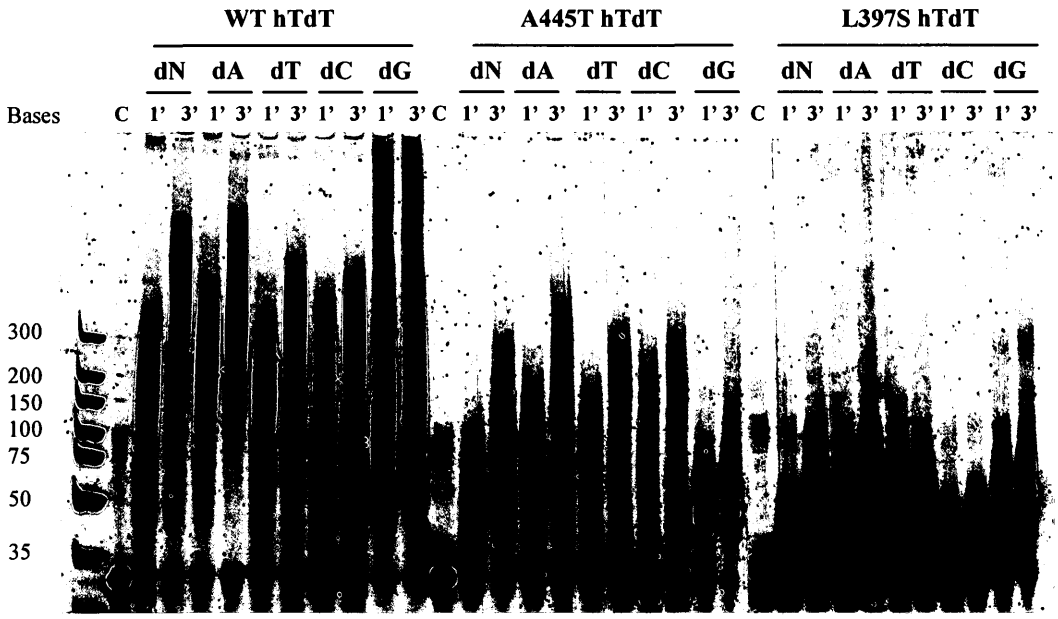
3.4.3 dNTP substrate preference polymerase activity assay

Once the polymerase activities of hTdTS variants were determined through the time course assays, the usage of a single type of dNTP was tested to observe if there is any preference of hTdTS variants to use certain nucleotide for the polymerization reaction. Figure 3.12 shows the TdT polymerization reactions supplemented with a dNTP mix, dATPs, dTTPs, dCTPs or dGTPs. The dNTPs generated by far the best reactions, each single deoxy-nucleotide resulted in some change in activity of the polymerases, with dCTP and dGTP causing the most variation. When comparing the activity of either of hTdTS variants to wild type protein, few key points can be made. First, five hTdTS variants supplemented with dCTPs produce short oligo products as compared to wild type hTdTS. In addition, dCTPs reactions with all hTdTS variants, including wild type protein, result in lower size band than the control Cy3-labelled oligo substrate. Second, all dGTPs supplemented reactions with all six hTdTS variants show a predominant band at around 50 bases. Lastly, dNTPs and dATPs supplemented reactions result in relatively long TdT oligo products as compared to dTTPs, dCTPs and dGTPs reactions by all six hTdTS variants. R431C hTdTS showed no polymerase activity with either of the supplemented dNTPs.

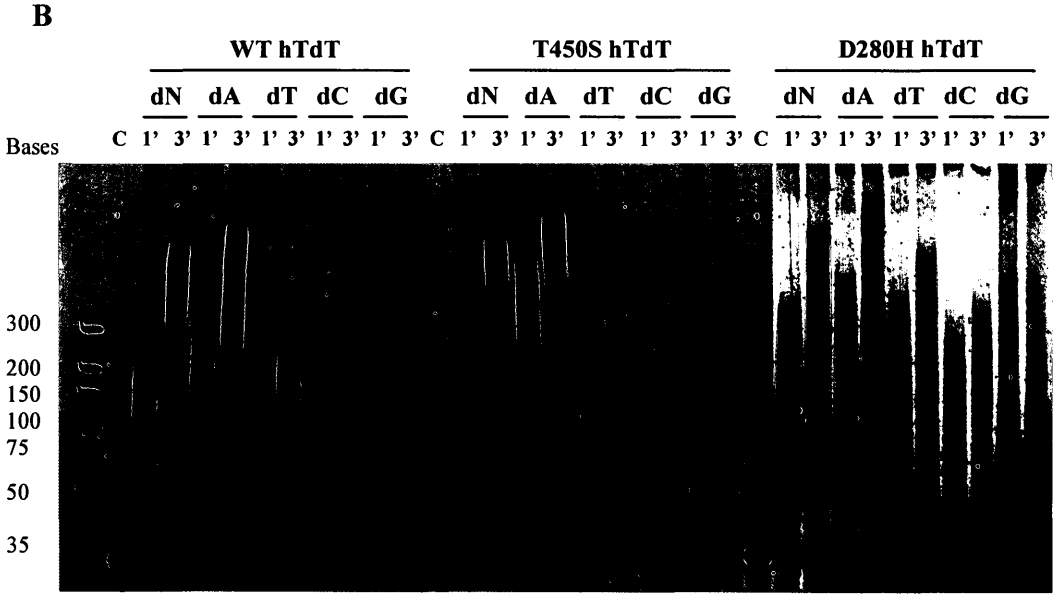
To investigate dGTPs supplemented reactions, an additional time course assay was performed in Figure 3.13. As the wild type hTdTS reaction was allowed to proceed, no extension of the Cy3-labelled oligo substrate was observed beyond the predominant band at around 50 bases.

A


hTdT
*5'
+ dATPs + Co2+
*5'
AAAAAA


WT hTdT
A445T hTdT
L397S hTdT
dN dA dT dC dG
Bases
C 1' 3' 1' 3' 1' 3' 1' 3' 1' 3' C 1' 3' 1' 3' 1' 3' 1' 3' 1' 3' C 1' 3' 1' 3' 1' 3' 1' 3' 1' 3'
300
200
150
100
75
50
35

B


WT hTdT
T450S hTdT
D280H hTdT
dN dA dT dC dG
Bases
C 1' 3' 1' 3' 1' 3' 1' 3' 1' 3' C 1' 3' 1' 3' 1' 3' 1' 3' 1' 3' C 1' 3' 1' 3' 1' 3' 1' 3' 1' 3'
300
200
150
100
75
50
35

C

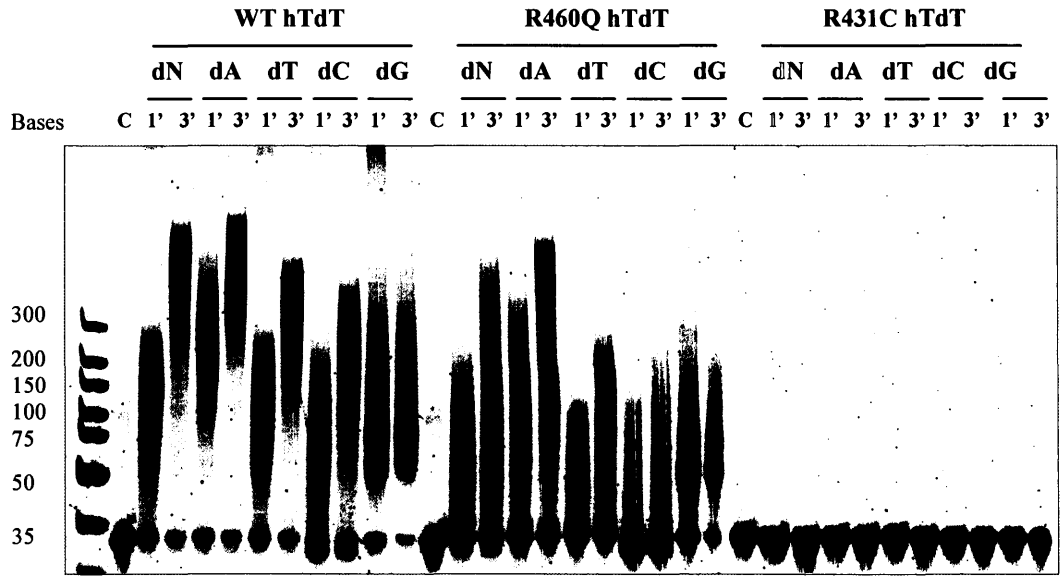


Figure 3.12: *In vitro* dNTPs preference polymerase activity of purified hTdTS SNP variants. Panel A assays the activity of wild type, A445T and L397S hTdTS. Panel B assays the activity of wild type, T450S and D280H hTdTS. Panel C assays the activity of wild type, R460Q and R431C hTdTS.
Control reaction contained 0.3nM Cy3 oligo, 1xTdT reaction buffer, 0.25mM $CoCl_2$, 0.15mM dNTP mix, dH_2O up to 30µl. Time course reactions were set up as a master mix. Each polymerase reaction contained 0.3nM Cy3 oligo, 1xTdT reaction buffer, 0.25mM $CoCl_2$, 0.15mM dNTPs (either dNTP mix, dATPs, dTTPs, dCTPs or dGTPs), 3.4µM purified hTdTS protein, dH_2O up to 30µl. Reactions were set up as a master mix and incubated at 37°C. Aliquots of 30µl were taken at 1 and 3 minutes and added to 30µl TdT stopping buffer. The completed reactions were resolved on a denaturing 10% TBE-urea gel and the image obtained by Typhoon Imager scanning.

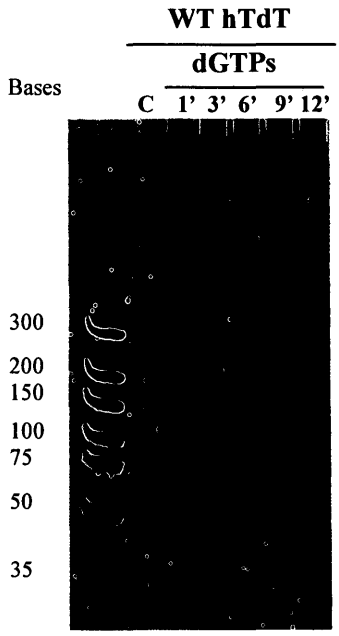
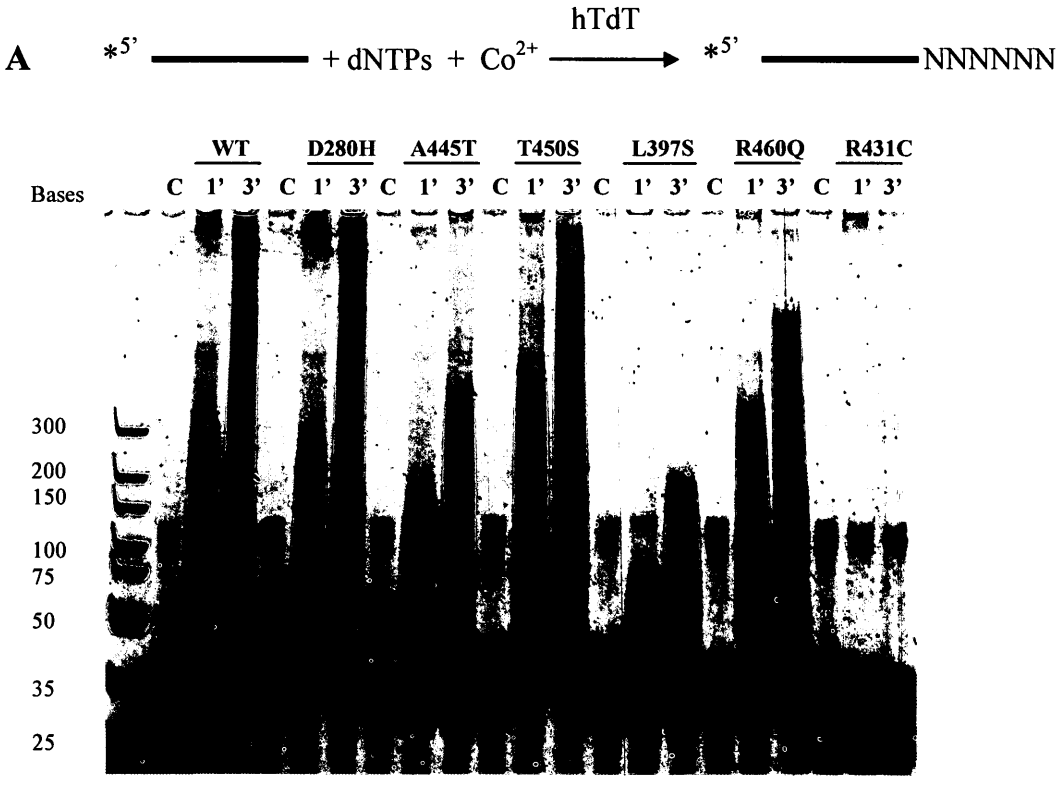


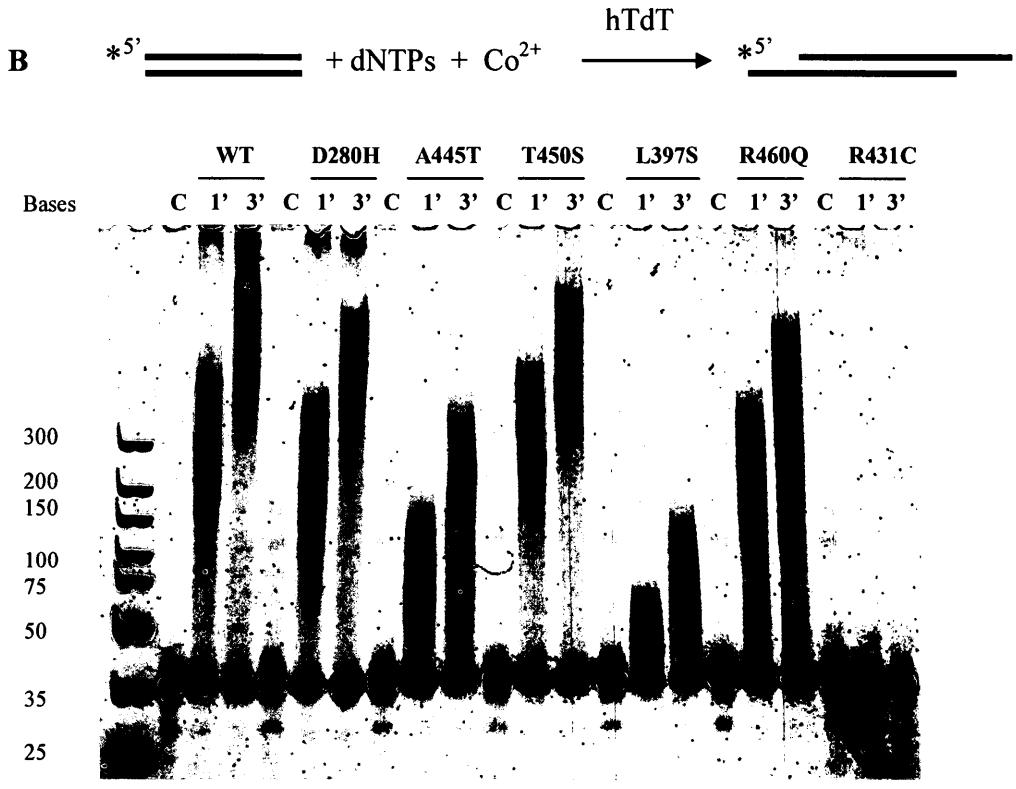
Figure 3.13: *In vitro* polymerase activity of purified wild type hTdTS supplemented with dGTPs. Control reaction contained 0.3nM Cy3 oligo, 1xTdT reaction buffer, 0.25mM $CoCl_2$, 0.15mM dNTP mix, dH_2O up to 30µl. Time course reactions were set up as a master mix. Each polymerase reaction contained 0.3nM Cy3 oligo, 1xTdT reaction buffer, 0.25mM $CoCl_2$, 0.15mM dGTPs, 3.4µM purified wild type hTdTS protein, dH_2O up to 30µl. Reactions were set up as a master mix and incubated at 37°C. Aliquots of 30µl were taken at 1, 3, 6, 9 and 12 minutes and added to 30µl TdT stopping buffer. The completed reactions were resolved on a denaturing 10% TBE-urea gel and the image obtained by Typhoon Imager scanning.

3.4.4 DNA substrate preference polymerase activity assays

The activity of hTdTS variants was tested in terms of their preference to utilize different types of Cy3-labelled DNA substrate, specifically single stranded oligo, a double stranded oligo having a blunt end or a double stranded oligo having a 3' end overhang (Figure 3.14). The DNA substrate assay shows a clear preference by all hTdTS variants to utilize double stranded 3' overhang oligo (panel C) due to the absence of any residual starting DNA substrate oligo within the first minute of reaction, as compared to

single stranded and double stranded blunt end oligos that show the presence of the starting material DNA substrate (panels A and B). The double stranded blunt end DNA substrate is the least preferred by all hTdTS variants when looking at the left-over starting DNA substrate after 1 and 3 minutes. In addition, R431C hTdTS demonstrated no polymerase activity with either of the DNA substrate oligos.




B
*5'
+ dNTPs + Co2+
hTdT
*5'
WT
D280H
A445T
T450S
L397S
R460Q
R431C
Bases
C 1' 3' C 1' 3' C 1' 3' C 1' 3' C 1' 3' C 1' 3' C 1' 3'
300
200
150
100
75
50
35
25

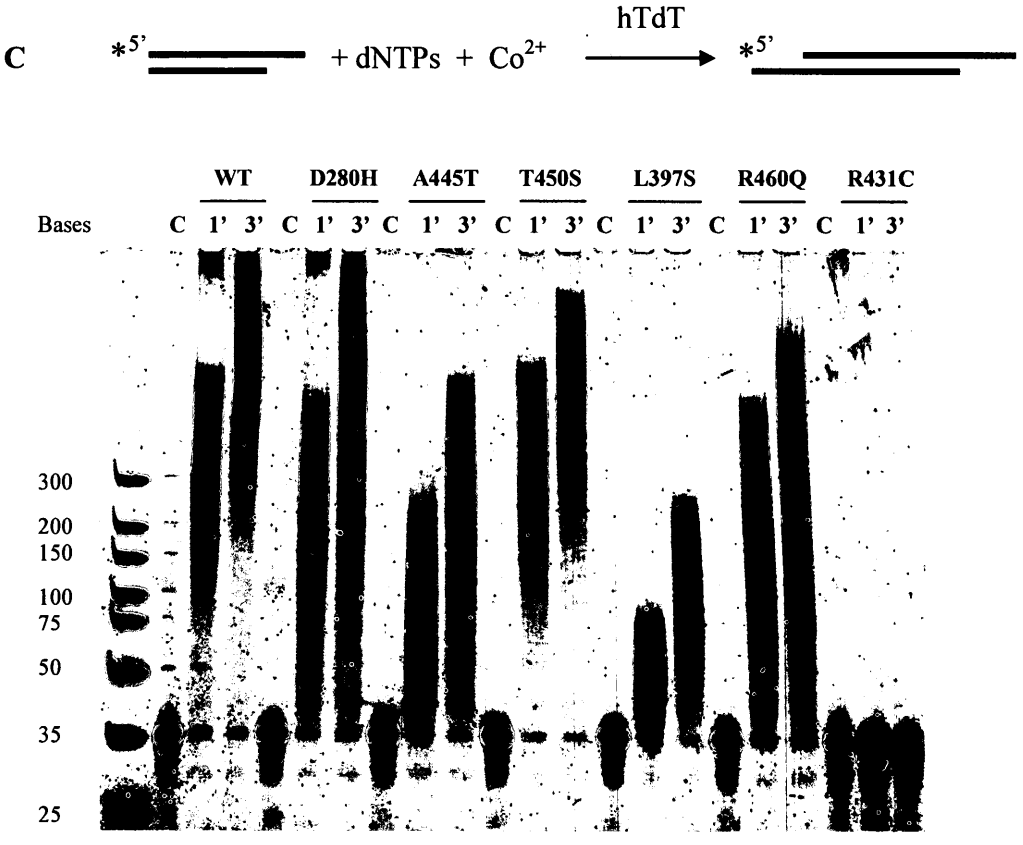


Figure 3.14: *In vitro* DNA substrate preference polymerase activity of purified hTdTS SNP variants. Panel A assays the activity using single stranded Cy3-labelled oligonucleotide. Panel B assays the activity using double stranded Cy3-labelled blunt end oligonucleotide. Panel C assays the activity using double stranded Cy3-labelled 3' overhang oligonucleotide.
Control reaction contained 0.3nM Cy3 oligo, 1xTdT reaction buffer, 0.25mM $CoCl_2$, 0.15mM dNTP mix, dH_2O up to 30µl. Time course reactions were set up as a master mix. Each polymerase reaction contained 0.3nM DNA oligo substrate (single stranded, double stranded blunt end or double stranded 3' overhang), 1xTdT reaction buffer, 0.25mM $CoCl_2$, 0.15mM dNTP mix, 3.4µM purified hTdTS protein, dH_2O up to 30µl. Reactions were set up as a master mix and incubated at 37°C. Aliquots of 30µl were taken at 1 and 3 minutes and added to 30µl TdT stopping buffer. The completed reactions were resolved on a denaturing 10% TBE-urea gel and the image obtained by Typhoon Imager scanning.

