CHAPTER 1

INTRODUCTION
INTRODUCTION

1.1 Toxoplasma gondii

*Toxoplasma gondii* is a widespread obligate protozoan parasite of genus *Toxoplasma* which infects humans and animals, causing Toxoplasmosis. It was first identified in the tissues of birds and mammals. It infects about one-third world population. It is an intracellular apicomplexan parasite capable of infecting humans. It may lead to congenital birth defects and encephalitis in immuno-suppressed individuals (Remington JS et al., 1989; Luft BJ et al., 1992). The parasite can be transmitted by vertical transmission of the rapidly growing tachyzoites. It is mostly acquired in humans by the oral ingestion of the tissue cysts containing bradyzoites. Ingestion of oocysts containing sporozoites which are the products of the sexual cycle in cat intestines also results in transmission. Once cysts or oocysts are ingested, they divide rapidly to form tachyzoites. These differentiate into latent bradyzoites (reviewed in Kim K and Weiss LM, 2008). Type I strains are highly virulent and result in acute outbreaks. Both GT1 and RH strain belong to this class. Type II strain, ME49 which was first *T. gondii* strain sequenced, is comparatively less virulent in mice. Type III strain, VEG strain is a type III strain which is also less virulent in mice (reviewed in Kim K and Weiss LM, 2008).

The parasite genome is 63 Mb in size and consists of 14 chromosomes (Kissinger JC et al., 2003). This protozoan parasite has a number of structural similarities to *Plasmodium, Cryptosporidium* etc. In addition to the importance of *T. gondii* as a human and veterinary pathogen, the parasite accessibility to experimental manipulation made *Toxoplasma* as a useful model for the study of other disease causing microbial pathogens and apicomplexans like *Babesia, Cryptosporidium, Eimeria, Neospora, Plasmodium, Sarcocystis, Theileria* etc. It has thus become a
model organism for genetic studies. Owing to high degree of non-homologous end
joining (NHEJ), gene targeting efficiency is very low in this parasite. It is puzzling to
know that although NHEJ is prevalent in eukaryotes, no functional NHEJ has been
reported in any other protozoan parasite. *Toxoplasma gondii* is the only protozoan
parasite which harbors NHEJ. Gene knockouts and allelic replacement is a challenge
in *Toxoplasma* (Donald RG et al., 1998). High frequency of non-homologous
recombination can be exploited for insertional mutagenesis but at the same time,
targeted gene integration is complicated in this parasite (Donald RGK et al., 1995,
1996). In *Ku80* knock out genetic background, an increase of 300 to 400 fold was
observed in targeted gene replacement efficiency in *Toxoplasma gondii*. On the other
hand, integration of transgenes appears to occur exclusively by homologous
recombination in other protozoan parasites, like *Leishmania*, *Trypanosoma* and
*Plasmodium* (Kapler GM et al., 1990; Lee MG et al., 1990; Wu Y et al.,
1996).

1.2 DNA damage

DNA damage is inevitable. It is caused by ionizing radiation, UV light, replication
across a nick, oxidative free radicals, topoisomerase failures, inadvertent enzyme
action at fragile sites and mechanical stress (reviewed in Lieber MR., 2010). Figure 1
displays the possible DNA damaging agents. The DNA damage signaling cascade
finally reaches the effector molecules which triggers the following cellular processes-
DNA repair, cell cycle checkpoint activation, transcriptional program activation,
apoptosis and senescence (reviewed in Rupnik A et al., 2010).

DNA damaging agent, a type of alkylating chemical, methyl methane sulphonate
(MMS) exposure results in the methylation of DNA bases and generates double strand
breaks (DSBs) in a time and dose-dependent manner (Zhou C et al., 2006) upon
collision of progressing replication forks with SSB intermediates of base excision repair (Pascucci B et al., 2005; Wyatt M D et al., 2006).