3.4.5 DNA substrate sequence preference polymerase assays

In order to observe whether TdT polymerase activity may be affected by the specific nucleotide sequence of the DNA substrate oligo, a time course assay was undertaken using two Cy3-labelled 33 nucleotide long DNA substrate oligos that are complementary to one another in their base sequence. Figure 3.15 panel A shows the base sequence of the two tested oligos. No difference in terms of the length of TdT reaction products or the banding pattern was observed between the two tested oligos in the presence of wild type hTdTS (Figure 3.15, panel B). In addition, the usage of both DNA substrate oligos in the polymerization reaction in the presence of dGTPs resulted in the production of a predominant band (Figure 3.15, panel C). This was similarly seen in dNTP substrate preference assay depicted in Figure 3.12.

A

Substrate oligo 1 5' Cy3-cta ctg gta ctt cga tct ctg ggg ccg tga cgc 33nt long

Substrate oligo 2 5' Cy3-gtg cca cgg ccc cag aga tcg aag tac cag tag 33nt long

Nanomer 23-bp heptamer IGHJ2

5' gtg ttt ttg tat ggg aga agc agg agg gca gag gct gtg *cta ctg gta ctt cga tct ctg ggg ccg tgg cac 3'…

3' cac aaa aac ata ccc tct tcg tcc tcc cgt ctc cga cac gat gac cat gaa gct aga gac ccc ggc acc gtg*5' …

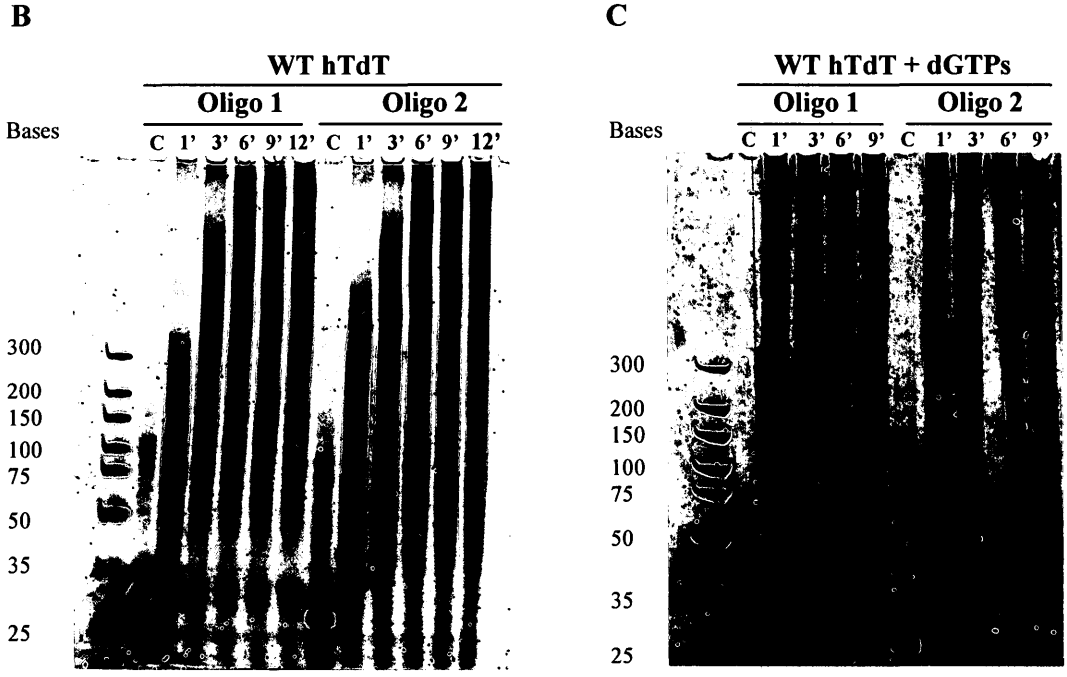


Figure 3.15: *In vitro* DNA substrate sequence preference polymerase activity assay.

A. Sequences of Cy3-labelled oligo 1 and 2 used for *in vitro* DNA substrate sequence preference polymerase activity assay. Germline sequence of human immunoglobulin heavy chain J2 region flanked by 23-RSS region (adapted from NCBI database) shown at the bottom.

B. Utilization of oligo 1 versus oligo 2 by purified wild type hTdTS.
Control reaction contained 0.3nM Cy3-labelled oligo 1 or 2, 1xTdT reaction buffer, 0.25mM $CoCl_2$, 0.15mM dNTP mix, dH_2O up to 30µl. Time course reactions were set up as a master mix. Each polymerase reaction contained 0.3nM Cy3-labelled oligo 1 or 2, 1xTdT reaction buffer, 0.25mM $CoCl_2$, 0.15mM dNTP mix, 3.4µM purified wild type hTdTS protein, dH_2O up to 30µl. Reactions were set up as a master mix and incubated at 37°C. Aliquots of 30µl were taken at 1, 3, 6, 9 and 12 minutes and added to 30µl TdT stopping buffer. The completed reactions were resolved on a denaturing 10% TBE-urea gel and the image obtained by Typhoon Imager scanning.

C. Utilization of oligo 1 versus oligo 2 by purified wild type hTdTS in the presence of dGTPs. Reactions were carried out as described in caption to panel B. The reactions were supplemented with 0.15mM dGTPs instead of dNTP mix.

3.4.6 Cofactor preference polymerase activity assays

TdT requires a divalent cofactor for function *in vitro*. Therefore, TdT dependency of the divalent metal cofactors (Co^{2+} and Mg^{2+}) was examined. Figure 3.16 shows cofactor preference assay of wild type hTdTS and R431C hTdTS variants. No differences in the lengths of wild type hTdTS reaction products were observed between the three conditions (with Co^{2+}, without Co^{2+} and with Mg^{2+}). However, the banding pattern of the reactions supplemented with cobalt (Co^{2+}) resulted in tighter smears as compared to reactions in the absence of cobalt or the presence of magnesium (Mg^{2+}). R431C hTdTS reactions in any of the three tested conditions did not result in polymerization products, however smearing was observed below the control band in the absence of cobalt (Co^{2+}) or the presence of magnesium (Mg^{2+}).

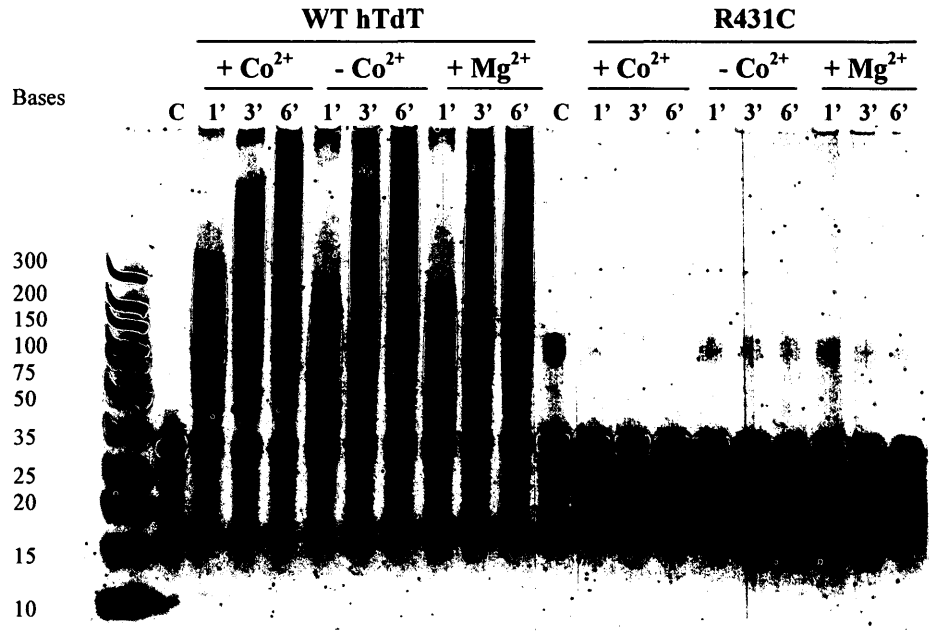


Figure 3.16: *In vitro* cofactor polymerase activity assay.
Control reaction contained 0.3nM Cy3 oligo, 1xTdT reaction buffer, 0.15mM dNTP mix, dH_2O up to 30µl. Time course reactions were set up as a master mix. Each polymerase reaction contained 0.3nM Cy3 oligo, 1xTdT reaction buffer, 0.25mM cofactor (either $CoCl_2$, no $CoCl_2$ or $MgCl_2$), 0.15mM dNTP mix, 3.4µM purified wild type or R431C hTdTS, dH_2O up to 30µl. Reactions were set up as a master mix and incubated at 37°C. Aliquots of 30µl were taken at 1, 3 and 6 minutes and added to 30µl TdT stopping buffer. The completed reactions were resolved on a denaturing 10% TBE-urea gel and the image obtained by Typhoon Imager scanning.

Section 3.3 *In vivo* V(D)J recombination assays

To investigate the functional differences between polymorphic hTdTS variants, *in vivo* V(D)J extrachromosomal recombination assays were performed. The assay focused on A445T, L397S and R431C hTdTS from the six hTdTS SNP candidates, since these three demonstrated marked differences in *in vitro* functional assays as compared to wild type hTdTS.

3.3.1 HEK293T cell line testing

Endogenous hTdT expression was tested in HEK293T cells prior to performing recombination assay. This test was important since the differences among transfected polymorphic forms of hTdT were assayed and any background hTdT expression would otherwise interfere with the assay. The cell line was tested at the RNA level through RT-PCR and also at the protein level through Western blotting.

The results of the RT-PCR experiment are depicted in Figure 3.17. Total RNA isolated from HEK293T cells that were transfected with either substrate pGG51 plasmid, empty pcDNA expression vector, hTdTS pcDNA expression vector or co-transfected with substrate pGG51 plasmid, hRAG1 and hRAG2 expression vectors along with hTdTS pcDNA expression vector. RAG1 and RAG2 expression vectors were co-transected since HEK293T cells do not express endogenous RAG proteins that are required at the first stage of V(D)J recombination. Reverse transcription reaction was performed using total RNA as template along with hTdTS gene-specific primers either in the presence or absence of reverse transcriptase to test for possible genomic DNA contamination. The presence of a DNA band around 1Kb following the PCR reaction

indicates the presence of hTdTS RNA. Non-transfected HEK293T cells did not produce hTdTS PCR product either in the absence or presence of reverse transcriptase. The same results were observed in HEK293T cell transfection with substrate pGG51 plasmid and empty pcDNA expression vector. The RT-PCR reaction of HEK293T cells transfected with wild type hTdTS pcDNA expression vector resulted in a single predominant band around 1Kb in the presence of reverse transcriptase. In the absence of reverse transcriptase the DNA band was not present, suggesting that no genomic DNA contamination occurred in the reaction. Lastly, HEK293T cell co-transfection of four plasmids required for recombination assay resulted in hTdTS PCR product in the presence of reverse transcriptase. In the absence of reverse transcriptase, no PCR product was observed.

Human TdT Western blot results are shown in Figure 3.18. Endogenous hTdT protein was not detected in non-transfected HEK293T cells as shown in the far left lane. However, upon transfection of hTdTS pcDNA expression vector hTdTS protein was detected using rabbit polyclonal TdT specific antibody. The protein expression level of the four plasmid co-transfection was lower as compared to the single hTdTS pcDNA expression vector transfection, as shown by the fainter protein signal. In addition, TdT specific antibody detected purified His-tagged wild type hTdTS protein used in the *in vitro* functional assays. This result is shown in the far right lane and served as a positive control for the activity of the TdT specific antibody.

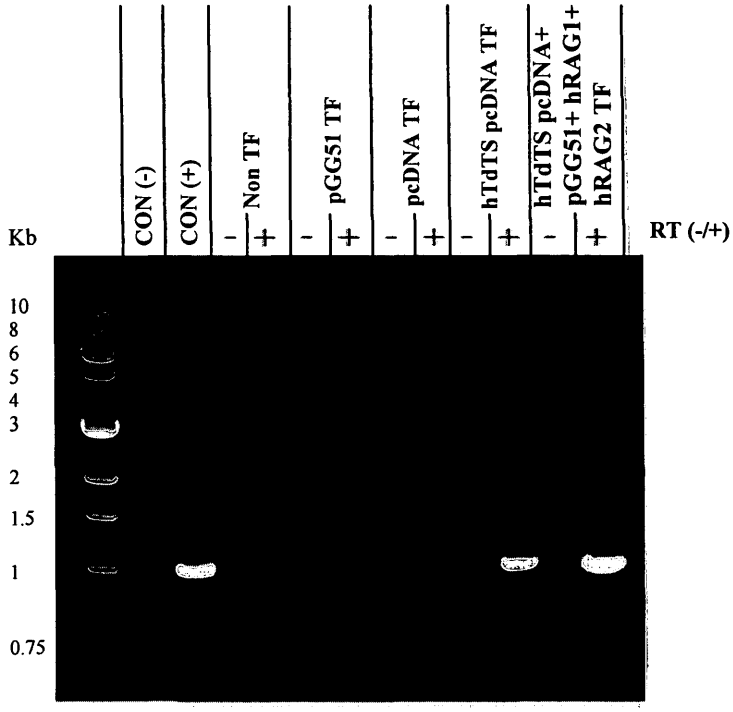


Figure 3.17: Reverse transcriptase PCR products using hTdTS gene-specific primers, separated on 0.7% agarose gel.

HEK293T cells were transfected with the indicated vectors; substrate recombination plasmid pGG51, empty pcDNA^{TM}4/myc-His A vector, wild type hTdTS pcDNA^{TM}4/myc-His A expression vector or co-transfection of pGG51 plasmid, wild type hTdTS pcDNA, pEBG-hRAG1 and hRAG2 expression vectors. Non-transfected HEK293T cells were used as experimental control.

Total RNA was extracted 48 hours post transfection using TRIzol® Reagent. Total isolated RNA was reverse transcribed (either + or - RT) to generate cDNA using oligo $(dT)_{20}$, as per ThermoScript^{TM} RT-PCR system. Generated cDNA was PCR amplified using hTdTS gene-specific primers (DNTT_F and DNTT_R).

PCR negative control included no cDNA template. PCR positive control reaction included wild type hTdT pcDNA^{TM}4/myc-His A expression vector as DNA template.

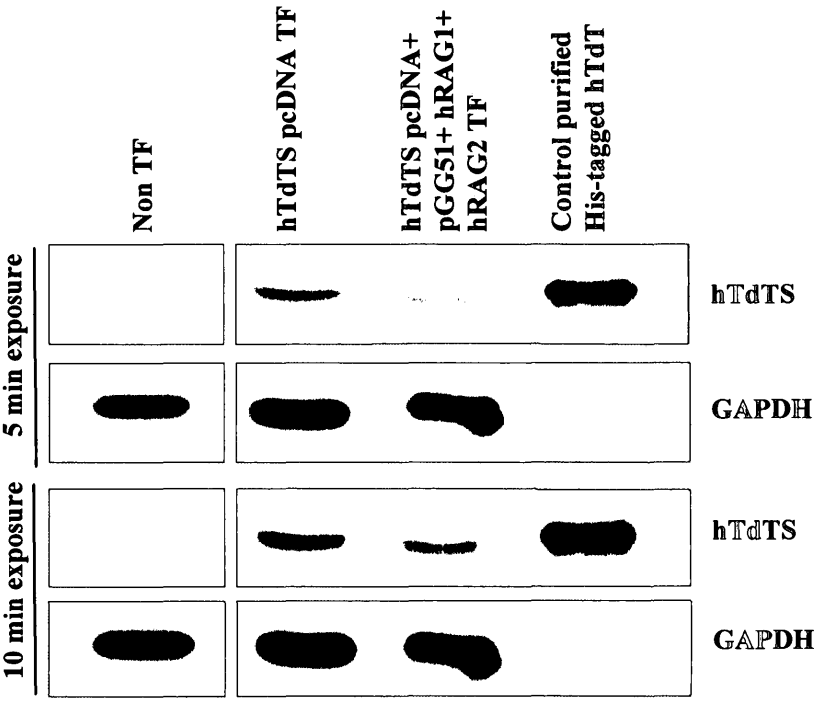


Figure 3.18: Western blot image of total HEK293T cell lysates probed with rabbit polyclonal TdT specific antibody and mouse monoclonal IgG GAPDH specific antibody. Total HEK293T cell lysate (85μg of total protein) harvested 48 hours post-transfection (wild type hTdTS pcDNA expression vector or co-transfection of wild type hTdTS pcDNA expression vector, substrate pGG51 plasmid, pEBG-hRAG1 and hRAG2 expression vectors) ran on 12% SDS-PAGE gel and transferred onto activated protein membrane. Cell lysate of non-transfected HEK293T cells was also loaded onto SDS-PAGE gel. The membrane was then blocked and probed with primary antibodies (rabbit polyclonal TdT specific antibody and mouse monoclonal IgG GAPDH specific antibody) followed by probing with secondary antibody (anti-rabbit HRP-conjugated antibody and anti-mouse IgG HRP-conjugated antibody). The signal on the membrane was developed using HRP substrate for either 5 or 10 minute exposure. GAPDH utilized as loading control. Bacterial purified His-tagged wild type hTdTS served as hTdTS antibody control.

3.3.2 Recombination frequencies of the substrate pGG51 plasmid

The recombination frequencies (R-values, detailed in Chapter 2, Materials and Methods) for every HEK293T cell transfection were computed based on the number of ampicillin resistant versus chloramphenicol and ampicillin resistant bacterial colonies detected. The ampicillin resistant colonies represent the total number of bacterial cells that have been transformed with the substrate pGG51 plasmid, while the double-resistant colonies represent the total number of substrate plasmids that undergone RAG-mediated recombination. The raw data from the recombination assay are summarized in Appendix C. The graph in Figure 3.19 plots individual R-values and the mean R-value calculated based on eight independent HEK293T cell transfections. The mean R-value for L398S hTdTS experimental condition is 1.5 fold higher when compared to wild type hTdTS experimental condition (0.00498% versus 0.00318%). No marked differences were observed among A445T and R431C hTdTS and wild type hTdTS. The mean R-value for control experimental conditions (no hTdTS expression plasmid added) is two fold higher compared to wild type hTdTS experimental condition (0.00662% versus 0.00318%).

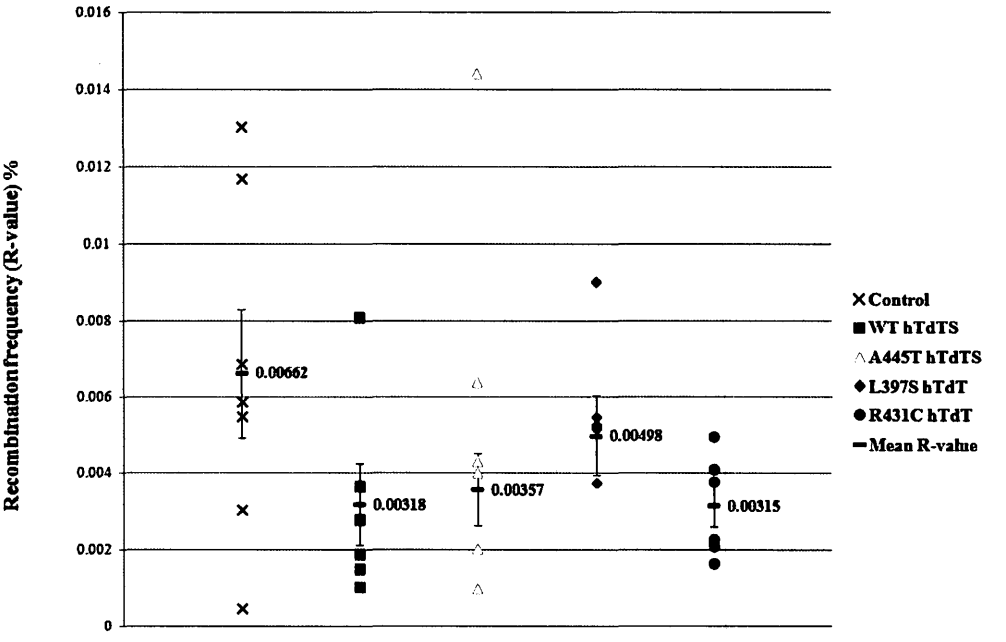


Figure 3.19: Recombination frequencies of *in vivo* V(D)J recombination assay. The individual R-values computed based on 8 independent HEK293T cell transfections per single experimental condition. The mean R-values are shown along with the standard error, represented by error bars.
The mean R-values are as follows;
Control= 0.00662 ± 0.00169 %
WT hTdTS= 0.00318 ± 0.00106 %
A445T hTdTS= 0.00357 ± 0.00094 %
L397S hTdTS= 0.00498 ± 0.00104 %
R431C hTdTS= 0.00315 ± 0.00054 %

Experimental conditions are as follows;
Control = transfection with substrate pGG51 plasmid + pEBG-hRAG1 + pEBG-hRAG2 expressing vectors
WT = transfection with wild type hTdTS pcDNA expressing vector + substrate pGG51 plasmid + pEBG-hRAG1+ pEBG-hRAG2 expressing vectors
A445T = transfection with A445T hTdTS pcDNA expressing vector + substrate pGG51 plasmid + pEBG-hRAG1+ pEBG-hRAG2 expressing vectors
L397S = transfection with L397S hTdTS pcDNA expressing vector + substrate pGG51 plasmid + pEBG-hRAG1+ pEBG-hRAG2 expressing vectors
R431C = transfection with R431C hTdTS pcDNA expressing vector + substrate pGG51 plasmid + pEBG-hRAG1+ pEBG-hRAG2 expressing vectors

3.3.3 Analysis of recombined substrate pGG51 sequences

Analysis of *in vivo* functional activity differences between wild type hTdTS and its polymorphic forms was carried out using the extrachromosomal V(D)J recombination substrate assay in HEK293T cells. Recombination substrate pGG51 plasmid was allowed to undergo RAG-mediated recombination in mammalian cells either in the presence of hTdTS (wild type, A445T, L397S or R431C) or its absence. The substrate plasmid was harvested from the mammalian cells and sequenced to analyze the junctional region where recombination occurred. A total of 276 unique recombined substrate plasmid sequences were obtained from the tested experimental conditions (Appendix D). Repeated recombined sequences within the same mammalian transfection were excluded from the analysis due to the possibility of plasmid replication in mammalian cells after recombination (Appendix C, Table 3).

The sequences were subjected to analysis in terms of number of nucleotide deletions, palindromic nucleotide additions (P-additions) and random nucleotide additions (N-additions). In addition, the frequency of N-additions as well as their AT versus GC base content was investigated. The distribution of N-nucleotide additions per recombined joint was examined. Finally, the length of the arbitrary complementary-determining region (CDR) was compared amongst the experimental conditions.

Nucleotide deletions

Nucleotide deletions at the recombined joints were analyzed as experimental control, since hTdTS is known to catalyze nucleotide additions and not deletions. The summary of the nucleotide deletions observed is shown in Table 3.2. There was no

marked difference between the control experiment, lacking hTdTS, and other tested conditions including those with hTdTS variants. The average number of nucleotide deletions in L397S hTdTS transfection was determined to be 5.17 nucleotides, which is higher than 4.53 nucleotides as compared to WT hTdTS. However, this difference is not statistically significant (p-value ≥ 0.05).

Table 3.2: Recombined substrate pGG51 plasmid nucleotide deletions.

	Control	WT hTdTS	A445T hTdTS	L397S hTdTS	R431C hTdTS
Total number of unique sequences obtained	48	57	65	72	34
Total number of nucleotide deletions	230	258	298	372	148
Average number of deletions *	4.79	4.53	4.58	5.17	4.35
P-value		0.4977**	0.8643	0.0717	0.6608

* Average number of deletions = total number of nucleotide deletions / total number of unique sequences
** P-value for nucleotide deletions for WT hTdTS experimental condition computed using comparison to the control condition. All other p-values computed through comparison to WT hTdTS experimental condition.
*** The statistical analysis was performed using a standard T-test.

P-nucleotide additions

Palindromic nucleotide additions (P-additions) result from off-center hairpin opening of DNA during the process of V(D)J recombination. P-additions at the recombined joints were analyzed as experimental control, since hTdTS does not play a role in hairpin openining *in vivo*. Table 3.3 summarizes the experimental results of P-nucleotide additions. The proportions of P-addition range from 17.54% in WT hTdTS

transfection to 25% in control transfection. The average number of P-additions per recombined joint range from 1.38 to 1.8 nucleotides. No significant differences were observed between A445T and R431C hTdTS transfections when compared to WT hTdTS. However, average number of P-additions observed in L397S hTdTS transfection is significantly different, as compared to control transfection (p-value 0.0487).

Table 3.3: Recombined substrate pGG51 plasmid P-nucleotide additions.

	Control	WT hTdTS	A445T hTdTS	L397S hTdTS	R431C hTdTS
Total number of unique sequences obtained	48	57	65	72	34
Total number of P-additions	19	18	25	18	12
Number of sequences containing P-additions	12	10	16	13	8
Proportion of sequences containing P-additions (%) *	25	17.54	24.62	18.06	23.53
Average number of P-additions *	1.58	1.8	1.56	1.38	1.5
P-value		0.2996**	0.2319	**0.0487**	0.2011

* Proportion of sequences containing P-additions (%) = (number of sequences containing P-additions / total number of unique sequences) * 100%
Average number of P-additions = total number of P-additions / number of sequences containing P-additions
** P-value for P-additions for WT hTdTS experimental condition computed using comparison to the control condition. All other p-values computed through comparison to WT hTdTS experimental condition.
*** The statistical analysis was performed using a standard T-test. P-values shown in bold are <0.05, thus found to be statistically significant.

N-nucleotide additions

The addition of random non-templated nucleotides (N-additions) at the recombined joints is catalyzed by TdT. Therefore, the average number and type of N-additions observed at the joints provides a direct measure to the activity of the polymerase. The analysis of random nucleotide additions (N-additions) is summarized in Table 3.4 and all computed statistical p-values are shown in Figure 3.20. Average number of N-additions observed in control transfection is 1 nucleotide (only 3 out of 48 sequences contained a single N-nucleotide). The WT hTdTS transfection resulted in the highest average number of N-additions as compared to other three hTdTS variants (average of 3 nucleotides per joint). Thus, the number of N-additions was different for WT hTdTS transfection as compared to control experiment (p-value 0.0336). Another significant difference was observed in R431C hTdTS transfection. Only 1.38 N-nucleotides were observed on average in that transfeciton compared to 3 N-nucleotides in WT hTdTS transfection (p-value 0.0062).

The proportion of sequences containing N-additions was also computed (Table 3.4). For control transfection only 6.25% of all recovered recombined sequences contained N-additions, compared to 73.68% of sequences recovered from WT hTdTS transfection (p-value <0.0001). A445T hTdTS transfection did not result in a differenent frequency of N-containing joints, as compared to WT hTdTS transfection (78.46% versus 73.68%). However, both L397S and R431C hTdTS transfections generated a significantly lower number of N-containing joints as compared to WT hTdTS transfection (about 23% versus

73.68%, p-values <0.0001). Therefore, there is over a 3 fold difference among these frequencies of N-containing recombined sequences.

The recombined pGG51 sequences containing N-additions were further investigated in terms of their AT versus GC base content (Table 3.4). In WT hTdTS transfection, over 60% of N-nucleotides were G or C bases. A445T hTdTS transfeciton also shows a clear preference for GC content over AT of N-additions (72% GC content). L397S hTdTS transfection 47% of N-additions were GC bases, which was determined not to be significant when compared to WT hTdTS transfection (p-value 0.1691). The major difference was observed in R431C hTdT transfection. Only 27% of all the recovered N-containing joints contained GC bases. This indicated a significant shift towards the preference for AT bases, compared to WT hTdTS transfection (p-value 0.0164).
Note; the statistical analysis performed using numbers of bases counted and not percentages.

The distribution of N-nucleotide additions per single recombined joint was characterized (Figure 3.21). The frequencies of N-nucleotide occurrence were computed as a ratio of the number of joints having a certain number of N-nucleotides, for instance 2 N-nucleotide additions per joint, to the number of joints containing N-additions. WT hTdTS transfection resulted in even distribution between 2, 3 and 4 N-nucleotides per joint. Lower frequencies were observed for longer N-nucleotides (for example 5, 6 and 8 N-nucleotides per joint). A445T hTdTS transfection showed an even distribution for 1-4 N-nucleotides per joint. There were no N-additions longer than 4 nucleotides. L397S hTdT transfection showed a preference for 1 or 2 N-nucleotide additions per joint (over

30% of joints had 1 N-nucleotide and over 40% of joints contained 2 N-nucleotides per joint). A clear preference was observed in R431C hTdTS transfection in which about 75% of all N-containing joints were in fact 1 N-nucleotide additions.

Table 3.4: Recombined substrate pGG51 plasmid N-nucleotide additions, their frequency and AT versus GC base content.

	Control	WT hTdTS	A445T hTdTS	L397S hTdTS	R431C hTdTS
Total number of unique sequences obtained	48	57	65	72	34
Number of N-additions					
Total number of N-additions	3	126	122	34	11
Number of sequences containing N-additions	3	42	51	17	8
Average number of N-additions *	1	3	2.39	2	1.38
P-value		**0.0336****	**0.0335**	**0.0218**	**0.0062**
Frequency of N-additions					
Number of sequences containing N-additions	3	42	51	17	8
Number of sequences not containing N-additions	45	15	14	55	26
Proportion of sequences containing N-additions (%) *	6.25	73.68	78.46	23.61	23.53
P-value		**<0.0001****	0.6704	**<0.0001**	**<0.0001**
Frequency of AT versus GC bases in N-additions					
Number of GC bases in N-additions (percentage of total N-additions) *	3 (100%)	77 (61.11%)	88 (72.13%)	16 (47.06%)	3 (27.27%)
Number of AT bases in N-additions (percentage of total N-additions) *	0	49 (38.89%)	34 (27.87%)	18 (52.94%)	8 (72.73%)
P-value		0.2913**	0.0814	0.1691	**0.0164**

* Average number of N-additions = total number of N-additions / number of sequences containing N-additions. Proportion of sequences containing N-additions (%) = (number of sequences containing N-additions / total number of unique sequences) * 100%
Percentage of AT or GC bases in N-additions (%) = (total number of AT or GC bases / total number of N-additions) * 100%

** P-value for N-additions for WT hTdTS experimental condition computed using comparison to the control condition. All other p-values computed through comparison to WT hTdTS experimental condition.

*** The statistical analysis was performed using a standard T-test for the number of N-additions. A chi-square test was used to analyze frequency of N-additions and frequency of AT versus GC bases in N-additions. P-values shown in bold are <0.05, thus found to be statistically significant.

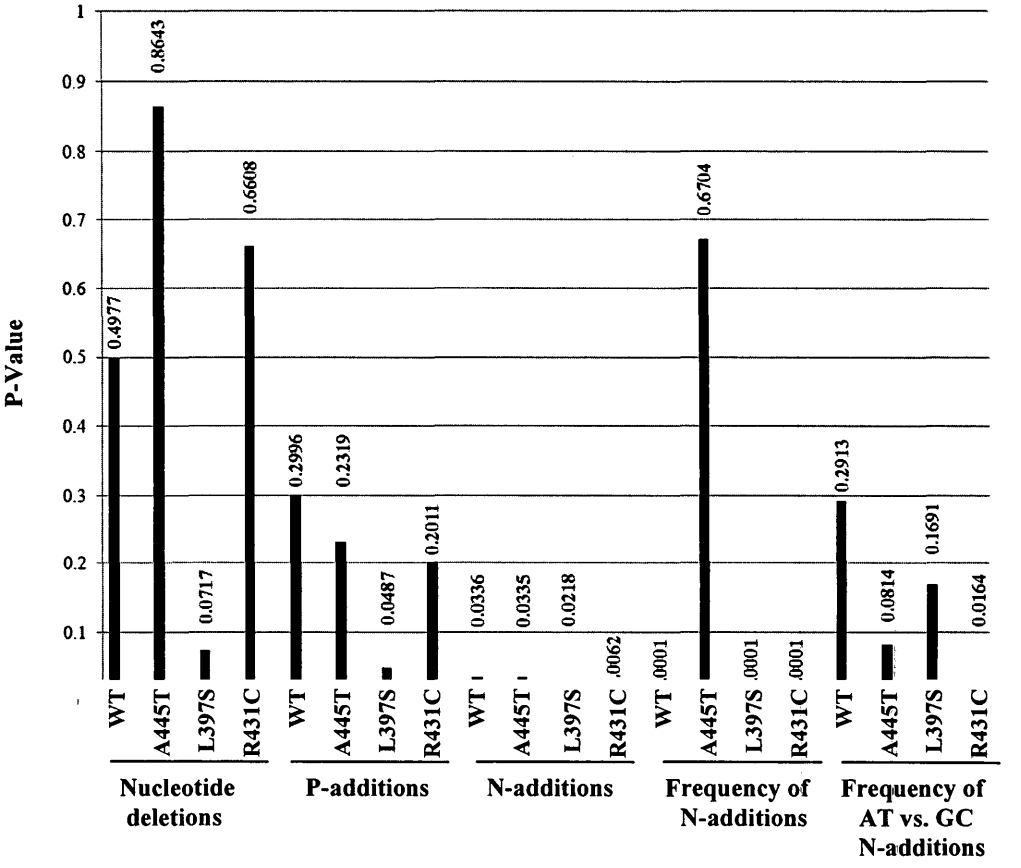


Figure 3.20: P-values obtained from standard T-test or chi-square test. The p-values, labeled above each bar, are subdivided to few categories; nucleotide deletions, P-additions, N- additions, frequency of N-additions and frequency of AT versus GC bases in N-additions. The p-values were acquired from Table 3.4.

* P-value for N-additions for WT hTdTS experimental condition computed using comparison to the control condition. All other p-values computed through comparison to WT hTdTS experimental condition. P-values below 0.05 were found to be statistically significant.

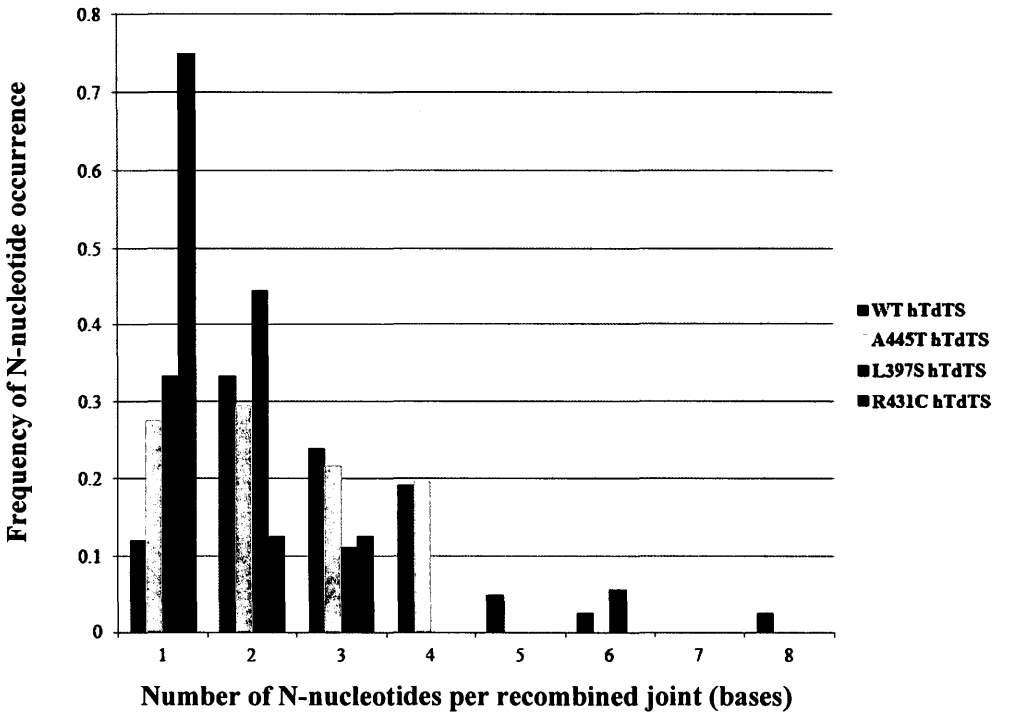


Figure 3.21: Distribution of N-nucleotide additions per single recombined joint.
The frequencies of N-nucleotide occurrence were computed as a ratio of the number of joints having a certain number of N-nucleotides to the number of joints containing N-additions.

Length of arbitrary complementary-determining region (CDR)

The length of arbitrary complementary-determining region (CDR) was chosen to include 15 bases flanking the RSS-region of recombination substrate pGG51 plasmid on the 5' and the 3' end. Therefore, the length of a non-recombined germline arbitrary CDR region is 30 bases. Figure 3.22 shows the mean length of arbitrary CDR for every experimental condition tested. The mean length was computed by summing up all the lengths of arbitrary CDRs per experimental condition (include nucleotide deletions and additions) and dividing those by the total number of unique sequences. WT hTdTS

transfection generated significantly longer arbitrary CDRs as compared to control transfection (28 versus 25.67 bases, p-value <0.0001). In addition, L397S and R431C hTdTS transfections resulted in lower arbitrary CDRs as compared to WT hTdTS transfection (25.56 and 26.32 bases versus 28 bases, p-values <0.0001).