1.3 DNA repair

DNA damage can occur in single or both the strands of DNA. Damaged bases or nucleotides can be repaired by excision repair mechanisms: base excision repair (BER) or nucleotide excision repair (NER). DNA mismatch is repaired via mismatch repair mechanism (MMR). Double stranded DNA breaks are more severe and difficult to repair. Double stranded DNA break repair mechanisms are of 3 types viz. homologous recombination (HR), non-homologous end joining (NHEJ) and single stranded annealing (SSA). Genomic integrity of cell depends on the fidelity of DNA repair. Figure 2 shows the consequences of the unrepaired breaks in an organism. In unicellular organisms, failure to repair DNA leads to cell death by blocking essential cellular functions; whereas in multicellular organisms, it results in apoptosis. Unrepaired DNA breaks result in genome instability, cancer, neuro-degeneration and other pathologies.
DNA damage

Sources

Exogenous
UV light exposure
Ionizing radiation
Chemicals

Endogenous
Replication errors
Cellular metabolism

Figure 1. Different sources of DNA damage.
Figure 2. Illustration showing the consequences of unrepaired DNA breaks.
1.4 Repair of damaged or mismatched bases

A1) Base excision repair of DNA (BER)

Base excision repair occurs when the DNA is oxidized, deaminated or alkylated. It is well conserved from prokaryotes to eukaryotes. In general, it is of 2 types- pol β dependent short (1 nucleotide) patch and PCNA dependent long (10-12 nucleotides) patch BER.

In general, DNA glycosylases first recognize and excise the damaged base from the sugar phosphate backbone of DNA by cleaving N-glycosidic bond between the target base and deoxyribose. This creates an abasic site (AP-apurinic/apyrimidinic site). A group of enzymes called AP endonucleases recognize the abasic site and make an incision at the 5′ or 3′ phosphodiester bond of the AP site. This generates a nucleotide gap. These enzymes belong to 2 categories viz. class I and class II. Flap endonuclease (FEN-1) resolves the single stranded DNA flap intermediates that form during long patch DNA base excision repair. Long patch BER involves DNA polymerase I and DNA ligase I. DNA polymerase I fills in the abasic site and DNA ligase I catalyzes the phosphodiester bond formation.

In yeast, AP site during short patch BER pathway is created by APN1. Rad27 (FEN-1) trims the DNA at the double strand break and polymerase ε synthesizes the new strand and Cdc9 carries out ligation step (Kelley MR et al., 2003). In mammals, APE1 creates AP site. β-pol with DNA ligase I or DNA ligase III/XRCC1 carries out polymerization and nick sealing (Kelley MR et al., 2003).

A2) Base excision repair of DNA (BER) in apicomplexa

Short patch BER eliminates 80% of the damaged bases in most of the eukaryotes. Exceptionally in Plasmodium falciparum, BER is predominantly by long patch pathway (Haltiwanger BM et. al., 2000). In P. falciparum, AP endonuclease (APN1),
replication factor C (RFC), DNA polymerase I and DNA ligase are involved in BER. Homologue of DNA glycosylase is apparently absent in *Plasmodium falciparum*. But, partial purification and characterization of DNA pol β-like enzyme was done and it was observed that it can repair a stretch of 3-5 nucleotides (Nunthawarasilp P et al., 2007). DNA glycosylase and class I AP endonuclease are apparently absent in the genome of *Plasmodium*. In *P. falciparum*, class II AP endonuclease Mg\(^{2+}\) dependent and independent pathways have been reported (Haltiwanger BM et. al., 2000). Mammalian cells contain only Mg\(^{2+}\) dependent AP endonuclease (HAP1, ExoIII) which indicates that the malarial parasite Mg\(^{2+}\) independent AP endonuclease might be an important chemotherapeutic target (Haltiwanger BM et. al., 2000). The PCNA-binding domain with amino acid residues 350-359 is highly conserved in *Plasmodium* (Casta LJ et al., 2007). PfFEN-1 showed endonuclease activity which was stimulated by divalent cations (Mg\(^{2+}\) and Mn\(^{2+}\)) and inhibited by monovalent ions (Casta LJ et al., 2007). The PCNA (proliferating cell nuclear antigen) binding site is internally located in the C-domain of PfFEN-1, rather than the extreme C-terminal end as seen in FEN-1 from other organisms. PfLig I has a variable N-terminal and a constant C-terminal end. DNA binding domain is present towards the inner side of N-terminal end (Buguliskis JD et al., 2007). Onyango DO et al., 2011 reported that *Toxoplasma gondii* possesses two apurinic/apyrimdinic (AP) endonucleases, TgAPE and TgAPN that function in DNA BER. The authors reported that the knockout of TgAPN was not possible and TgAPN is critical for the parasite to recover from the DNA damage. *Toxoplasma* genome database revealed the presence of homologues of DNA glycosylase, FEN-1, PCNA, DNA ligase I and DNA polymerase I.

**B1) Nucleotide excision repair**
It occurs when the pyrimidine dimers are formed in DNA due to the exposure to UV light and/or environmental carcinogens. In general, UvrA protein identifies the pyrimidine dimers due to UV irradiation. Two subunits of UvrA with one subunit of UvrB binds to the DNA and unwinds it around the damaged site. UvrA subunits leave; UvrB forms a stable complex with the DNA and UvrC subunit is recruited. UvrB nicks on one side and UvrC on the other side of the DNA lesion.

In yeast, it was shown that there was no preformed repairosome complex for damage recognition and dual incision of DNA lesion but the sequential assembly of proteins takes place. Rad14p recognizes the damaged DNA and forms a ternary complex with Rad1/Rad10 proteins (Guzder SN et al., 1993). It shows strong association with Rad1 but weakly associated with Rad10 protein. This ternary protein complex is referred to as nucleotide excision repair factor or designated as NEF-1. Rad1/10 proteins show endonuclease activity (Sung P et al., 1993; Tomkinson AE et al., 1993). Rad3 shows ATP dependent binding to the UV damaged DNA (Sung P et al., 1994) and it associates with Rad25 resulting in unwinding of the damaged duplex DNA. This facilitates the RPA to bind to the single stranded damaged DNA. Rad2 shows endonuclease activity (Habraken Y et al., 1993). Rad2 interacts with TFIIH and RPA and makes dual incision in the damaged DNA strand. In humans, genes encoding RPA, TFIIH, XPA, XPC, PCNA, RFC and ERCC1 are present. PARP1 is involved in single strand break repair mechanisms, BER or NER, together with XRCC1, DNA ligase III, as well as polynucleotide kinase (PNK), PCNA and FEN1 (Okano S et al., 2003; Frouin I et al., 2003).

B2) Nucleotide excision repair in apicomplexa

Interestingly, the homologues of UvrA, UvrB, UvrC, XPD, XPE and XPF proteins seem to be absent in *Plasmodium falciparum* and *Toxoplasma gondii*. The genome of
*P. falciparum* and *T. gondii* exhibited homologues of UvrD, XPA, XPB, XPC, XPG and CSA.

**C1) Mismatch repair**

It is a system for recognizing and repairing the erroneous insertion, deletion and misincorporation of bases that can arise during DNA recombination and DNA replication. It is also conserved from prokaryotes to eukaryotes. It is strand specific.

In eukaryotes, the newly synthesized strand is probably recognized by nascent strand discontinuities during replication (Holmes JJ et al., 1990). MSH2 first recognizes and initiates the mismatch repair process at the site of DNA lesion. MLH1 and PMS1 form a heterodimer complex. DNA polymerase synthesizes the correct bases and DNA ligase I seals the nicks to complete the mismatch repair process. PCNA (proliferating cell nuclear antigen) is a cofactor of DNA polymerase and it increases the activity of it. Single stranded binding protein binds to the unwound DNA.

In yeast, all the six MSH1 to MSH6 proteins are present. MutSa (MSH2/MSH6) and MutSβ (MSH2/MSH3) complexes recognize and bind to the mispaired bases in DNA (Acharya S et al., 1996, Habraken Y et al., 1996). Mlh3 and Pms1 are the MutL homologs in yeast. Three different exonuclease activities have been identified in yeast viz. Exo I shows 5′ to 3′ exonuclease activity; polymerases δ and ε show proofreading activity. PCNA stabilizes the MutSa and MutLa heterodimers at the mismatched sites. In humans, hMSH1 protein has not been reported till date. MutSa (hMSH2 / hMSH6) and MutSβ (hMSH2 / hMSH3) recognize the incorrect bases involved in pairing. MutSa binds to many base/base mismatches and one base pair insertion/deletion loop (IDL), whereas MutSβ binds to 2-4 basepair IDLs (Jiricny J, 1998, Kolodner RD and Marsischky GT, 1999). MutLa (MLH1/PMS2) acts as a
facilitator by coordinating all the events in mismatch repair. MutLβ (MLH1/PMS1) and MutLγ (MLH1/MLH3) complexes are less understood in humans. PCNA is required at the excision and resynthesis steps (Umar A et al., 1996, Gu L et al., 1998). EXOI/HEXI is the human homolog of Exo I. It interacts with MSH2.