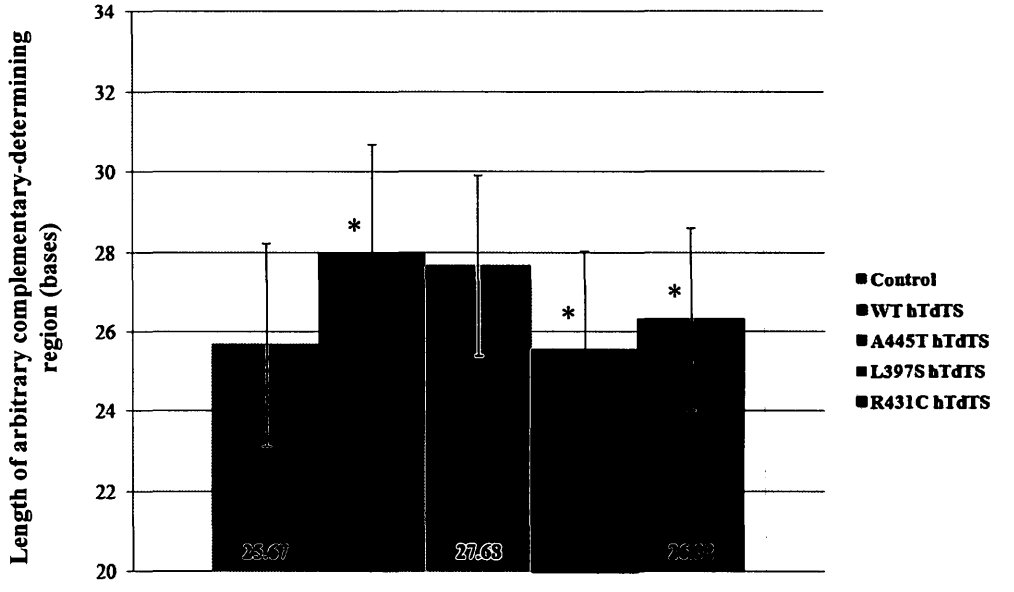


Figure 3.22: Mean length of arbitrary complementary-determining region.
The mean values are shown at the bottom of the bars. The error bars represent the standard deviation.
The mean length of CDRs are as follows;
Control= 25.67 ± 2.55 bases
WT hTdTS= 28 ± 2.69 bases (p-value **<0.0001**)*
A445T hTdTS= 27.68 ± 2.26 bases (p-value 0.4719)
L397S hTdTS= 25.56 ± 2.49 bases (p-value **<0.0001**) *
R431C hTdTS= 26.63 ± 2.28 bases (p-value **<0.0001**) *

* P-value for mean length of CDR for WT hTdTS experimental condition computed using comparison to the control condition. All other p-values computed through comparison to WT hTdTS experimental condition. The statistical analysis was performed using a standard T-test. P-values shown in bold are <0.05, thus found to be statistically significant.

Chapter 4: Discussion

V(D)J recombination is an essential process in the generation of the wide range of diverse antigen receptors needed to recognize pathogens and activate the adaptive immune response to fight infection. The process requires a number of protein players, including TdT. TdT is a template-independent DNA polymerase responsible for the addition of random nucleotides to B and T cell receptors during V(D)J recombination. In its role, TdT contributes to junctional diversity of antigen receptors.

I identified a number of naturally occurring polymorphic variants of the hTdT short isoform that potentially may alter TdT's polymerase activity. The criteria for selection were based on amino acid residue conservation among species, the chemical nature of the amino acid change and the position of the SNP relative to TdT's catalytic site. The activity differences of selected hTdTS SNPs compared to wild type hTdTS were tested using an *in vitro* polymerase assays and substrate preferences. Specifically, dNTPs usage and DNA substrate primer preferences were examined. The activity of hTdTS variants was also evaluated using an *in vivo* extrachromosomal V(D)J recombination assay. The implications for the development of a broad antigen receptor repertoire in homozygous individuals with hTdTS SNPs are elaborated in this chapter.

4.1 Selection of hTdTS SNPs

The NCBI SNP database was used to identify a total of six non-synonymous hTdTS SNPs that, by my criteria, were predicted to be the best candidates to have altered polymerase activity compared to wild type hTdTS (Table 3.1). The active site of hTdT contains three highly conserved aspartic acid residues (D343, D345 and D434) and a

cation binding site that aids in stabilizing the negative charge associated with a triphosphate group of an incoming dNTP (Figure 3.2). As described in Chapter 1, hTdTS structure has not been solved to date. It would be beneficial to investigate the structure-function relationship of the identified SNPs utilizing the solved structure of the polymerase rather than that of murine TdT.

The threonine to serine amino acid change at position 450 does not result in an amino acid chemical property change since both residues are polar (they contain an hydroxyl group -OH). Therefore, the new residue, serine, is expected to maintain polar contacts in the same manner as the original amino acid residue, threonine. Threonine contains an extra methyl side chain group ($-CH_3$) as compared to serine, thus this SNP results in a substitution for a smaller residue with no change to the chemical nature of the residue. An amino acid change from a big to relatively small molecule is not expected to create steric hindrance at the structural protein level. However, T450 is located within one of two incoming dNTP-binding sites as identified by Delarue *et. al.* (5). The dNTP-binding site conserved consensus sequence ALLGWTGSR (residues 445-453) is part of a hinge region first identified in polymerase β that is involved in open-closed conformation transition of the enzyme (5, 15) (Figure 4.1). A445T SNP is also found within the incoming dNTP-binding sites (ALLGWTGSR, residues 445-453). This SNP substitutes a small non-polar alanine residue for a bigger polar threonine amino acid residue. This mutation might create new polar interactions with surrounding residues that were otherwise absent.

A second SNP (R460Q), is also in close proximity to the active site of TdT. This substitution results in a change of a basic amino acid residue to a polar uncharged residue. This substitution may disrupt original side chain interactions with surrounding amino acid residues, thus potentially affecting the microenvironment in the vicinity of this SNP.

Arginine at position 431 (R431) participates in a salt bridge between D343 and cobalt cation (Co^{2+}). Specifically, positively charged amine group of arginine non-covalently interacts with the negatively charged oxygen atom of carboxyl group (oxygen at position δ2) of aspartate, while positively charged cobalt cation interacts with negatively charged oxygen (position δ1) of the same catalytic aspartate residue. The amine group of R431 also interacts with negatively charged oxygen of D434 (oxygen at position δ1) (Figure 4.1). Thus, the salt bridges created contribute to the overall stability of TdT at the active site. Loss of positive charge at R431 will therefore disrupt this salt bridge, as in the case of R431C variant, rendering TdT's catalytic site less stable.

L397S SNP is located at the unstructured lariant-like Loop1 region (ESTFEQPSRVKDALDH, residues 382-401) (Figure 4.3) that has been shown to associate with template independent polymerase activity. Romain *et. al.* demonstrated that a 13 residue deletion of Loop1, including L397, managed to switch the template independent activity of TdT to template dependent activity comparable to polymerase λ that lacks the loop region (17). Thus, it is though that TdT's Loop1 region sterically excludes double stranded template DNA from the active site and promotes template independent polymerase activity.

Lastly, D280H SNP is located at the periphery of TdT, away from the active site. Thus, it served as an experimental control that was hypothesized not to affect the polymerase activity.

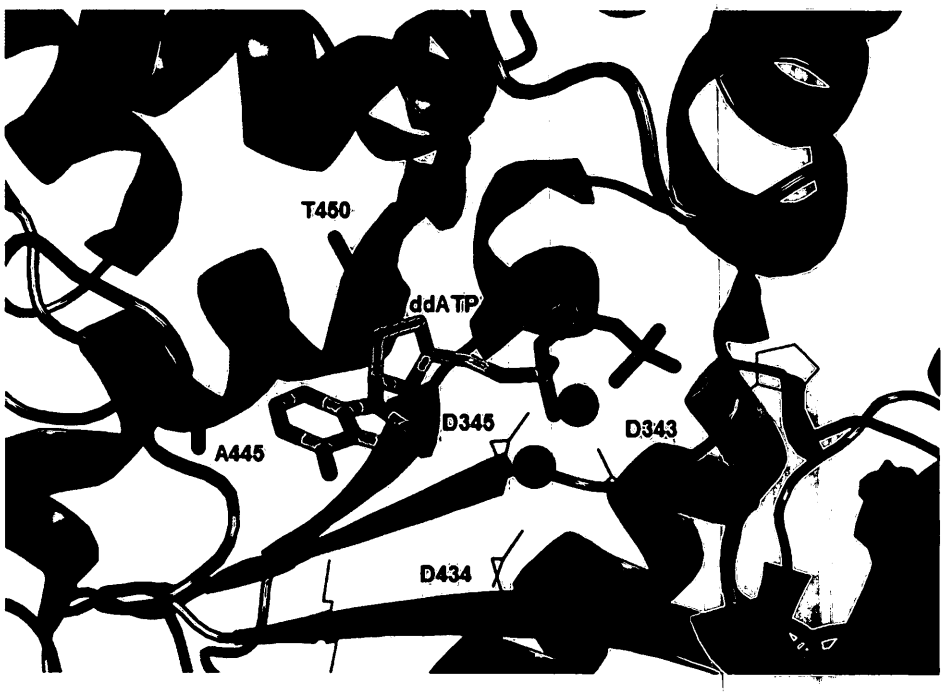


Figure 4.1: Crystal structure of murine TdT (PDB ID: 1KEJ) and location of A445 and T450 residues within incoming dNTP-binding site. Alpha helical structure highlighted in green represents incoming dNTP-binding site (ALLGWTGSR, residues 445-453). The locations of A445 and T450 are shown in magenta. The second incoming dNTP-binding site is highlighted in red (TGGFRRG, residues 331-337). Three conserved aspartic acid residues are highlighted in blue (D343, D345 and D434). The incoming nucleotide substrate, ddATP, is shown in yellow. Two cobalt ions are colored cyan. Figure generated using *PyMol* software.

Figure 4.2: Crystal structure of murine TdT (PDB ID: 1KEJ) and salt bridge interactions of R431 with catalytic aspartate residues. Three conserved aspartic acid residues are highlighted in blue (D343, D345 and D434). Salt bridge interactions between R431 amine group and carboxyl group of D434 and D343 are shown in dashed lines (atomic distances of 3.0Å and 3.2Å are labeled). The incoming nucleotide substrate, ddATP, is shown in yellow. Two cobalt ions are colored cyan. Figure generated using *PyMol* software.

Figure 4.3: Crystal structure of murine TdT in complex with ddATP (PDB ID: 1KEJ) is superimposed onto the murine TdT structure in complex with primer single stranded DNA (PDB ID: 1KDH). Unstructured Loop1 is shown in yellow (ESTFEQPSRVKDALDH, residues 382-401) and L397 position is highlighted in magenta. Three conserved aspartic acid residues are highlighted in blue (D343, D345 and D434). The incoming nucleotide substrate, ddATP, is shown in yellow. Two cobalt ions are colored cyan. The primer single stranded DNA represents three nucleotides at the active site, specifically 5-bromo-2'-deoxyuridine-5'-monophosphate, shown in green. Figure generated using *PyMol* software.

4.2 hTdTS SNPs demonstrate varied polymerization activities *in vitro*

Following the identification of six hTdTS SNP candidates, *in vitro* polymerase activity assays were conducted utilizing purified bacterial His-tagged hTdTS protein variants. Figure 3.10 depicts the results of the time course polymerization assay comparing WT hTdTS activity versus all six hTdTS SNPs. Five out of six hTdTS SNPs are active in the biochemical assay, while R431C hTdTS is a dead mutant under the assayed conditions. The polymerase activity of all assayed hTdTS variants varied based on the maximum length of oligo products produced and the banding pattern of the products. The graph in Figure 3.11 plots the change in maximum size of hTdTS oligo product versus time of reaction. D280H and T450S polymerization curve overlaps that of WT hTdTS, especially within the first 3 minutes of reaction. As detailed earlier, neither D280H nor T450S were expected to alter hTdTS activity. D280H SNP is located outside the catalytic core, at the periphery of the enzyme, and T450S because the change did not result in a notable amino acid change since both the original and the changed residues are polar. The banding pattern of D280H and T450S was similar to that of WT protein, meaning that most of the DNA signal concentrated at the top of the smear. The broadness or tightness of bands demonstrates the heterogeneity in the lengths of the oligo products. This heterogeneity may be interpreted as differences in processivity of the polymerase, meaning the ability to carry out continuous nucleotide polymerization per single association/dissociation event. A broad banding pattern suggests a great heterogeneity of oligo product lengths and frequent association/dissociation events of TdT from the DNA primer. On the other hand, tight banding pattern implies less heterogeneity of oligo

lengths and thus a processive mode of polymerization. Figure 3.10 illustrates a tight banding pattern for WT hTdTS, D280H and T450S SNPs (observed in three independent assays), suggesting the processive mechanism.

It is important to note that the kinetic mechanism of polymerization by TdT has not been conclusively established (1, 2). In 2001, Boulé *et. al.* concluded that it is highly probable that DNA synthesis catalyzed by TdT proceeds in a strictly distributive fashion, meaning that the oligo product dissociates from TdT following a single nucleotide addition (74). They observed that the size of TdT oligo product decreased as lower TdT concentrations were used in the *in vitro* reaction (all reactions carried out under excess primer $(dA)_{10}$ initiator and dATP substrate conditions) (74). The distributive polymerization mode has not been agreed upon by researchers in the field.

The initial polymerization rate of R460Q SNP is slower as compared to WT protein, however the most significant difference was observed for A445T and L397S SNPs. In fact, A445T initial rate is 3 fold lower and L397S initial rate is 5 fold lower as compared to WT hTdTS (comparing the slopes of polymerization curves in Figure 3.11). Therefore, A445T and L397S variants are least active out of six SNPs tested. In addition, broad banding pattern is observed for these two variants and is distinctive from the tight banding pattern of WT hTdTS oligo products. Based on these results I conclude that the polymerase mechanism of A445T and L397T appears non-processive under the assayed *in vitro* conditions.

4.3 hTdTS activity differs according to dCTPs and dGTPs usage

A single type of free deoxynucleotide triphosphate was added to the *in vitro* polymerization reaction to observe whether the hTdTS SNPs have a substrate preference in utalizing substrate dNTPs (Figure 3.12). The dNTP mix (all four nucleotides) always resulted in longest polymerization products by all assayed hTdTS variants. With each SNP the dNTP supplement always resulted in the most active polymerase reactions. Interestingly, usage of dCTPs or of dGTPs in the reactions resulted in the greatest variation among all hTdTS variants when compared to WT hTdTS usage of dNTP, dATP and dTTP. Specifically, dCTP polymerization reactions resulted in shorter oligo products compared to the starting Cy3-labelled oligo by all assayed hTdTS variants. This phenomenon was observed in repeated assays and may be attributed to a 3' to 5' exonuclease activity by the polymerase. In spite of this observation, *in vitro* 3' to 5' exonuclease activity has only been shown to associate with long isoforms of hTdT (hTdTL1 and hTdTL2) (12). Thus, this result is the first report of potential hTdTS *in vitro* 3' to 5' exonculease activity and this activity is only seen in dCTPs supplemented reactions. Due to the fact that the short isoform of hTdT is primarily responsible for nucleotide additions at coding joints, it is difficult to speculate the origin of this phenomenon. Further experiments are required to understand these results. For example, performing *in vitro* polymerization reactions with purified hTdT long isoforms in presence of dCTPs and comparing oligo products to hTdT short isoform reactions would be informative.

Polymerization reactions supplemented with dGTPs result in a predominant band of about 50 bases long in all assayed hTdTS variants (Figure 3.12). Even extending the reaction times of WT hTdTS in presence of dGTPs did not allow polymerization beyond 75 base oligo products (Figure 3.13). In order to investigate whether this phenomenon is DNA sequence dependent, a complementary nucleotide sequence to the Cy3-labelled oligo DNA primer was supplemented in the reaction (Figure 3.15). The usage of either of the Cy3-labelled oligo sequences resulted in the production of aborted products in the presence of dGTPs. Thus, the assay ruled out my initial hypothesis that the aborted dGTP reaction product was caused by the sequence of the DNA primer. It is important to note that *in vivo* studies have shown a clear preference for the incorporation of dGTPs and dCTPs by TdT versus dATPs and dTTPs (1, 2). Lieber *et. al.* have demonstrated that as much as two thirds of base composition is G+C at coding joints isolated from V(D)J recombination products and only about one third is A+T (75). A plausible explanation for the clear G+C preference by TdT has not been established. However, favorable base stacking interactions between purine-purine and pyrimidine-pyrimidine bases were suggested to explain the observed trend of TdT to add purines or pyrimidines in succession to promote DNA stability (71).

4.4 Usage of double stranded 3' overhang DNA substrate primer is preferred by hTdTS

Three different DNA substrate primers were used in the *in vitro* polymerase reaction; single stranded DNA, blunt ended double stranded DNA and double stranded DNA with a 3' nucleotide overhang. Five of the assayed hTdTS variants (R431C, the "dead" mutant, remained inactive) were able to polymerize all three DNA substrates

(Figure 3.14). This result supports the mechanism of hairpin opening process *in vivo* during V(D)J recombination. TdT requires a free 3' hydroxyl end group of a DNA primer in order to catalyze the formation of a new phosphodiester bond with an incoming dNTP. Off-center hairpin opening of double stranded DNA results in generation of short nucleotide overhangs, known as P nucleotides. A blunt ended DNA results if the hairpin opening occurs at the tip. Thus, both types of double stranded DNA primers may be utilized by TdT in a lymphoid cell.

A clear preference for double stranded 3' overhang DNA primer by all hTdTS variants is demonstrated in Figure 3.14. The blunt end double stranded DNA primer is the least preferred substrate of hTdTS. Therefore, 3' overhang DNA primer is the more efficient substrate for TdT within the *in vitro* setting. Overall, the data lead to the conclusion that there are no marked difference between hTdTS SNPs and WT hTdTS to polymerize the three DNA primer substrates. It is important to note that *in vivo* TdT may only add up to 15 nucleotides at the coding joint between recombined gene segments in order to properly maintain the shape and size of newly formed antigen receptor. Excess nucleotide additions between V(D)J segments would result in a misfolded and non-functional receptor that would be degraded in the cell. A fact that has found a viable and advantageous market for biotech companies marketing TdT. On the other hand, the *in vitro* polymerization reaction is able to promote hundreds of nucleotide additions in the presence of excess amounts of substrate dNTPs. Therefore, although the *in vitro* activity does not reflect the *in vivo* ability of hTdT to carry its function the assay does reveal the fundamental polymerase activity of the enzyme.

4.5 Cobalt versus magnesium cofactor use

TdT has been shown to use a variety of divalent cations for *in vitro* polymerization reactions, including cobalt, magnesium, zinc and manganese. At physiological conditions it is thought that magnesium is the primary divalent cation promoting TdT's activity (1, 2). Based on these considerations the polymerization activity in the presence of cobalt or magnesium was investigated (Figure 3.16). Wild type hTdTS and R431C variant reactions are shown. R431C hTdTS remained a dead mutant, as shown previously, in presence of either cofactors. However, smears were observed below the control Cy3-labelled oligo band in the absence of cobalt or the presence of magnesium that have not been seen previously. As detailed earlier, this activity may be interpreted at 3' to 5' exonuclease activity that is related to the cobalt cofactor in the reaction. These smears were not present in the WT hTdTS reactions, suggesting that this observation is unique to the inactive R431C variant. Structurally, R431C mutant was implicated in coordination of the negative charge associated with triphosphate group of incoming dNTP (Figure 4.2). Thus, the effect of R431C mutant to drive depolymerization reaction in absence of cobalt remains unknown.