C2) Mismatch repair in apicomplexa

MSH2 homologue is present in *Plasmodium falciparum*. MutH homolog is absent in *Plasmodium falciparum*. PfMLH has been cloned and biochemically characterized from *Plasmodium falciparum* 3D7 strain (Tarique M et al., 2012). It contains endonuclease and ssDNA-dependent ATPase activity and is expressed in schizont stage at a peak level in *P. falciparum* 3D7 and Dd2 strains (Tarique M et al., 2012). Bethke L et al., 2007 research group disrupted *P. berghei* MSH2-2 gene and found that this gene is not essential for parasite growth in both asexual and sexual life cycle. They identified an unusual complement of two mismatch repair genes, PfMSH2-1 and PfMSH2-2. Moreover PfMSH2-2 showed an elevation in mutation frequency and it was indirectly demonstrated that this gene played a role in mismatch repair. Antimalarial drug resistance can result from defective mismatch DNA repair (Castellini MA et al., 2011). Exo I with 5’ to 3’ exonuclease activity is present in the malarial parasite. In *Toxoplasma*, homologue of MutS DNA damage repair enzyme (TgMSH-1) is present. Lavine MD and Arrizabalaga G, 2011 reported that the antibiotic monensin caused cell death of *Toxoplasma gondii* through an unknown MSH-1 dependent cell cycle disruption.

1.5 Double strand break repair mechanisms

DNA double strand breaks (DSB) are mainly repaired either by homologous recombination (HR) pathway or by non-homologous end joining (NHEJ) mechanism.
When a DSB is formed, HR repair depends on searching of extensive homologous stretches of DNA; whereas NHEJ of broken DNA ends depends on little or no homology (Figure 3). Both pathways compete with each other for repair. The whole process of cell’s decision making on which pathway to proceed for repair is not yet known. The double strand break repair pathways in *E. coli* and *S. cerevisiae* have been used as the model systems to understand the repair pathways in other organisms.

Interestingly, the usage frequency of HR over NHEJ is different in different organisms (Figure 4). In higher eukaryotes, like in humans, NHEJ is predominant whereas in lower eukaryotes, like in *Saccharomyces cerevisiae*, HR is the major pathway. Interestingly, two lower eukaryotes, very closely related apicomplexan parasites; *Plasmodium falciparum* and *Toxoplasma gondii* show strikingly opposite choice of the repair pathway. *P. falciparum* apparently lacks NHEJ (based on the genome information) and shows HR and *T. gondii* uses NHEJ as the predominant pathway. *Plasmodium* is dependent on only homologous recombination in case of a double strand break repair in DNA. From the genome information, the NHEJ and SSA proteins are apparently absent in the malarial parasite (Gardner MJ et. al., 2002). In *T. gondii*, both HR and NHEJ pathways exist. This parasite surprisingly chooses NHEJ as its major repair mechanism (Fox BA et al., 2009; Huynh MH et al., 2009).