No marked difference was observed between the sizes of WT hTdTS oligo products under the different conditions (Figure 3.16). The banding pattern differs when comparing WT hTdTS reactions in presence versus absence of cobalt cofactor. Broad smears reveal great heterogeneity in lengths of oligo products, such as those reactions in absence of cobalt or magnesium supplemented reactions. Tight smears imply less heterogeneity of reaction products in presence of cobalt. These observations may point to

greater efficiency of WT hTdTS to utilize cobalt versus magnesium in the *in vitro* polymerization reactions. Differences in metal utilization have been investigated in the past. In fact, cobalt was shown to increase the catalytic polymerization efficiency of pyrimidines, while magnesium facilitates preferential addition of purines (76-78).

4.6 Examining R-values in V(D)J recombination assay

Following *in vitro* polymerization assays of six hTdTS SNP candidates, V(D)J recombination assay was employed. The assay models RAG-mediated V(D)J recombination process on a substrate plasmid and the ability of TdT to contribute to junctional diversity (Figures 2.4, 2.5, 2.6). Three hTdTS SNPs were investigated, namely A445T, L397S and R431C, since they demonstrated significant functional differences as compared to WT hTdTS in *in vitro* polymerization assays. Recombination efficiency was examined through calculating R-values from *in vivo* recombination assay. R-value represents the ratio of recombined substrate plasmid over all recombination substrate plasmids, which indirectly represents the efficiency of recombination during transfection. As R-values are ratios, the efficiencies from multiple transfections can be compared (Appendix C). Figure 3.19 plots the calculated R-values from of eight independent HEK293T cell transfections. No marked difference in mean R-values was observed between WT hTdTS, A445T and R431C variants. The mean R-value for L397S variant was the 1.5 fold higher than that of WT hTdTS, suggesting highest recombination efficiency of these transfections. In addition, L397S variant generated the highest number of ampicillin and chloramphenicol resistant colonies in repeated transfections, which supports favorable recombination. Recombination efficiency has been shown to correlate

to changes in RSS DNA sequence (25, 27), such that RAG proteins showed preference to utilize certain RSSs more than others. As TdT is known to catalyze nucleotide additions to already cleaved double stranded DNA breaks *in vivo*, its role in recombination efficiency has not been established to date. Thus, the direct effect of L397S mutation of TdT is difficult to point. In addition, control transfections resulted in a 2 fold higher mean R-value compared to WT hTdTS transfections. The presence of TdT decreases recombination efficiency, suggesting that TdT may interfere with initiation of V(D)J recombination process in the cell.

4.7 Analysis of recombined joints

In vivo recombination assay included co-transfection of recombination substrate plasmid, RAG proteins and hTdTS variants into HEK293T cells lacking endogenous TdT expression. The recombined substrate plasmids were harvested from each independent transfection and the junction region was sequenced (Appendix D) to examine nucleotide deletions, P-additions and N-additions. Statistical analysis was performed on the generated data. First, on average 4-5 nucleotide deletions were observed for all experimental conditions and no significant difference was noted among them (Table 3.2). Thus, the presence of TdT does not affect nucleotide deletions at recombined joints. This conclusion agrees with the primary function of hTdTS to add random nucleotides and not to promote nucleotide deletions (although hTdT long isoforms may catalyze nucleotide deletions due to 3' to 5' exonuclease activity). Second, an average of 1.5 P-additions was observed per recombined joint for all experimental conditions (Table 3.3). These two

parameters served as internal controls for the recombination assay and were found to be independent of hTdTS.

The number of random N-additions was analyzed among the three hTdTS variants and compared to WT hTdTS transfections. This analysis provides a measure for TdT's polymerase activity and allows for comparison among polymorphic forms of the enzyme. WT hTdTS demonstrated the greatest activity through the addition of 3 nucleotides on average per recombined joint (Table 3.4). Three of the 48 recovered joints from control transfections (lack TdT expression) each contained one N-addition. This type of activity can be attributed to the action of polymerase μ that possess both template dependent and template independent polymerase activity. In fact, TdT knockout mice were shown to have template independent additions at nearly 5% of their V(D)J junctions. It is thought that random additions are beneficial for annealing of the coding ends during non-homologous end jointing due to 1 or 2 nucleotides of terminal microhomology (45, 52, 55).

A445T hTdTS demonstrated *in vivo* polymerase activity comparable to WT hTdTS. The recombination assay revealed no significant difference in the number of N-containing joints compared to WT hTdTS (78 versus 73%, Table 3.4). A445T recombined joint were GC-rich, as were the WT hTdTS joints. An even distribution was observed between the occurrences of 1-4 N-additions per A445T recombined joint (Figure 3.21). Finally, the mean length of arbitrary CDR region in the presence of A445T was not significantly different from that of WT hTdTS (27.68 versus 28 bases, p-value 0.4719, Figure 3.22). Therefore, I conclude that A445T SNP does not to alter hTdTS

polymerase activity *in vivo* despite its location within the incoming dNTP-binding site of the polymerase.

The proportion of N-containing recombined joints revealed a significant difference for L397S SNP. About 73% of WT hTdTS joints contained N-additions, versus only 23% of L397S hTdTS (p-value <0.0001) (Table 3.4). Figure 3.22 depicts the lengths of arbitrary CDR regions of recombined joints. The L397S variant generated significantly shorter CDR regions compared to WT hTdTS (25.56 versus 28 bases), suggesting less overall additions of non-templated nucleotide and subsequent diversity at the junction region. As detailed earlier, L397S position is located within unstructured Loop1 region that is thought to prevent double stranded DNA primer substrate from diffusing into the active site of TdT (Figure 4.3). It is possible that non-polar leucine to polar serine amino acid change forms new intra-molecular polar contacts within the Loop1 region, rendering it less flexible. This in turn may decrease the efficiency of DNA primer substrate diffusion into the active site, thus resulting in lower frequency of N-containing joints.

R431C variant that was found to be inactive *in vitro*, however demonstrated N-addition ability in the *in vivo* recombination assay. However, R431C was the least active polymerase adding an average of only 1.38 nucleotides per joint. Only about 23% of R431C recombined joints contained N-additions versus 73% of WT joints (p-value <0.0001) (Table 3.4). Most of R431C joints had a single nucleotide addition compared to WT hTdTS joints (Figure 3.21). In fact, over 70% of the R431C joints that contained N additions, contained 1 N-addition while WT hTdTS demonstrated somewhat even

distribution between 2, 3 and 4 N-additions per joint. Referring back to Figure 4.2, it is plausible that an essential salt bridge between amine group of arginine and catalytic aspartate residues (namely D343 and D434) was disrupted, causing loss of stability at the active site. *In vitro*, this mutation turned out to be deleterious for the polymerase, while *in vivo* it resulted in the least active polymerase out of all assayed. The amino acid change did not eliminate polymerase activity all together, however. Potentially, the R431C variant was having difficulty initiating the polymerase reaction in terms of coordination of cations in addition to the polymerization reaction since significantly short N-additions were observed. In addition, significantly shorter arbitrary CDR region were generated in R431C experiments versus WT hTdTS (26.32 versus 28 bases, p-value <0.0001) (Figure 3.22). The type of nucleotide added by hTdTS was also investigated in terms of AT versus GC content. As detailed earlier, TdT shows preferential addition of GC bases to V(D)J junctions *in vivo* (1, 2). Indeed, WT hTdTS recombined joints contained mostly GC bases (77%, Table 3.4). Surprisingly, the abundance of AT bases in R431C recombined joints was greater than GC bases (about 73% AT versus 27% GC bases). This result demonstrated a clear switch in terms of dNTP preference by R431C. The specific SNP is not located in the dNTP binding site of hTdTS, thus it is difficult to speculate the reason behind the switch. However, it was clearly demonstrated that R431C SNP has a profound effect on polymerase activity, both *in vitro* and *in vivo*.

4.8 Implications of hTdTS SNPs on adaptive immune diversity

Both L397S and R431C SNPs of hTdTS demonstrated low polymerase activities compared to WT hTdTS *in vitro* and *in vivo* assays. Decreased polymerase activity in turn affects junctional diversity at V(D)J junctions of antigen receptors of B and T cells. As TdT is partially responsible for antigen receptor diversity allowing recognition of broad pathogens by the adaptive immunity, its diminished activity may impair the development of a diverse antigen receptor repertoire. Failure to recognize invading pathogens lowers the chances of the immune system to promote an adequate response to fight the infection. Thus, it is plausible to hypothesize that otherwise healthy individuals homozygous for these hTdTS SNPs may encounter difficulty in mounting an adaptive immune response due to the lack of a specific high affinity antigen receptors. This may promote susceptibility for pathogens and subsequently increase the risk of infection.

The specific role of TdT in autoimmune models remains controversial. Absence of TdT has been shown to correlate with decreased severity of autoimmune diseases, such as lupus and diabetes (57-59). In fact, TdT-deficiency has been linked to decreased number of autoantibody producing cells. This deficiency may provide a protective role against autoimmunity due to lower polyreactivity towards self-antigens and reduction in autoantibody affinities. Researchers speculate that this autoimmune protection is the reason for the delayed onset of TdT activity during fetal development (57-60). The junctional diversity brought on by TdT is thought to produce pathogenic high affinity lymphocytes that protect individuals from foreign pathogens. Thus, autoimmune-prone individuals homozygous for the investigated SNPs (L397S and R431C) may experience

symptom relief associated with autoimmunity and potentially control the disease. It is likely that TdT represents one form of activity involved in autoimmunity, as TdT generates B and T cell receptors that are not-self. Other factors influencing onset of autoimmunity and its progression remain unknown. However, TdT suppression represents a new therapeutic avenue to manage autoimmune-related disorders.

It is now evident that single nucleotide polymorphisms, present in all individuals, contribute to individual phenotypes, susceptibility to disease or even cause a disease. Studies focused on understanding genetic diversity among individuals, such as this, will provide information on the variation seen within a population as well as allow researchers to explore personalized medicine to tailor individual genetic backgrounds.

Conclusion

In this thesis, the *in vitro* and *in vivo* functional polymerase activities of few single nucleotide polymorphisms of hTdTS were investigated. The aim was to characterize the possible effects of SNPs on the diversity of V(D)J joints during recombination process of B and T cell receptors. A445T and L397S variants were observed to affect TdT polymerization reactions *in vitro*, while R431C SNP demonstrated no polymerase activity whatsoever. Analysis of the *in vivo* activity was done via extrachromosomal V(D)J recombination assay. L397S SNP resulted in highest recombination efficiency out the three assayed variants. In addition, L397S substitution resulted in low *in vivo* polymerase activity that may have been caused by loss of flexibility of the unstructured Loop1 region. *In vivo* R431C hTdTS demonstrated lowest activity in terms of N-additions. The proportion of N-containing joints and the number of nucleotide additions per recombined joint were significantly lower compared to WT enzyme. This may be attributed to disruption of an essential salt bridges caused by the amino acid substitution within the catalytic site of the polymerase. It is evident that hTdTS SNPs revealed variable effects on both the *in vitro* and the *in vivo* TdT activities. Thus, homozygous individual for these SNPs may alter the V(D)J junctional diversity at B and T cell receptor and subsequently affect the adaptive immune response. Evidence of TdT's connection to autoimmunity has sparked an interest to investigate the role of the polymerase and its therapeutic potential in immune-related diseases.

References

1. Fowler, J. D., and Suo, Z. (2005) Biochemical, Structural, and Physiological Characterization of Terminal Deoxynucleotidyl Transferase. Chem. Rev. **106**(6): 2092-2110.
2. Motea, E. A., and Berdis, A. J. (2010) Terminal deoxynucleotidyl transferase: The story of a misguided DNA polymerase. Biochemica et Biophysica Acta- Proteins & Proteomics. **1804**(5):1151-1166.
3. Bollum, F. J. (1960) Calf thymus polymerase. J. Biol. Chem. **235**:2399–2403.
4. Chang, L. M., and Bollum, F. J. (1971) Deoxynucleotide-polymerizing enzymes of calf thymus gland. V. Homogeneous terminal deoxynucleotidyl transferase. J. Biol. Chem. **246**(4):909–916.
5. Delarue, M., Boule, J. B., Lescar, J., Expert-Bezancon, N., Jourdan, N., Sukumar, N., Rougeon, F. and Papanicolaou, C. (2002) Crystal Structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase. EMBO J. **21**(3):427-439.
6. Lewis, S. M. (1994). The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. Adv. Immunol. **56**:27-150.
7. Li, Y. S., Hayakawa, K., and Hardy, R. R. (1993) The regulated expression of B lineage associated genes during B cell differentiation in bone marrow and fetal liver. J. Exp. Med. **178**(3):951-960.
8. Bogue, M., Gilfillan, S., Benoist, C., Mathis, D. (1992) Regulation of N-region diversity in antigen receptors through thymocyte differentiation and thymus ontogeny. Proc. Natl. Acad. Sci. USA. **89**(22):11011-11015.
9. Peralta-Zaragoza, O., Recillas-Targa, F., and Madrid-Marina, V. (2004) Terminal deoxynucleotidyl transferase is down-regulated by AP-1 like regulatory elements in humanlymphoid cells. Immunology. **111**(2):195-203.
10. Elias, L., Longmire, J., Wood, A., Ratliff, R. (1982) Phosphorylation of terminal deoxynucleotidyl transferase in leukemic cells. Biochem Biophys Res Commun. **106**(2):458-65.

11. Sandor, Z., Calicchio, M. L., Sargent, R. G., Roth, D. B., Wilson, J.H. (2004) Distinct requirements for Ku in N nucleotide addition at V(D)J-and non-V(D)J-generated double-strand breaks. Nucleic Acids Res. **32**(6):1866-73.
12. Thai, T, H., Purugganan, M. M., Roth, D. B., Kearney, J. F. (2002) Distinct and opposite diversifying activities of terminal transferase splice variants. Nat Immunol. **3**(5):457-462.
13. Huyton, T., Bates, P. A., Zhang, X., Sternberg, M. J., Freemont, P. S. (2000) The BRCA1 C-terminal domain: structure and function. Mutat Res. **460**(3-4):319-332.
14. Yamtich, J., and Sweasy J. B. (2010) DNA Polymerase Family X: Function, Structure, and Cellular Roles. Biochim Biophys Acta. **1804**(5):1136–1150.
15. Beard, W. A, and Wilson, S. H. (2006) Structure and mechanism of DNA polymerase β. Chem Rev. **106**(2):361-82.
16. Steitz, T. A., and Steitz, J. A. (1993) A general two-metal-ion mechanism for catalytic RNA. Proc. Natl. Acad. Sci. USA. **90**:6498-6502.
17. Romain, F., Barbosa, I., Gouge, J., Rougeon, F., and Delarue, M. (2009) Conferring a template-dependent polymerase activity to terminal deoxynucleotidyltransferase by mutations in the Loop1 region. Nucl. Acids Res. **37**(14):4642-4656.
18. Tonegawa, S. (1983) Somatic generation of antibody diversity. Nature 302:575-581.
19. Hardy, R. R., and Hayakawa, K. (2001) B cell development pathways. Annu Rev Immunol. **19**:595-621.
20. Schatz, D. G., and Spanopoulou, E. (2005) Biochemistry of V(D)J recombination. Curr. Top. Microbiol. Immunol. **290**:49-85.
21. Lewis, S. M., and Wu, G. E. (2000) The old and the restless. J Exp Med. **191**(10):1631-1636.
22. Grawunder, U., Harfst, E. (2001) How to make ends meet in V(D)J recombination. Curr Opin Immunol. **13**(2):186-194.
23. Wilson, A., Held, W., MacDonald, H. R. (1994) Two waves of recombinase gene expression in developing thymocytes. J Exp Med. **179**(4):1355-1360.

24. Rolink, A. G., ten Boekel, E., Yamagami, T., Ceredig, R., Andersson, J., Melchers, F. (1999) B cell development in the mouse from early progenitors to mature B cells. Immunol Lett. **68**(1):89-93.
25. Akira, S., Okazaki, K., and Sakano, H. (1987) Two pairs of recombination signals are sufficient to cause immunoglobulin V-(D)-J joining. Science. **238**(4830):1134-1138.
26. Hesse, J. E., Lieber, M. R., Mizuuchi, K., Gellert, M. (1989) V(D)J recombination: a functional definition of the joining signals. Genes Dev. **3**(7):1053-1061.
27. Akamatsu, Y., Tsurushita, N., Nagawa, F., Matsuoka, M., Okazaki, K., Imai, M., Sakano, H.(1991) Essential residues in V(D)J recombination signals. J Immunol. **153**(10):4520-4529.
28. Ramsden, D. A., and Wu, G. E. (1991) Mouse kappa light-chain recombination signal sequences mediate recombination more frequently than do those of lambda light chain. Proc Natl Acad Sci USA. **88**(23):10721-10725.
29. West, K. L., Singha, N. C., De Ioannes, P., Lacomis, L., Erdjument-Bromage, H., Tempst, P., Cortes, P. (2005) A direct interaction between the RAG2 C terminus and the core histones is required for efficient V(D)J recombination. Immunity. **23**(2):203-212.
30. Matthews, A. G., Kuo, A. J., Ramón-Maiques, S., Han, S., Champagne, K. S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D. N. *et. al.* (2007) RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. Nature. **450**(7172):1106-1110.
31. McBlane, J. F., van Gent, D. C., Ramsden, D. A., Romeo, C., Cuomo, C. A., Gellert, M., Oettinger, M. A. (1995) Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. Cell. **83**(3):387-395.
32. Roth, D. B., Menetski, J. P., Nakajima, P. B., Bosma, M. J., Gellert, M. (1992) V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mousethymocytes. Cell. **70**(6):983-991.

33. Swanson, P. C. (2002) A RAG-1/RAG-2 tetramer supports 12/23-regulated synapsis, cleavage, and transposition of V(D)J recombination signals. Mol Cell Biol. **22**(22):7790-801.

34. Grundy, G. J., Ramón-Maiques, S., Dimitriadis, E. K., Kotova, S., Biertümpfel, C., Heymann, J. B., Steven, A. C., Gellert, M., Yang, W. (2009) Initial stages of V(D)J recombination: the organization of RAG1/2 and RSS DNA in the postcleavage complex. Mol Cell. **35**(2):217-227.

35. Rivera-Munoz, P., Malivert, L., Derdouch, S., Azerrad, C., Abramowski, V., Revy, P., Villartay, J. P. (2007) DNA repair and the immune system: from V(D)J recombination to aging lymphocytes. Eur J Immunol. **37** Suppl 1:S71-82.

36. Malu, S., Malshetty, V., Francis, D., Cortes, P. (2012) Role of non-homologous end joining in V(D)J recombination. Immunol Res. **54**(1-3):233-246.

37. Wiler, R , Leber, R., Moore, B. B., VanDyk, L. F., Perryman, L. E., Meek, K. (1995) Equine severe combined immunodeficiency: a defect in V(D)J recombination and DNA-dependent protein kinase activity. Proc Natl Acad Sci USA. **92**(25):11485-11489.

38. Rooney, S., Sekiguchi, J., Zhu, C., Cheng, H. L., Manis, J., Whitlow, S., De Vido., J., Foy, Dm, Chaudhuri, J., Lombard, D., Alt, F. W. (2002) Leaky Scid phenotype associated with defective V(D)J coding end processing in Artemis-deficient mice. Mol Cell. **10**(6):1379–1390.

39. Gottlieb, T. M., Jackson, S. P. (1993) The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. Cell. **72**(1):131-142.

40. Ma, Y., Pannicke, U., Schwarz, K., Lieber, M. R. (2002) Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell. **108**(6):781-794.

41. Shockett, P. E., and Schatz, D. G. (1999) DNA hairpin opening mediated by the RAG1 and RAG2 proteins. Mol Cell Biol. **19**(6):4159-4166.

42. Paull, T. T., and Gellert, M. (1999) Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. Genes Dev. **13**(10):1276-1288.

43. Desiderio, S. V., Yancopoulos, G. D., Paskind, M., Thomas, E., Boss, M.A., Landau, N., Alt, F. W., Baltimore, D. (1984) Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. Nature. **311**(5988):752–755.

44. Bentolila, L.A., Wu, G.E., Nourrit, F., Fanton d'Andon, M., Rougeon, F., Doyen, N. (1997) Constitutive expression of terminal deoxynucleotidyl transferase in transgenic mice is sufficient for N region diversity to occur at any Ig locus throughout B cell differentiation. J Immunol. **158**(2):715-23.

45. Komori, T., Okada, A., Stewart, V., Alt, F. W. (1993) Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. Science. **261**(5125):1171-1175.

46. Benedict, C. L., Gilfillan, S., Thai, T. H., Kearney, J. F. (2000) Terminal deoxynucleotidyl transferase and repertoire development. Immunol Rev. **175**:150-157.

47. Bentolila, L. A., Olson, S., Marshall, A., Rougeon, F., Paige, C. J., Doyen, N., Wu, G. E. (1999) Extensive junctional diversity in Ig light chain genes from early B cell progenitors of mu MT mice. J Immunol. **162**(4):2123-2128.

48. Schroeder, H. W. Jr, Mortari, F., Shiokawa, S., Kirkham, P. M, Elgavish, R. A., Bertrand, F. E. 3rd. (1995) Developmental regulation of the human antibody repertoire. Ann N Y Acad Sci. **764**:242-260.

49. Feeney, A. J. (1990) Lack of N regions in fetal and neonatal mouse immunoglobulin V-D-J junctional sequences. J Exp Med. **172**(5):1377-1390.

50. Mahajan, K. N., Nick McElhinny, S. A., Mitchell, B.S., Ramsden, D. A. (2002) Association of DNA polymerase μ (pol μ) with Ku and ligase IV: role for pol μ in end-joining double-strand break repair. Mol Cell Biol. **22**(14):5194-5202.

51. Bertocci, B., De Smet, A., Berek, C., Weill, J. C., Reynaud, C. A. (2003) Immunoglobulin kappa light chain gene rearrangement is impaired in mice deficient for DNA polymerase μ. Immunity. **19**(2):203–211.
52. Bertocci, B., De Smet, A., Weill, J. C., Reynaud, C. A. (2006) Nonoverlapping functions of DNA polymerases mu, lambda, and terminal deoxynucleotidyltransferase during immunoglobulin V(D)J recombination *in vivo*. Immunity. **25**(1):31-41.
53. Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K., Lieber, M. R. (1998) DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. Mol Cell. **2**(4):477–484.
54. Frank, K. M., Sekiguchi, J. M., Seidl, K. J., Swat, W., Rathbun, G. A., Cheng, H. L., Davidson, L., Kangaloo, L., Alt, F. W. (1998) Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. Nature. **396**(6707):173-177.
55. Gilfillan, S., Dierich, A., Lemeur, M., Benoist, C., Mathis, D. (1993) Mice lacking TdT: mature animals with an immature lymphocyte repertoire. Science. **261**(5125):1175-1178.
56. Gavin, M. A., Bevan, M. J. (1995) Increased peptide promiscuity provides a rationale for the lack of N regions in the neonatal T cell repertoire. Immunity. **3**(6):793–800.
57. Conde, C., Weller, S., Gilfillan, S., Marcellin, L., Martin, T., Pasquali, J. L. (1998) Terminal deoxynucleotidyl transferase deficiency reduces the incidence of autoimmune nephritis in (New Zealand Black x New Zealand White) F1 mice. J Immunol. **161**(12):7023-7030.
58. Freeney, A.J., Lawson, B.R., Kono, D.H. and Theofilopoulos, A.N. (2001) Terminal Deoxynucleotidyl Transferase Deficiency Decreases Autoimmune Disease in MRL-Fas^{lpr} Mice. J Immunol. **167**(6):3486-3493.
59. Robey, I. F., Peterson, M., Horwitz, M. S., Kono, D. H., Stratmann, T., Theofilopoulos, A. N., Sarvetnick, N., Teyton, L., Feeney, A. J. (2004) Terminal deoxynucleotidyltransferase deficiency decreases autoimmune disease in diabetes-prone nonobese diabetic mice and lupus-prone MRL-Fas(lpr) mice. J Immunol. **172**(7):4624-4629.

60. Molano, I.D., Redmond, A., Sekine, H., Zhang, X.K., Reilly, C., Hutchison, F., Ruiz, P. and Gilkeson, G.S. (2003) Effect of genetic deficiency of terminal deoxynucleotidyl transferase on autoantibody production and renal disease in MRL/lpr mice. Clinical Immunology. **107**:186-197.

61. Kodama, E. N., McCaffrey, R. P., Yusa, K., Mitsuya, H. (2000) Antileukemic activity and mechanism of action of cordycepin against terminal deoxynucleotidyl transferase- positive (TdT+) leukemic cells. Biochem Pharmacol. **59**(3):273-281.

62. Locatelli, G. A., Di Santo, R., Crespan, E., Costi, R., Roux, A., Hübscher, U., Shevelev, I., Blanca, G., Villani, G., Spadari, S., Maga, G. (2005) Diketo hexenoic acid derivatives are novel selective non-nucleoside inhibitors of mammalian terminal deoxynucleotidyl transferases, with potent cytotoxic effect against leukemic cells. Mol Pharmacol. **68**(2):538-550.

63. Syvänen, A. C. (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet. **2**(12):930-942.

64. Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J. *et. al.* (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science. **280**(5366):1077-1082.

65. Bunn, H. F. (1997) Pathogenesis and treatment of sickle cell disease. N Engl J Med. **337**(11):762-769.

66. Higgs, D. R., Wood, W. G. (2008) Genetic complexity in sickle cell disease. Proc Natl Acad Sci USA. **105**(33):11595-11596.

67. Niehues, T., Perez-Becker, R., Schuetz, C. (2010) More than just SCID--the phenotypic range of combined immunodeficiencies associated with mutations in the recombinase activating genes (RAG) 1 and 2. Clin Immunol. **135**(2):183-192.

68. Villa, A., Santagata, S., Bozzi, F., Giliani, S., Frattini, A., Imberti, L., Gatta, L. B., Ochs, H. D., Schwarz, K., Notarangelo, L. D. *et. al.* (1998) Partial V(D)J recombination activity leads to Omenn syndrome. Cell. **93**(5):885-896.

69. Spanopoulou, E., Cortes, P., Shih, C., Huang, C. M., Silver, D.P., Svec, P., Baltimore, D. (1995) Localization, interaction, and RNA binding properties of the V(D)J recombination-activating proteins RAG1 and RAG2. Immunity. **3**(6):715-26.

70. Gauss, G. and Lieber, M. R. (1993) Unequal Signal and Coding Joint Formation in Human V(D)J Recombination. Mol. Cell. Biol. **13**(7):3900-3906.

71. Gauss, G.H., and M.R. Lieber (1996). Mechanistic Constraints on Diversity in Human V(D)J Recombination. Mol. Cell. Biol. **16**(1):258-269.

72. Yuan, S.W., Agard, E. A., Larijani, M., Wu, G.E. (2005) Coding joint diversity in mature and immature B-cell lines. Scandinavian J. Immunology **62**(s1):114-118.

73. Menetski, J. P., and Gellert, M. (1990) V(D)J recombination activity in lymphoid cell lines is increased by agents that elevate cAMP. Proc Natl Acad Sci USA. **87**(23):9324-9328.

74. Boulé, J. B., Rougeon, F., Papanicolaou, C. (2001) Terminal deoxynucleotidyl transferase indiscriminately incorporates ribonucleotides and deoxyribonucleotides. J Biol Chem. **276**(33):31388-31393.

75. Lieber, M. R., Hesse, J. E., Mizuuchi, K., Gellert, M. (1988) Lymphoid V(D)J recombination: nucleotide insertion at signal joints as well as coding joints. Proc Natl Acad Sci USA. **85**(22):8588-8592.

76. Johnson, D., Morgan, A. R. (1976) The isolation of a high molecular weight terminal deoxynucleotidyl transferase from calf thymus. Biochem Biophys Res Commun. **72**(3):840-849.

77. Deibel, M. R. Jr, and Coleman, M. S. (1980) Biochemical properties of purified human terminal deoxynucleotidyltransferase. J Biol Chem. **255**(9):4206-4212.

78. Chang, L. M., and Bollum, F. J. (1990) Multiple roles of divalent cation in the terminal deoxynucleotidyltransferase reaction. J Biol Chem. **265**(29):17436-17440.

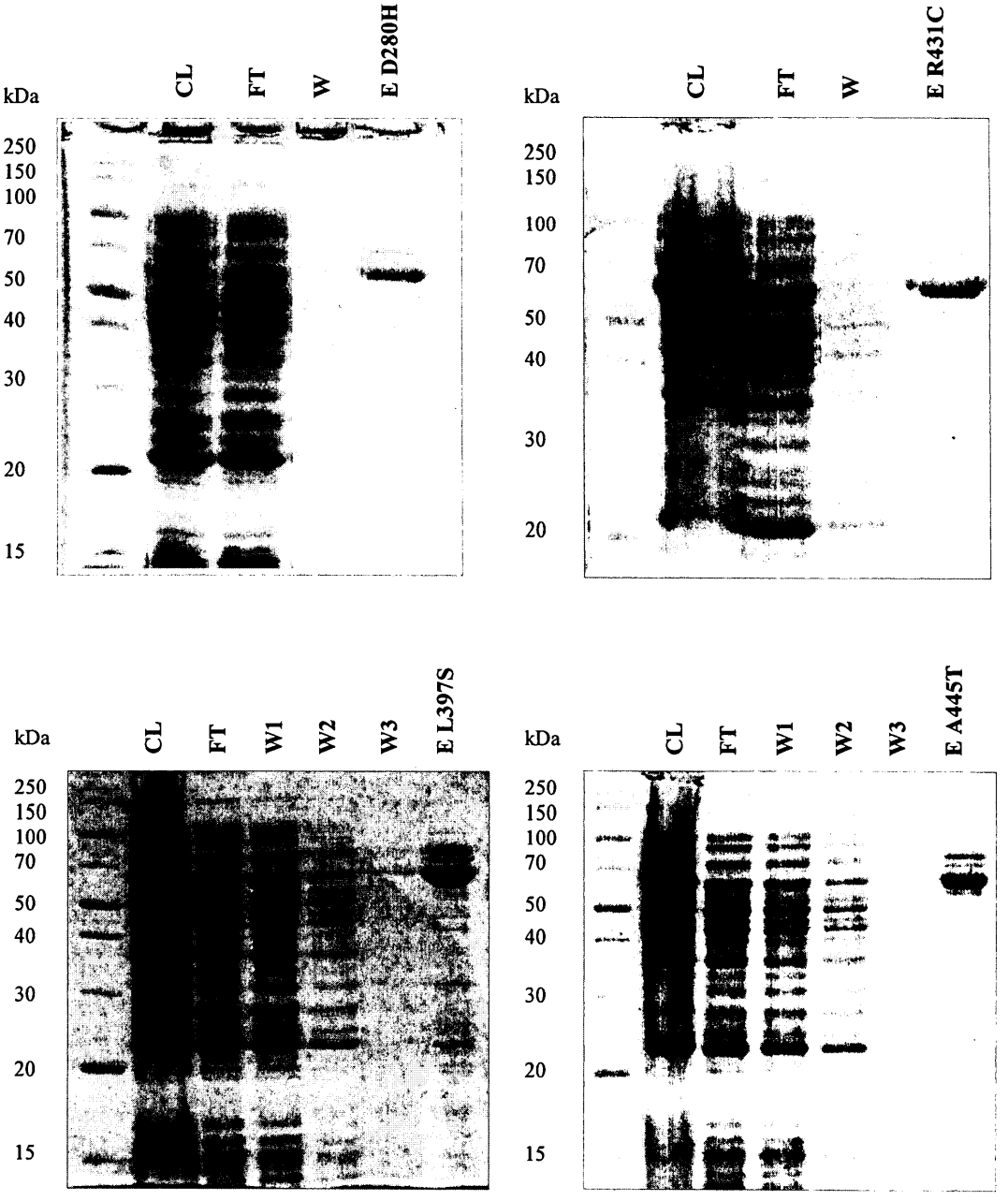
Appendix A: Complete list of hTdT variant 1 SNPs categorized according to hTdT mRNA position, base change, amino acid position and change, the chemical nature of amino acid change and the type of mutation. Chosen SNPs for the study are highlighted in bold. Adapted from NCBI SNP database.

	SNP ID	mRNA position	Base change	Amino acid position	Amino acid change	Chemical nature of amino acid change	Type of mutation
1	rs144113836	180	C to G	4	Pro to Ala	No change; non-polar to non-polar	Missense
2	rs199671984	197	G to C	9	Leu to Phe	No change; non-polar to non-polar	Missense
3	rs144375473	203	T to C	11	Pro to Pro	No change	Synonymous
4	rs148791733	216	C to T	16	Pro to Ser	Non-polar to polar	Missense
5	rs183314099	219	C to T	17	Arg to Trp	Positively charged to non-polar	Missense
6	rs188914324	245	C to T	25	Ser to Ser	No change	Synonymous
7	rs142422958	259	T to C	30	Ile to Thr	Non-polar to polar	Missense
8	rs112589253	279	G to A	37	Val to Ile	No change; non-polar to non-polar	Missense
9	rs139248550	301	T to C	44	Met to Thr	Non-polar to polar	Missense
10	rs201054340	311	C to T	47	Thr to Thr	No change	Synonymous
11	rs144026779	313	G to A	48	Arg to His	No change; positively charged to positively charged	Missense
12	rs144017257	348	G to C	60	Gly to Arg	Non-polar to positively charged	Missense
13	rs7081385	392	C to T	74	Ile to Ile	No change	Synonymous
14	rs36126211	439	C to T	90	Ala to Val	No change; non-polar to non-polar	Missense
15	rs150564632	455	C to T	95	Val to Val	No change	Synonymous
16	rs201835824	471	C to T	101	Leu to Phe	No change; non-polar to non-polar	Missense
17	rs143503401	476	C to T	102	Leu to Leu	No change	Synonymous
18	rs6584066	504	A to G	112	Arg to Gly	Positively charged to non-polar	Missense
19	rs148898308	511	G to A	114	Gly to Glu	Non-polar to negatively charged	Missense

20	rs61758435	557	A to G	129	Arg to Arg	No change	Synonymous
21	rs150972091	587	C to A	139	Pro to Pro	No change	Synonymous
22	rs141660410	588	C to A	140	Pro to Thr	Non-polar to polar	Missense
23	rs147054012	589	C to T	140	Pro to Leu	No change; non-polar to non-polar	Missense
24	rs201827729	590	G to A	140	Pro to Pro	No change	Synonymous
25	rs138317908	622	C to T	151	Ser to Phe	Polar to non-polar	Missense
26	rs34111381	665	C to T	165	Asn to Asn	No change	Synonymous
27	rs150189459	733	T to A	188	Cys to Tyr	No change; polar to polar	Missense
28	rs138972209	756	T to C	196	Ser to Pro	Polar to non-polar	Missense
29	rs181822026	773	G to A	201	Leu to Leu	No change	Synonymous
30	rs112342086	827	C to T	219	Ser to Ser	No change	Synonymous
31	rs183180123	872	T to C	234	Ser to Ser	No change	Synonymous
32	rs149429691	888	G to T	240	Val to Leu	No change; non-polar to non-polar	Missense
33	rs140275341	893	A to G	241	Leu to Leu	No change	Synonymous
34	rs144787508	902	A to G	244	Glu to Glu	No change	Synonymous
35	rs201883920	921	C to T	251	Leu to Phe	No change; non-polar to non-polar	Missense
36	rs139656736	924	T to C	252	Phe to Leu	No change; non-polar to non-polar	Missense
37	rs141822346	932	T to C	254	Ser to Ser	No change	Synonymous
38	rs200431503	975	A to T	269	Met to Leu	No change; non-polar to non-polar	Missense
39	**rs41291616**	**1008**	**G to C**	**280**	**Asp to His**	**Negatively charged to positively charged**	**Missense**
40	rs61729855	1010	C to A	280	Asp to Glu	No change; negatively charged to negatively charged	Missense
41	rs77358696	1046	A to C	292	Gly to Gly	No change	Synonymous
42	rs34175187	1091	A to C	307	Glu to Asp	No change; negatively charged to negatively charged	Missense
43	rs200138780	1127	C to G	319	Val to Val	No change	Synonymous
44	rs181202846	1144	A to T	325	Asp to Val	Negatively charged to non-polar	Missense
45	rs202209727	1217	C to A	349	Thr to Thr	No change	Synonymous

46	rs201092394	1220	C to T	350	Ser to Ser	No change	Synonymous
47	rs139853147	1248	C to T	360	Leu to Phe	No change; non-polar to non-polar	Missense
48	rs183025867	1255	A to G	362	Gln to Arg	Polar to positively charged	Missense
49	rs201883290	1283	G to A	371	Lys to Lys	No change	Synonymous
50	**rs142078021**	**1360**	**T to C**	**397**	**Leu to Ser**	**Non-polar to polar**	**Missense**
51	rs151142894	1402	G to A	411	Arg to His	No change; positively charged to positively charged	Missense
52	**rs139202533**	**1461**	**C to T**	**431**	**Arg to Cys**	**Positively charged to polar**	**Missense**
53	rs142389547	1462	G to A	431	Arg to His	No change; positively charged to positively charged	Missense
54	**rs149947558**	**1503**	**G to A**	**445**	**Ala to Thr**	**Non-polar to polar**	**Missense**
55	**rs145011041**	**1518**	**A to T**	**450**	**Thr to Ser**	**No change; polar to polar**	**Missense**
56	**rs142141382**	**1549**	**G to A**	**460**	**Arg to Gln**	**Positively charged to polar**	**Missense**
57	rs188851180	1548	C to T	460	Arg to Trp	Positively charged to non-polar	Missense
58	rs145922527	1563	C to T	465	His to Tyr	Positively charged to polar	Missense
59	rs147702836	1570	G to A	467	Arg to Gln	Positively charged to polar	Missense
60	rs200644125	1578	A to T	470	Ile to Phe	No change; non-polar to non-polar	Missense
61	rs190218659	1663	G to A	498	Gly to Glu	Polar to negatively charged	Missense
62	rs55671548	1674	A to T	502	Ile to Phe	No change; non-polar to non-polar	Missense

Appendix B:


kDa
CL
FT
W
E D280H
250
150
100
70
50
40
30
20
15
kDa
CL
FT
W
E R431C
250
150
100
70
50
40
30
20
kDa
CL
FT
W1
W2
W3
E L397S
250
150
100
70
50
40
30
20
15
kDa
CL
FT
W1
W2
W3
E A445T
250
150
100
70
50
40
30
20
15

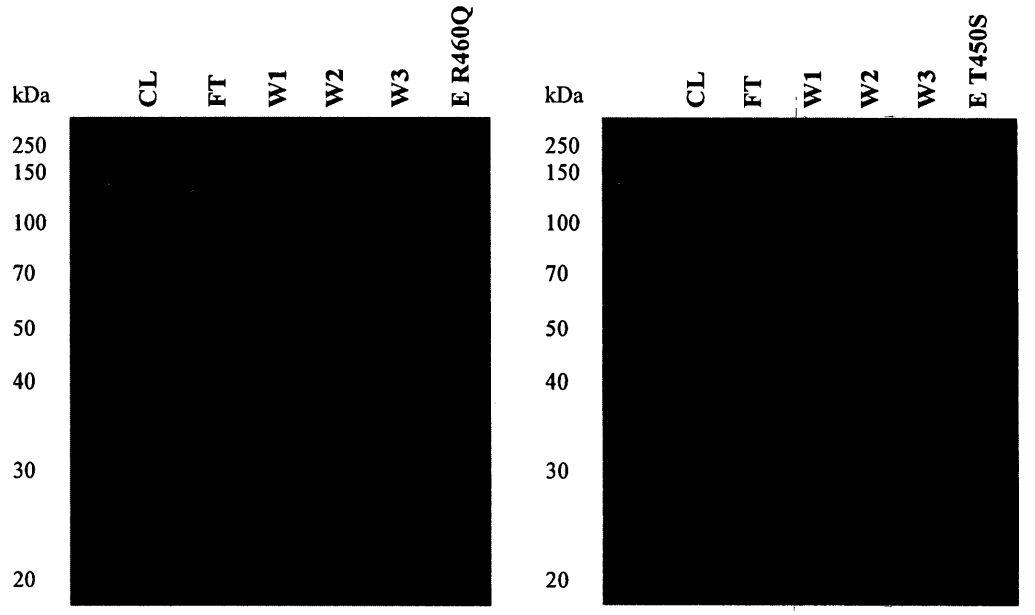


Figure 1: Coomassie stained 12% SDS-PAGE gels containing His-tag purification samples of D280H, R431C, L397S, A445T, R460Q and T450S hTdTS protein variants. Crude lysate (CL) fraction was collected after cell lysis. Flow-through (FT) fraction was collected after incubation of crude lysate with the nickel resin. The wash (W) fractions were collected to test for removal of non-specific nickel resin binding. The elution (E) fraction depicts single protein band at around 58kDa, corresponding to hTdTS.