If Ku70/80 heterodimeric protein binds to DSB, the repair takes place by NHEJ. But if the ends are resected by 5' to 3' exonuclease, it forms long 3' single stranded DNA where in HR mediated repair occurs (Figure 5). In yeast, it was shown that ExoI and/or Sgs1 generate long stretches of ssDNA that are essential for HR (Mimitou EP and Symington LS, 2008). Chung W et al., 2010 demonstrated that about 2 to 10 Kb of ssDNA is generated depending on the kinetics of repair in yeast. Long ssDNA overhang inhibits recruitment of Ku protein and in turn incorporates recombination.
proteins. The first step in DNA resection where in the intermediate 3' tail is formed requires Mre11-Rad50-Xrs2 complex and Sae2, while the second step where the early intermediate is rapidly processed to generate the extensive 3' overhang employs exonuclease Exo1 and/or Sgs1 (Mimitou EP and Symington LS, 2008). In eukaryotes, within S and G2 phases of cell cycle, homologous recombination is very active because the two sister chromatids are directly adjacent and further, NHEJ is active in G1 phase of the cell cycle (reviewed in Lieber MR, 2010).
Figure 3. Illustration showing the two main DNA double strand break repair pathways: Homologous recombination (HR) and Non homologous end joining (NHEJ).
Figure 4. Illustration shows the DNA double strand break repair pathway preference in prokaryotes and eukaryotes.
Figure 5. Choice between HR and NHEJ pathways [Modified from Mimitou EP and Symington LS. DNA repair, 2009].
In DNA repair, Mre11 (Meiotic recombination 11 protein) is recruited very early to the broken ends. Mre11p is a multifunctional player and has several roles like recognition of DNA damage, binds to the broken chromosomal ends and starts the repair process, stabilizes replisome as well as activates the DNA damage checkpoint proteins. It exists as a MRX complex which is made up of two molecules each of Mre11, Rad50 (Radiation sensitive 50) and one molecule of Xrs2 (X-ray sensitive 2) protein. The human homolog of Xrs2 is Nbs1 (Nijmegen breakage syndrome 1). Mre11 protein is known to dimerize. The Mre11 is core member of the complex and interacts with itself and both Rad50 and Xrs2/Nbs1. Mre11p resects the broken DNA ends in 5′ to 3′ direction and tethers the broken ends in the form of MRX complex. It has 3′ to 5′ exonuclease and endonuclease activities that aids in processing of DNA ends during DNA repair and activation of cell checkpoint proteins (Jazayeri A et al., 2008; Lee JH and Paull TT, 2005; Paull TT and Gellert M, 1998). Studies from the *Xenopus laevis* extracts revealed that DNA resection results in production of ssDNA oligos that associate with the MRN complex and influence cell signaling protein, ATM activity (Jazayeri A et al, 2008). Mre11 binds duplexed DNA (Hopfner KP et al., 2002) and to the ends of the linearized DNA (de Jager M et al., 2001 and Usui T et al., 1998). Once recruited to the site of DNA damage, ATM physically interacts with Mre11 and this interaction stimulates ATM to phosphorylate the downstream target substrates (Bakkenist CJ and Kastan MB., 2003; Lee JH and Paull TT., 2004). Mre11p is also known to play a role in telomeric end processing (Haber JE et al., 1995, Deng Y et al., 2009). The work done by Williams B et al., 2005 indicate that the A and B DNA binding domains of MRE11 repress telomere rapid deletion. Recently, it was proposed that MRX or MR complex disrupts the G rich quadruplex telomeric DNA in the presence of ATPase activity of Rad50. This generates the
telomeric ends which are free for the recruitment of telomerase (Ghosal G and Muniyappa K, 2007). Mre11 nuclease and C-terminal DNA damage response functions are required for initiating yeast telomere healing (Bhattacharyya MK et al., 2008). Tittel-Elmer M et al., 2009 investigated that the MRX complex of yeast is recruited to replication forks during HU-induced pausing and it stabilizes the replisome independent of the S-phase checkpoint and Mre11 nuclease activity. This study revealed the molecular function of MRX in DNA replication. In higher eukaryotes, null mutations in components of Mre11 complex proved lethal (Luo G et al., 1999; Yamaguchi-Iwai Y et al., 1999; Zhu J et al., 2001).

A1) Homologous recombination

Homologous recombination is a process in which the genetic material is exchanged between two similar or identical strands of DNA. This repair mechanism is operational when the double strand break occurs in one of the homologous chromosomes. The main proteins involved are Rad50, Rad51, Rad52, Rad54, Rad55 and RPA (replication protein A). Rad51 is one of the key proteins involved in the HR process. It is produced in larger amounts so that it immediately binds to the DNA double strand break much before other repair proteins. Repair by HR involves 5’ to 3’ resection of the broken DNA ends, and generates 3’ single stranded DNA ends. These ends get coated by Rad51 to form the DNA-Rad51 filament. This filament searches for and base pairs with a homologous donor, forming a three stranded structure called displacement loop or D-loop. When both the ends of the broken DNA share homology with a sister chromatid, the repair takes place by gene conversion (GC). It involves the synthesis of short patch of new DNA. When homology is present with only one of the DNA DSB ends, the repair occurs by break-induced replication (BIR). The
presence of only one end can arise from stalled and/or broken replication forks (Haber JE et al., 1999) or at the telomeres that have lost their end protection (Lundblad V and Blackburn EH., 1993). When homology at both ends is present, BIR is outcompeted by GC (Malkova A et al., 2005); however what actually inhibits BIR is not clear. A DSB flanked by direct repeats can be repaired by another recombination pathway, single stranded annealing (SSA). The homologous sequences simply anneal with each other, the non-complementary tails are removed, new DNA synthesized to fill the single-stranded gaps and annealing occurs. GC, BIR and SSA are kinetically and mechanistically quite different from each other. Jain S et al., 2009 demonstrated that the initial steps of strand pairing events are similar in these three pathways. Their data suggested that a “recombination execution checkpoint” (REC) regulates the pathway choice. REC is sensitive to the location, orientation and distance between the homologous sequences used for the DSB repair.