Appendix C: Raw data of *in vivo* V(D)J extrachromosomal recombination assay.

Table 1: Summary of *in vivo* V(D)J recombination assay depicting 8 independent HEK293T cell transfection results. Bacterial counts of ampicillin resistant (AMP) and chloramphenicol and ampicillin resistant (CAM/AMP) colonies are shown along with number of picked double-resistant colonies and the number of true recombined plasmids obtained following colony PCR and plasmid sequencing.

TF #	Experimental Condition	AMP Plates			CAM/AMP Plates				Total # Picked CAM/ AMP Colonies	Number of True Recombined Plasmids (colony PCR)	Number of True Recombined Plasmids (sequencing)
		Volume Plated	Number of Colonies		Volume Plated	Number of Colonies					
			200X Dilution	400X Dilution		Plate A	Plate B	Plate C			
1	Control	100µl	207	111	450µl	0	2	-	2	0	0
	WT *	100µl	2	1	450µl	0	0	-	0	0	0
	A445T	100µl	32	39	450µl	1	0	-	1	0	0
	L397S	100µl	238	124	450µl	10	7	-	17	5	5
2	Control	100µl	TNTC	768	300µl	4	9	3	16	4	1
	WT	100µl	TNTC	800	300µl	4	3	1	8	2	2
	A445T	100µl	TNTC	560	300µl	3	2	3	8	4	3
	L397S	100µl	TNTC	868	300µl	3	9	4	16	11	11
	R431C	100µl	TNTC	652	300µl	3	5	4	12	10	6
3	Control	100µl	TNTC	118	450µl	151	250	-	27	15	14
	WT	100µl	TNTC	130	450µl	13	12	-	30	16	15
	A445T	100µl	TNTC	TNTC	450µl	16	7	-	23	15	13
	L397S	100µl	TNTC	TNTC	450µl	67	101	-	31	15	15
	R431C	100µl	320	197	450µl	6	14	-	22	9	9

TF #	Experimental Condition	AMP Plates			CAM/AMP Plates				Total # Picked CAM/ AMP Colonies	Number of True Recombined Plasmids (colony PCR)	Number of True Recombined Plasmids (sequencing)
		Volume Plated	Number of Colonies		Volume Plated	Number of Colonies					
			100X Dilution	500X Dilution		Plate A	Plate B	Plate C			
4	Control	100µl	780	468	300µl	30	43	42	18	11	11
	WT	100µl	451	105	300µl	1	2	6	9	4	4
	A445T	100µl	342	107	300µl	4	4	2	10	7	5
	L397S	100µl	936	394	300µl	14	34	19	30	19	19
	R431C	100µl	290	88	300µl	8	5	3	16	11	10
5	Control	100µl	860	351	300µl	25	28	55	15	7	7
	WT	100µl	285	59	300µl	18	2	2	15	4	4
	A445T	100µl	425	136	300µl	0	3	3	6	0	0
	L397S	100µl	760	255	300µl	29	9	22	15	10	9
	R431C	100µl	227	79	300µl	6	0	2	8	0	0
6	Control	100µl	960	401	300µl	71	79	61	30	16	14
	WT	100µl	452	226	300µl	0	5	6	11	7	6
	A445T	100µl	607	60	300µl	3	2	7	12	8	7
	L397S	100µl	712	289	300µl	23	22	26	50	35	29
	R431C	100µl	97	174	300µl	2	2	3	7	2	2
7	Control	100µl	TNTC	447	300µl	114	91	57	30	13	11
	WT	100µl	511	430	300µl	30	24	17	45	18	16
	A445T	100µl	520	83	300µl	14	27	28	60	34	33
	L397S	100µl	TNTC	792	300µl	131	110	81	30	16	13
	R431C	100µl	648	259	300µl	17	13	14	45	12	10

TF #	Experimental Condition	AMP Plates			CAM/AMP Plates				Total # Picked CAM/ AMP Colonies	Number of True Recombined Plasmids (colony PCR)	Number of True Recombined Plasmids (sequencing)
		Volume Plated	Number of Colonies		Volume Plated	Number of Colonies					
			100X Dilution	500X Dilution		Plate A	Plate B	Plate C			
8	Control	100µl	TNTC	504	300µl	22	29	17	24	14	11
	WT	100µl	736	237	300µl	5	7	4	16	13	13
	A445T	100µl	1068	288	300µl	2	10	8	20	14	12
	L397S *	100µl	332	133	300µl	2	1	4	7	4	4
	R431C	100µl	448	114	300µl	6	4	1	11	6	6

Experimental conditions of HEK293T cell transfections;
Control = transfection with substrate pGG51 plasmid + pEBG-hRAG1 + pEBG-hRAG2 expressing vectors
WT = transfection with wild type hTdTS pcDNA expressing vector + substrate pGG51 plasmid + pEBG-hRAG1+ pEBG-hRAG2 expressing vectors
A445T = transfection with A445T hTdTS pcDNA expressing vector + substrate pGG51 plasmid + pEBG-hRAG1+ pEBG-hRAG2 expressing vectors
L397S = transfection with L397S hTdTS pcDNA expressing vector + substrate pGG51 plasmid + pEBG-hRAG1+ pEBG-hRAG2 expressing vectors
R431C = transfection with R431C hTdTS pcDNA expressing vector + substrate pGG51 plasmid + pEBG-hRAG1+ pEBG-hRAG2 expressing vectors

* The bacterial transformation via electroporation of those transfection conditions marked by a star (*) has arched, indicated by a popping sound, thus reducing transformation efficiency due to uneven electrical current applied.

** TNCT: Too numerous to count.

Table 2: Calculating recombination frequency values (R-value) of 8 independent HEK293T cell transfections. Bacterial counts of ampicillin resistant and chloramphenicol and ampicillin resistant colonies are shown for 1ml of undiluted transformation mixture. The mean R-value and standard deviation are calculated from 8 independent HEK293T cell transfections.

TF #	Control	WT hTdTS	A445T hTdTS	L397S hTdTS	R431C hTdTS
Ampicillin resistant colonies					
1	444000	---	15600	49600	---
2	307200	320000	224000	347200	260800
3	47200	52000	TNCT	TNCT	78800
4	2340000	525000	535000	1970000	440000
5	1755000	295000	680000	1275000	395000
6	2005000	1130000	300000	1445000	485000
7	2235000	2150000	415000	3960000	1295000
8	2520000	1185000	40000	665000	570000
Chloramphenicol and ampicillin resistant colonies					
1	2	---	1	19	---
2	18	9	9	18	13
3	446	28	26	187	22
4	128	10	11	74	18
5	120	24	7	67	9
6	234	12	13	79	8
7	291	79	60	358	49
8	76	18	22	8	12

TF #	Control	WT hTdTS	A445T hTdTS	L397S hTdTS	R431C hTdTS
			R-value (%)		
1	0.00045045	---	0.006410256	0.038306452 *	---
2	0.005859375	0.0028125	0.004017857	0.005184332	0.004984663
3	0.944915254 *	0.053846154 *	---	---	0.027918782 *
4	0.005470085	0.001904762	0.002056075	0.003756345	0.004090909
5	0.006837607	0.008135593	0.001029412	0.005254902	0.002278481
6	0.011670823	0.001061947	0.004333333	0.005467128	0.001649485
7	0.013020134	0.003674419	0.014457831 *	0.009040404	0.003783784
8	0.003015873	0.001518987	0.05500000 *	0.001203008	0.002105263
Mean R-value (%)	0.00662	0.00318	0.00357	0.00498	0.00315
Standard deviation	0.00446	0.0026	0.0021	0.00255	0.00132

* Values marked with a star (*) were disregarded in mean R-value and standard deviation calculations, due to their significant deviation from other values. The bacterial transformations associated with these values were arched, indicated by a popping sound, thus reducing transformation efficiency due to uneven electrical current applied.
** Dashed lines (---): value is not available due to arched bacterial transformation.
*** TNCT: Too numerous to count.
****The R-values calculated in this table are represented on a scatter-plot graph in Results section, Figure 3.19. Number of ampicillin resistant colonies was calculated using the lower dilution available; either 400X or 500X. Sample calculation is shown below.

Sample calculation of R-value (%)

$$\text{R-value (\%)} = \frac{\text{The total number of AMP + CAM resistant colonies in a specific volume of undiluted transformation sample}}{\text{The total number of AMP resistant colonies in the same volume of undiluted transformation sample}} \times 100\%$$

Sample calculation from Control TF #1 (Appendix C, Table 1);

AMP resistant colonies in 400X diluted transformation sample = 111
AMP resistant colonies in 1ml of undiluted transformation sample = 44,400

CAM+AMP resistant colonies in 900µl undiluted transformation sample = 2
CAM/AMP resistant colonies in 1ml of undiluted transformation sample = 2
[(2x1000µl)/900µl=2]

R-value (%) = (2/44,400) * 100 % = **0.0045045 %**

Table 3: Total number of unique substrate pGG51 plasmid sequences obtained in 8 independent mammalian cell transfections in *in vivo* extrachromosomal V(D)J recombination assay.

	Control	WT hTdTS	A445T hTdTS	L397S hTdTS	R431C hTdTS
TF 1	0 (0)	0 (0)	0 (0)	45 (1)	0 (0)
TF 2	2 (2)	2 (0)	3 (0)	7 (3)	5 (1)
TF 3	11 (3)	15 (0)	11 (2)	10 (5)	6 (3)
TF 4	10 (1)	4 (0)	4 (1)	12 (7)	6 (4)
TF 5	4 (3)	4 (0)	0 (0)	7 (2)	0 (0)
TF 6	8 (6)	7 (1)	7 (0)	21 (8)	2 (0)
TF 7	8 (3)	15 (1)	29 (4)	7 (6)	9 (1)
TF 8	5 (6)	12 (1)	11 (1)	4 (0)	6 (0)
Total number of unique sequences	48 (22)	57 (3)	65 (8)	72 (32)	34 (9)

* TF stands for HEK293T cell transfection number.
** The number of repeated sequences generated within the same transfection, shown in brackets, were excluded from the joint analysis.

Experimental conditions of HEK293T cell transfections;
Control = transfection with substrate pGG51 plasmid + pEBG-hRAG1 + pEBG-hRAG2 expressing vectors
WT hTdTS = transfection with wild type hTdTS pcDNA expressing vector + substrate pGG51 plasmid + pEBG-hRAG1+ pEBG-hRAG2 expressing vectors
A445T hTdTS = transfection with A445T hTdTS pcDNA expressing vector + substrate pGG51 plasmid + pEBG-hRAG1+ pEBG-hRAG2 expressing vectors
L397S hTdTS = transfection with L397S hTdTS pcDNA expressing vector + substrate pGG51 plasmid + pEBG-hRAG1+ pEBG-hRAG2 expressing vectors
R431C hTdTS = transfection with R431C hTdTS pcDNA expressing vector + substrate pGG51 plasmid + pEBG-hRAG1+ pEBG-hRAG2 expressing vectors

Appendix D: Analysis of all obtained recombined substrate pGG51 plasmid joint sequences from *in vivo* V(D)J recombination assay.

Sample		5′ flanking region	P	N	P	3′ flanking region
Germline		TGGCTGCAGGTCGAC				GGATCCCCGGGGATC
1	2-C1	TGGCTGCAGGTCG--				-GATCCCCGGGGATC
2	2-C2	TGGCTGCAGGTCGAC		G		---TCCCCGGGGATC
3	3-C1	TGGCTGCAGGTCGAC	GT			-----CCCGGGGATC
4	3-C2	TGGCTGCAGGTCGAC	GT			----CCCCGGGGATC
5	3-C3	TGGCTGCAGG-----		C		-GATCCCCGGGGATC
6	3-C4	TGGCTGCAGGT----				--ATCCCCGGGGATC
7	3-C5	TGGCTGCAGGTC---				---TCCCCGGGGATC
8	3-C6	TGGCTGCAGGTCG--				----CCCCGGGGATC
9	3-C7	TGGCTGCAGGTC---				-----CCCGGGGATC
10	3-C8	TGGCTGCAGGTCGAC				-GATCCCCGGGGATC
11	3-C9	TGGCTGCAGGTCGAC	G			---TCCCCGGGGATC
12	3-C10	TGGCTGCAG------				--ATCCCCGGGGATC
13	3-C11	TGGCTGCAGGTCG--				-GATCCCCGGGGATC
14	4-C1	TGGCTGCAGGTC---				-----CCCGGGGATC
15	4-C2	TGGCTGCAGGTCGAC	GT			--ATCCCCGGGGATC
16	4-C3	TGGCTGCAGG-----				-GATCCCCGGGGATC
17	4-C4	TGGCTGCAGGTC---				-----CCCGGGGATC
18	4-C5	TGGCTGCAGGTCGAC	GT			----CCCCGGGGATC
19	4-C6	TGGCTGCAGG-----		C		-GATCCCCGGGGATC
20	4-C7	TGGCTGCAGGT----				-GATCCCCGGGGATC
21	4-C8	TGGCTGCAGGTCGAC	G			---TCCCCGGGGATC
22	4-C9	TGGCTGCAGGTCGAC	GT			-GATCCCCGGGGATC
23	4-C10	TGGCTGCAGG-----				--ATCCCCGGGGATC

Sample		5′ flanking region	P	N	P	3′ flanking region
24	5-C1	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
25	5-C2	TGGCTGCAGGTCG--				-GATCCCCGGGGATC
26	5-C3	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
27	5-C4	TGGCTGCAGG-----				--ATCCCCGGGGATC
28	6-C1	TGGCTGCAGGTC---				-----CCCGGGGATC
29	6-C2	TGGCTGCAGGTCGAC	GT			----CCCCGGGGATC
30	6-C3	TGGCTGCAG------				--ATCCCCGGGGATC
31	6-C4	TGGCTGCAGGT----				--ATCCCCGGGGATC
32	6-C5	TGGCTGCAGGT----				-GATCCCCGGGGATC
33	6-C6	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
34	6-C7	TGGCTGCAGGTCGAC	G			---TCCCCGGGGATC
35	6-C8	TGGCTGCAGG-----				--ATCCCCGGGGATC
36	7-C1	TGGCTGCAGGTC---				----CCCCGGGGATC
37	7-C2	TGGCTGCAGGTCGAC				----CCCCGGGGATC
38	7-C3	TGGCTGCAGGTCGAC	G			---TCCCCGGGGATC
39	7-C4	TGGCTGCAGGTCGAC	GT			----CCCCGGGGATC
40	7-C5	TGGCTGCAGGTC---				-----CCCGGGGATC
41	7-C6	TGGCTGCAGGTCGAC				GGATCCCCGGGGATC
42	7-C7	TGGCTGCAGGTCG--				GGATCCCCGGGGATC
43	7-C8	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
44	8-C1	TGGCTGCAGGTC---				------CCGGGGATC
45	8-C2	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
46	8-C3	TGGCTGCAGGTCG--				-GATCCCCGGGGATC
47	8-C4	TGGCTGCAGGTGGA-				---TCCCCGGGGATC
48	8-C5	TGGCTGCAGGTCGAC	G			---TCCCCGGGGATC

Sample		5′ flanking region	P	N	P	3′ flanking region
1	2-W1	TGGCTGCAGGTCGAC	G	GTTCTT		--ATCCCCGGGGATC
2	2-W2	TGGCTGCAGG-----		CTT		GGATCCCCGGGGATC
3	3-W1	TGGCTGCAGGTCG--		CC		---TCCCCGGGGATC
4	3-W2	TGGCTGCAGG-----				--ATCCCCGGGGATC
5	3-W3	TGGCTGCAGGT----				--ATCCCCGGGGATC
6	3-W4	TGGCTGCAGGTCGAC		AAGA		----CCCCGGGGATC
7	3-W5	TGGCTGCAGGTCG--		GGGGTCTG		GGATCCCCGGGGATC
8	3-W6	TGGCTGCAGGTCGAC	GT	GG		---TCCCCGGGGATC
9	3-W7	TGGCTGCAGGTCGAC	GT	CGAA		-----CCCGGGGATC
10	3-W8	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
11	3-W9	TGGCTGCAGGTCG--				GGATCCCCGGGGATC
12	3-W10	TGGCTGCAGGTCG--		G		GGATCCCCGGGGATC
13	3-W11	TGGCTGCAGGTCGA-		AGG		----CCCCGGGGATC
14	3-W12	TGGCTGCAGGTCGA-		AGG		---TCCCCGGGGATC
15	3-W13	TGGCTGCAGG-----		GCCC		-GATCCCCGGGGATC
16	3-W14	TGGCTGCAGGTCGAC	GT	AA		-GATCCCCGGGGATC
17	3-W15	TGGCTGCAG------		ATC		--ATCCCCGGGGATC
18	4-W1	TGGCTGCAGGTCGA-		GG		-----CCCGGGGATC
19	4-W2	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
20	4-W3	TGGCTGCAGG-----		CCGC		-GATCCCCGGGGATC
21	4-W4	TGGCTGCAGGTCG--				-GATCCCCGGGGATC
22	5-W1	TGGCTGCAGGTCG--		GCCTT		--ATCCCCGGGGATC
23	5-W2	TGGCTGCAGG-----		CC		-GATCCCCGGGGATC
24	5-W3	TGGCTGCAGG-----		CGAT		-GATCCCCGGGGATC
25	5-W4	TGGCTGCAGGTCGA-				--ATCCCCGGGGATC

Sample		5′ flanking region	P	N	P	3′ flanking region
26	6-W1	TGGCTGCAGGTCGAC		CC		---ATCCCCGGGGAT
27	6-W2	TGGCTGCAGGTCGAC	G	CC		----CCCCGGGGATC
28	6-W3	TGGCTGCAGGT----		GGGT		--ATCCCCGGGGATC
29	6-W4	TGGCTGCAGGTC---		C		-GATCCCCGGGGATC
30	6-W5	TGGCTGCAGGTCG--				-GATCCCCGGGGATC
31	7-W1	TGGCTGCAGGTCGAC		AAAGGGG		----CCCCGGGGATC
32	7-W2	TGGCTGCAGG-----		CT	CC	GGATCCCCGGGGATC
33	7-W3	TGGCTGCAGGTCG--		TA		GGATCCCCGGGGATC
34	7-W4	TGGCTGCAGGTCGAC		AC		----CCCCGGGGATC
35	7-W5	TGGCTGCAGG-----		CC	CC	GGATCCCCGGGGATC
36	7-W6	TGGCTGCAGG-----			CC	GGATCCCCGGGGATC
37	7-W7	TGGCTGCAGGTCG--				-GATCCCCGGGGATC
38	7-W8	TGGCTGCAGG-----		CG		GGATCCCCGGGGATC
39	7-W9	TGGCTGCAGG-----		CCT		GGATCCCCGGGGATC
40	7-W10	TGGCTGCAGGTCGAC				--GATCCCGGGGATC
41	7-W11	TGGCTGCAGGT----		AGG		---TCCCCGGGGATC
42	7-W12	TGGCTGCAGGTCG--		CGGAA		-----CCCGGGGATC
43	7-W13	TGGCTGCAGG-----		CCTT		--ATCCCCGGGGATC
44	7-W14	TGGCTGCAGG-----		CACT		-GATCCCCGGGGATC
45	7-W15	TGGCTGCAGGTCGAC	GT	ATC		-GATCCCCGGGGATC
46	8-W1	TGGCTGCAGG-----			CC	GGATCCCCGGGGATC
47	8-W2	TGGCTGCAGGTC---				-----CCCGGGGATC
48	8-W3	TGGCTGCAGGTCG--		T		GGATCCCCGGGGATC
49	8-W4	TGGCTGCAGGTCG--				GGATCCCCGGGGATC
50	8-W5	TGGCTGCAGG-----		CCT		-GATCCCCGGGGATC

Sample		5′ flanking region	P	N	P	3′ flanking region
51	8-W6	TGGCTGCAGGT----		GG		GGATCCCCGGGGATC
52	8-W7	TGGCTGCAGGT----		T		---TCCCCGGGGATC
53	8-W8	TGGCTGCAGGT----		TCT		GGATCCCCGGGGATC
54	8-W9	TGGCTGCAGGTC---		CAT		--ATCCCCGGGGATC
55	8-W10	TGGCTGCAGG-----		CC		--ATCCCCGGGGATC
56	8-W11	TGGCTGCAGGTCGA-		A		------CCGGGGATC
57	8-W12	TGGCTGCAGGTGGA-				---TCCCCGGGGATC
1	2-A1	TGGCTGCAGGTCG--				----CCCCGGGGATC
2	2-A2	TGGCTGCAGGTCG--		GGG		-----CCCGGGGATC
3	2-A3	TGGCTGCAGGTCG--				---CCCCCGGGGATC
4	3-A1	TGGCTGCAGGTCGAC	G	AG		GGATCCCCGGGGATC
5	3-A2	TGGCTGCAGG-----		CC		-GATCCCCGGGGATC
6	3-A3	TGGCTGCAGG-----		CCTT		--ATCCCCGGGGATC
7	3-A4	TGGCTGCAGGTCGA-		A		----CCCCGGGGATC
8	3-A5	TGGCTGCAGGTCGAC	G	G		GGATCCCCGGGGATC
9	3-A6	TGGCTGCAGGT----		T		GGATCCCCGGGGATC
10	3-A7	TGGCTGCAGGTCGAC	GT	GGG		-----CCCGGGGATC
11	3-A8	TGGCTGCAGGTCGAC		AGGG		-----CCCGGGGATC
12	3-A9	TGGCTGCAGGTCGAC	GT	AGGG		----CCCCGGGGATC
13	3-A10	TGGCTGCAGG-----			CC	GGATCCCCGGGGATC
14	3-A11	TGGCTGCAGGTCG--				-GATCCCCGGGGATC
15	4-A1	TGGCTGCAGG-----			CC	GGATCCCCGGGGATC
16	4-A2	TGGCTGCAGGTCGAC				-----CCCGGGGATC
17	4-A3	TGGCTGCAGGTCG--		GA		--ATCCCCGGGGATC
18	4-A4	TGGCTGCAGGTCG--		CCCC		-GATCCCCGGGGATC

Sample		5′ flanking region	P	N	P	3′ flanking region
19	6-A1	TGGCTGCAGGTCG--				-GATCCCCGGGGATC
20	6-A2	TGGCTGCAGGTCG--		GG		----CCCCGGGGATC
21	6-A3	TGGCTGCAGGTC---				-GATCCCCGGGGATC
22	6-A4	TGGCTGCAGGTCGAC	GT	AAC		----CCCCGGGGATC
23	6-A5	TGGCTGCAGGTCGAC	G	C		-GATCCCCGGGGATC
24	6-A6	TGGCTGCAGGTC---				-----CCCGGGGATC
25	6-A7	TGGCTGCAGGTCGA-		GAG		----CCCCGGGGATC
26	7-A1	TGGCTGCAGGT----		G		GGATCCCCGGGGATC
27	7-A2	TGGCTGCAGGT----		TC		-GATCCCCGGGGATC
28	7-A3	TGGCTGCAGG-----		GCAT		--ATCCCCGGGGATC
29	7-A4	TGGCTGCAG------		CCCCC		-GATCCCCGGGGATC
30	7-A5	TGGCTGCAGGT----		T		--ATCCCCGGGGATC
31	7-A6	TGGCTGCAGGT----		TCCT		--ATCCCCGGGGATC
32	7-A7	TGGCTGCAGGTCG--		CC		-GATCCCCGGGGATC
33	7-A8	TGGCTGCAG------		TT	CC	GGATCCCCGGGGATC
34	7-A9	TGGCTGCAGGTCGAC	G			GGATCCCCGGGGATC
35	7-A10	TGGCTGCAGG-----		CCCC		-GATCCCCGGGGATC
36	7-A11	TGGCTGCAGGTC---		CC		----CCCCGGGGATC
37	7-A12	TGGCTGCAGGTC---		T		GGATCCCCGGGGATC
38	7-A13	TGGCTGCAGGTCG--		CCG		GGATCCCCGGGGATC
39	7-A14	TGGCTGCAGGT----		T		-GATCCCCGGGGATC
40	7-A15	TGGCTGCAGGTCG--		CCCC		--ATCCCCGGGGATC
41	7-A16	TGGCTGCAGGTCG--		CCC		-GATCCCCGGGGATC
42	7-A17	TGGCTGCAGG-----		GC		-GATCCCCGGGGATC
43	7-A18	TGGCTGCAGGTCGAC	G	G		----CCCCGGGGATC

Sample		5′ flanking region	P	N	P	3′ flanking region
44	7-A19	TGGCTGCAGG-----		CCC		-GATCCCCGGGGATC
45	7-A20	TGGCTGCAGGT----		ATT		--ATCCCCGGGGATC
46	7-A21	TGGCTGCAGGTCG--		C		-GATCCCCGGGGATC
47	7-A22	TGGCTGCAGGTCG--		GC		--ATCCCCGGGGATC
48	7-A23	TGGCTGCAGG-----		GG		GGATCCCCGGGGATC
49	7-A24	TGGCTGC--------			CC	GGATCCCCGGGGATC
50	7-A25	TGGCTGCAGGTCG--				-GATCCCCGGGGATC
51	7-A26	TGGCTGCAGGTCG--		C	CC	GGATCCCCGGGGATC
52	7-A27	TGGCTGCAGGTCG--			CC	GGATCCCCGGGGATC
53	7-A28	TGGCTGCAGGTCGAC		GGAA		----CCCCGGGGATC
54	7-A29	TGGCTGCAGG-----		CC		-GATCCCCGGGGATC
55	8-A1	TGGCTGCAGGTCG--		TTT		-GATCCCCGGGGATC
56	8-A2	TGGCTGCAGGTCGAC	G			---TCCCCGGGGATC
57	8-A3	TGGCTGCAGGTCG--		GG		----CCCCGGGGATC
58	8-A4	TGGCTGCAGGTCG--		GAG		-----CCCGGGGATC
59	8-A5	TGGCTGCAGGTCG--		TC		-GATCCCCGGGGATC
60	8-A6	TGGCTGC--------		CGCC		-GATCCCCGGGGATC
61	8-A7	TGGCTGCAG------		C		-GATCCCCGGGGATC
62	8-A8	TGGCTGCAGGTCGAC	G	GAG		----CCCCGGGGATC
63	8-A9	TGGCTGCAGGT----		T		-GATCCCCGGGGATC
64	8-A10	TGGCTGCAGGTCG--		GC		-GATCCCCGGGGATC
65	8-A11	TGGCTGCAGGT----		T		-GATCCCCGGGGATC
1	1-L1	TGGCTGCAGGTCGAC	GT			-----CCCGGGGATC
2	1-L2	TGGCTGCAGGTCG--				-GATCCCCGGGGATC
3	1-L3	TGGCTGCAGG-----		C		GGATCCCCGGGGATC

Sample		5′ flanking region	P	N	P	3′ flanking region
4	1-L4	TGGCTGCAGGT----				----CCCCGGGGATC
5	2-L1	TGGCTGCAGGTCGAC		AC		-GATCCCCGGGGATC
6	2-L2	TGGCTGCAGGTC---				-----CCCGGGGATC
7	2-L3	TGGCTGCAGGT----		A		--ATCCCCGGGGATC
8	2-L4	TGGCTGCAGGTC---				--ATCCCCGGGGATC
9	2-L5	TGGCTGCAGGTCGAC				-GATCCCCGGGGATC
10	2-L6	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
11	2-L7	TGGCTGCAGGTCGAC	G	AAA		----CCCCGGGGATC
12	3-L1	CCTTGCCAGGTCGA-		ACGTAA		-----CCCGGGGATC
13	3-L2	TGGCTGCAGGTCGA-		GT		-GATCCCCGGGGATC
14	3-L3	TGGCTGCAGGTCG--				-GATCCCCGGGGATC
15	3-L4	TGGCTGCAG------				--ATCCCCGGGGATC
16	3-L5	TGGCTGCAGG-----		C		-GATCCCCGGGGATC
17	3-L6	TGGCTGCAGGTC---				-----CCCGGGGATC
18	3-L7	TGGCTGCAG------				--ATCCCCGGGGATC
19	3-L8	TGGCTGCAGGT----				--ATCCCCGGGGATC
20	3-L9	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
21	3-L10	TGGCTGCAGGTCGAC	GT			----CCCCGGGGATC
22	4-L1	TGGCTGCAGG-----			C	GGATCCCCGGGGATC
23	4-L2	TGGCTGCAG------				------CCGGGGATC
24	4-L3	TGGCTGCAG------		TTT		-GATCCCCGGGGATC
25	4-L4	TGGCTGCAGGT----		GA		----CCCCGGGGATC
26	4-L5	TGGCTGCAGGTCGAC	G			---TCCCCGGGGATC
27	4-L6	TGGCTGCAGGTC---				----CCCCGGGGATC
28	4-L7	TGGCTGCAGGTC---			C	GGATCCCCGGGGATC

Sample		5′ flanking region	P	N	P	3′ flanking region
29	4-L8	TGGCTGCAGGTCG--		GG		----CCCCGGGGATC
30	4-L9	TGGCTGCAGGT----				-GATCCCCGGGGATC
31	4-L10	TGGCTGCAGGTCG--				-GATCCCCGGGGATC
32	4-L11	TGGCTGCAGGTC---				-----CCCGGGGATC
33	4-L12	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
34	5-L1	TGGCTGCAGGTC---				----CCCCGGGGATC
35	5-L2	TGGCTGCAGGTCG--				-GATCCCCGGGGATC
36	5-L3	TGGCTGCAGGTC---				-----CCCGGGGATC
37	5-L4	TGGCTGCAGGT----				GGATCCCCGGGGATC
38	5-L5	TGGCTGCAGG-----		C		-GATCCCCGGGGATC
39	5-L6	TGGCTGCAGGT----				-GATCCCCGGGGATC
40	5-L8	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
41	6-L1	TGGCTGCAGGTCG--			CC	GGATCCCCGGGGATC
42	6-L2	TGGCTGCAGG-----		C		--ATCCCCGGGGATC
43	6-L3	TGGCTGCAGGT----				---TCCCCGGGGATC
44	6-L4	TGGCTGCAGGT----				--ATCCCCGGGGATC
45	6-L5	TGGCTGCAGGTCG--				-GATCCCCGGGGATC
46	6-L6	TGGCTGCAGG-----		CC		--ATCCCCGGGGATC
47	6-L7	TGGCTGCAGG-----				-GATCCCCGGGGATC
48	6-L8	TGGCTGCAGGTCGAC				----CCCCGGGGATC
49	6-L9	TGGCTGCAGGTC---				-----CCCGGGGATC
50	6-L10	TGGCTGCAGGTCGAC	GT			----CCCCGGGGATC
51	6-L11	TGGCTGCAGG-----				--ATCCCCGGGGATC
52	6-L12	TGGCTGCAG------		TT		--ATCCCCGGGGATC
53	6-L13	TGGCTGCAGGTCG--				---TCCCCGGGGATC

Sample		5′ flanking region	P	N	P	3′ flanking region
54	6-L14	TGGCTGCAGGTC---				-----CCCGGGGATC
55	6-L15	TGGCTGCAGGTCGAC				---TCCCCGGGGATC
56	6-L16	TGGCTGCAGGTC---				--ATCCCCGGGGATC
57	6-L17	TGGCTGCAGG-----			CC	GGATCCCCGGGGATC
58	6-L18	TGGCTGCAGGTC---				--ATCCCCGGGGATC
59	6-L19	TGGCTGCAGGTCGAC		A		----CCCCGGGGATC
60	6-L20	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
61	6-L21	TGGCTGCAGGTCGAC	G			---TCCCCGGGGATC
62	7-L1	TGGCTGCAGGTCGAC	G			---TCCCCGGGGATC
63	7-L2	TGGCTGCAGGTCG--				-GATCCCCGGGGATC
64	7-L3	TGGCTGCAGGTCGAC				-GATCCCCGGGGATC
65	7-L4	TGGCTGCAGG-----		GC		-GATCCCCGGGGATC
66	7-L5	TGGCTGCAGGT----				-GATCCCCGGGGATC
67	7-L6	TGGCTGCAGGTC---				-----CCCGGGGATC
68	7-L7	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
69	8-L1	TGGCTGCAGGTCGAC	G	GA		----CCCCGGGGATC
70	8-L2	TGGCTGCAGGTCGAC	G			---TCCCCGGGGATC
71	8-L3	TGGCTGCAGGTC---				--ATCCCCGGGGATC
72	8-L4	TGGCTGCAGGT----				--ATCCCCGGGGATC
1	2-R1	TGGCTGCAGG-----				-GATCCCCGGGGATC
2	2-R2	TGGCTGCAGGTCGAC				----CCCCGGGGATC
3	2-R3	TGGCTGCAGGTC---				----CCCCGGGGATC
4	2-R4	TGGCTGCAGGTCGAC	G			---TCCCCGGGGATC
5	2-R5	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
6	3-R1	TGGCTGCAGGTCGAC	GT			----CCCCGGGGATC

Sample		5′ flanking region	P	N	P	3′ flanking region
7	3-R2	TGGCTGCAGGTCG--				---TCCCCGGGGATC
8	3-R3	TGGCTGCAGG-----		C		-GATCCCCGGGGATC
9	3-R4	CTTGCCAAGGTCG--		GTT		---TCCCCGGGGATC
10	3-R5	TGGCTGCAGGTCGAC				----CCCCGGGGATC
11	3-R6	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
12	4-R1	TGGCTGCAGGTCG--		T		GGATCCCCGGGGATC
13	4-R2	TGGCTGCAGGTCGAC	GT	AA		----CCCCGGGGATC
14	4-R3	TGGCTGCAGGTCGAC	GT	G		---TCCCCGGGGATC
15	4-R4	TGGCTGCAGGTCGAC				----CCCCGGGGATC
16	4-R5	TGGCTGCAGG-----				--ATCCCCGGGGATC
17	4-R6	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
18	6-R1	TGGCTGCAGGTCGAC	G	A		----CCCCGGGGATC
19	6-R2	TGGCTGCAGGTCG--				GGATCCCCGGGGATC
20	7-R1	TGGCTGCAGG-----				-GATCCCCGGGGATC
21	7-R2	TGGCTGCAGGTCGAC			C	GGATCCCCGGGGATC
22	7-R3	TG~~G~~CTGCAGGTC---		T		-GATCCCCGGGGATC
23	7-R4	TGGCTGCAGGTCGAC				-GATCCCCGGGGATC
24	7-R5	TGGCTGCAGG-----				--ATCCCCGGGGATC
25	7-R6	TGGCTGCAGGTCGA-				---TCCCCGGGGATC
26	7-R7	TGGCTGCAGGTC---				-----CCCGGGGATC
27	7-R8	TGGCTGCAGGTCGAC				----CCCCGGGGATC
28	7-R9	TGGCTGCAGGTC---				-----CCCGGGGATC
29	8-R1	TGGCTGCAGGTCGAC	G			---TCCCCGGGGATC
30	8-R2	TGGCTGCAGGT----				-GATCCCCGGGGATC
31	8-R3	TGGCTGCAGGTCGA-				---TCCCCGGGGATC

Sample		5′ flanking region	P	N	P	3′ flanking region
32	8-R4	TGGCTGCAGG-----			CC	GGATCCCCGGGGATC
33	8-R5	TGGCTGCAGGTCG--				GGATCCCCGGGGATC
34	8-R6	TGGCTGCAGGTCGA-		A		----CCCCGGGGATC

* The sequences were generated from 8 independent HEK293T cell transfections.
Germline sequence of non-recombined substrate recombination pGG51 plasmid, showing sequence at the 5' end of 12-RSS flanking region and the 3' end of 23-RSS flanking region.
Sample name; transfection number- hTdTS variant transfection.
Each dash (-) represents a single nucleotide deletion. P stands for P-nucleotide additions either from the 5' or the 3' flanking region.
N stands for N-nucleotide additions.

** Experimental conditions are as follows;
C= transfection with substrate pGG51 plasmid + pEBG-hRAG1 + pEBG-hRAG2 expressing vectors
W= transfection with wild type hTdTS pcDNA expressing vector + substrate pGG51 plasmid + pEBG-hRAG1+ pEBG-hRAG2 expressing vectors
A = transfection with A445T hTdTS pcDNA expressing vector + substrate pGG51 plasmid + pEBG-hRAG1+ pEBG-hRAG2 expressing vectors
L = transfection with L397S hTdTS pcDNA expressing vector + substrate pGG51 plasmid + pEBG-hRAG1+ pEBG-hRAG2 expressing vectors
R = transfection with R431C hTdTS pcDNA expressing vector + substrate pGG51 plasmid + pEBG-hRAG1+ pEBG-hRAG2 expressing vectors

*** Assignment of P-additions was considered prior to N-additions.