RPA of yeast binds to DSB and is replaced by RecA homologue, Rad51. Moreover, RPA helps Rad51 in strand invasion and facilitates Rad51 binding to ssDNA and possibly by displacing dsDNA (Wang X et al., 2004). Strand invasion is the central step of homologous recombination which requires Rad51 protein. The recombinase activity of Rad51 takes place in three different phases- presynaptic, synaptic and postsynaptic phases. RAD52 gene family is grouped into three classes- class I consists of RAD50, MRE11 and XRS2 genes (involved in processing of breaks); class II consists of RAD51, -52 and -54 (plays a role in homology search and DNA pairing) and class III consists of RAD55 and -57 genes (part of protein complex). Knockout of Rad51 showed embryonic lethality in mice (Tsuzuki T et al., 1996). This suggests that RAD51 is indispensable for the homology mediated DNA repair (HDR) pathway. It interacts with RPA, Rad51, Rad54 and Rad55 proteins. Rad51 self assembles on the
single and double stranded DNA to form the nucleoprotein filament. ATP binding is required for the nucleofilament formation. It forms helical filament similar to RecA. Replication protein A (RPA) bound ssDNA is recognized by Rad52 and Rad51-Rad52 complex replaces RPA. Rad51 with its motor activity searches for homologous sequences between ssDNA and dsDNA. Rad55-Rad57 heterodimer binds to the ssDNA and shows recombination mediator activity. Rad54 acts as a translocase and DNA helicase in this recombination pathway. Rad51 binding to ssDNA requires Rad52p; the absence of Rad55 delays Rad51 binding and further Rad54 is required at or before the initiation of DNA synthesis after synapsis occurs (Sugawara N et al., 2003). The kinetics of loading of Rad51 recombinase to the single HO-induced double strand break, DNA resection, homology search, strand invasion and initiation of repair DNA synthesis at a time point of every 10 min of homologous recombination during MATa switching in Saccharomyces cerevisiae was elucidated by Hicks WM et al., 2011.

Benson FE et al., 1994 research group first identified human Rad51. They reported that the purified HsRad51 protein binds to both ssDNA and dsDNA and shows DNA-dependent ATPase activity. Compared to RecA, it has got low ATP hydrolysis activity. In humans, HsRad51 forms a nucleoprotein filament with the DNA to be repaired. In humans, there is no evidence that RPA effects HsRad51 (Gupta RC et al., 1997). Two other Rad51 related proteins viz. Xrcc2 and Xrcc3 are known to express in human cells. Both of these interact with Rad51 (Liu N et al., 1998; Pierce AJ et al., 1999).

For DNA strand exchange, DNA binding ability by E.coli RecA is essential and there is no requirement for ATP hydrolysis. An allosteric effector consisting of ADP and Aluminium fluoride (ADP-AlF$_4^-$) covalent complex can induce the binding of RecA
to DNA and can promote strand exchange up to 800 to 900 base pairs. The needed free energy for strand exchange is supplied from binding of ligand ATP to DNA-protein RecA complex and the ATP hydrolysis destroys the effector ligand (Kowalczykowski SC., 1995). Zaitseva EM et al., 1999 reported on the DNA binding properties of Saccharomyces cerevisiae Rad51. This protein binds to both the single stranded and double stranded DNA in an ATP and Mg\(^{2+}\) dependent manner and one protein monomer binds to three or four nucleotides. They found that ScRad51 binds to ssDNA even in absence of ATP with the stoichiometry of one protein monomer for every seven to nine nucleotides. These two binding modes are not interconvertible and this behavior is similar to RecA protein. ScRad51 showed ssDNA dependent ATPase activity and strand exchange activity (Sung P., 1994). It was shown that the low efficiency of strand exchange activity of ScRad51 in \textit{in vitro} experiments was due to the enhanced DNA binding ability of this protein.

Site-directed mutagenesis of highly conserved residues (K133R) in ATPase domain of human RAD51 and strand exchange assay revealed that human Rad51 does not hydrolyse ATP for recombinalional repair. Weak mutant Rad51-K191R in human cannot hydrolyse ATP but capable of ATP binding and DNA strand exchange (Morrison C et al., 1999). Stark JM et al., 2002 research group generated mutants of hRAD51, hRAD51-K133A and hRAD51-K133R by dominant negative approach which were defective in ATP binding and ATP hydrolysis respectively. Mouse embryonic stem cells that express hRAD51-K133R were obtained which were hypersensitive to mitomycin C and ionizing radiation; showed decrease in spontaneous sister chromatid exchange and a defect in HDR pathway. Stark JM et al., 2004 reported that mammalian HDR and SSA pathways are inversely related. Expression of RAD51 ATP binding defective mutant drastically increased SSA and
Rad52−/− mouse cells showed no defect in HDR but SSA was decreased. The authors demonstrated that the genetic interplay of DSB repair factors is essential to limit the mutagenic potential of the repair process. Agmon N et al., 2009 showed that intrachromosomal homologous chromosomes are always preferred as donors for DSB repair in yeast. In yeast, different protein-protein interactions have been identified viz. Rad51p-Rad52p (Shinohara A et al., 1992), Rad51p-Rad54p (Clever B et al., 1997), Rad51p-Rad55p, Rad55p-Rad57p (Hays SL et al., 1995; Johnson RD and Symington LS, 1995) and Rad52p-RPA1p (Firmenich AA et al., 1995). In yeast, N-terminal amino acids of Rad54A (1-326 residues) interact with Rad51 and it appears that the C-terminal amino acids of Rad54B (327-898 residues) do not interact with Rad51 (Clever B et al., 1997). Strong interaction between Rad51 and Rad55 was identified. No interaction detected between Rad51 and Rad57 in yeast (Johnson RD and Symington LS, 1995). Rad51p, Rad52p, Rad55p, Rad57p and RPA participate in the formation of the “recombinosome” complex during double strand break repair in yeast (Hays SL et al., 1995; Johnson RD and Symington LS, 1995). Larger subunit of RPA interacts with Rad52p in the recombinosome (Hays SL et al., 1995).

A2) Homologous recombination in apicomplexa

PfRad51 protein was identified, biochemically characterized and its functional role in HR repair was demonstrated (Bhattacharyya MK et al., 2003, 2005). PfRad51 monomer binds to 3 nucleotides. PfRad51 shows 66-75% homology within the catalytic region of the Rad51p of human, yeast and other protozoan parasites viz. Trypanosoma and Leishmania (Bhattacharyya MK and Kumar N., 2003). ATPase activity of PfRad51 provides the requisite energy for the strand migration and homologous pairing. In Plasmodium, Rad54 acts as an accessory protein and is also incorporated after Rad51. Compared to RecA mediated homologous recombination,
PfRad51-PfRad54 show slower ATPase activity and lower strand exchange (Bhattacharyya MK et al., 2005). The purified recombinant PfRad51 protein exhibited three-strand exchange activity in presence of non-hydrolysable ATP-γS which showed ATP binding. However the kinetics of strand exchange was slower in presence of ATP-γS (Bhattacharyya MK et al., 2005). This is in accordance with the behavior of Saccharomyces cerevisiae and human Rad51 proteins. In presence of ATP and 10 mM MgCl₂, the kinetics of strand exchange reaction was slower for PfRad51 when compared to *E.coli* RecA. The products of the reaction appeared after 10 and 20 min for PfRad51. The efficiency of conversion of linear dsDNA into nicked circular dsDNA was lower in PfRad51 when compared to bacterial RecA (Bhattacharyya MK et al., 2005).

Homologues of Rad52 and RPA are apparently absent in the genome of *P. falciparum*. One possibility is that homologous recombination of *Plasmodium* does not need RPA and Rad52 proteins and another possibility is that the sequence of these 2 proteins is different. From the genome information of *Toxoplasma gondii*, putative orthologues of Rad51 and Rad54 are present. This suggests that HR is operational in this parasite.

**B) Single stranded annealing**

It is operational when the homology is limited to only few bases. Here, large chunks of DNA are eliminated at the double-strand break. SSA can occur with a small homology of 30 bp and can go on with a maximum of 200-400 bp. The efficiency of SSA improves with the increase in the homology. Rad52p recognizes and binds to the DNA lesion. It scans the damaged DNA for small stretches of homology. The damaged DNA is cleaved at the 5' region by the action of FEN-1. This leaves 3'
overhangs. Now, the DNA is resected so that the regions of homology can align with each other. Further endonuclease activity is carried out by Rad1/Rad10 proteins wherein the overlapping flaps are removed. The ultimate step of gap filling and nick sealing is accomplished by DNA polymerase I and DNA ligase IV respectively.

In yeast, Rad52, FEN-1 (Flap endonuclease-1) and Rad1/Rad10 proteins carry out SSA mechanism. There is evidence that the mismatch repair Msh2/Msh3 proteins are involved in cutting the small portions of overlapping DNA (Paques F et al., 1997, Sugawara N et al., 1997). In humans, the proteins operational in this pathway are RAD50, MRE11, NBS1, ERCC1 and XPF. In humans, XPF/ERCC1 perform the function of yeast Rad1/Rad10. It is yet to be investigated if hMsh2/hMsh3 proteins show dual role by involving in SSA of humans. Till now there is no report on SSA in any apicomplexa parasite.

C1) Non-homologous end joining

It is simple joining of the two broken ends of the double stranded DNA. Error incorporation is more in this method as it involves little or no homology in rejoining the broken ends. Small insertions or deletions in DNA are possible. NHEJ is mainly of two types- Classical NHEJ (C-NHEJ) and alternate NHEJ (A-NHEJ). A-NHEJ pathway is poorly understood. It is also called as microhomology-mediated end joining (MMEJ) or backup NHEJ (B-NHEJ).

In yeasts, the NHEJ core proteins are Ku70/80, XRCC4, DNA PKcs and DNA ligase IV. In budding yeast, Mre11p-Rad50p-Xrs2p protein complex shows exonuclease and endonuclease activity. In fission yeast, the absence of Mre11p has little effect. NHEJ is not a very efficient repair pathway in yeast cells. XRCC4 and DNA ligase IV are recruited for the ligation step. Domain analyses of Ku protein showed that both the N-
terminal and C-terminal domains are non-essential for its interaction with XLF protein. Only heterodimeric domain is involved. The last 10 amino acids of the C-terminal domain of XLF protein are crucial for XLF-Ku interaction. Though the N-terminal globular head is the site for interaction of XRCC4, the absence of XLF-Ku interaction reduced the interaction of XLF-XRCC4 (Yano K et al., 2011). In humans, the main proteins involved in classical NHEJ of humans are Ku70/80, XRCC4, DNA PKcs (DNA protein kinase catalytic subunit), MRX complex (Mre11/Rad 50/NBS1 proteins) and DNA ligase IV. There is another Ku-independent NHEJ pathway which is the backup NHEJ. It involves DNA ligase III, PARP1 and histone 1 (Iliakis G, 2009). Ku protein exists as a heterodimer of Ku70 and Ku80 and binds to the damaged DNA. Ku70/80 serves two functions. Firstly, it interacts with the damaged DNA and tethers the two broken ends. Secondly, it serves as a platform for the recruitment of additional NHEJ factors (reviewed in Lieber MR, 2010). Ku heterodimer interacts with the catalytic subunit of DNA-PK (Sipley JD et al., 1995). Artemis is a nuclease which is the substrate for DNA-PK. Further, its activity is modulated by DNA-PK. DNA-PKcs becomes active as a kinase and phosphorylates itself and Artemis. Phosphorylation of Artemis has no functional significance but autophosphorylation results in conformational change in DNA-PKcs resulting in its translocation; thereby exposing the DNA ends at the damaged site to Artemis. Then, Artemis cleans up these ends by nuclease activity and makes them suitable for ligation. Mre11-Rad50-Nbs1 complex is involved in the processing of the 3′ ends. Rad50 has ATPase activity and Nbs1 acts as a sensor for DNA damage response. FEN-1 enzyme processes the 5′ ends. DNA ligase IV is essential for non-homologous end joining (Timson DJ et al., 2000).
In humans, both the pathways can co-exist, although C-NHEJ is the major repair mechanism and widely characterized as a Ku protein dependent pathway. NHEJ constitutes about 70% whereas HR constitutes 30% of the total double strand break repair pathways in humans (Liang F et al., 1998).

Early studies in SV40 DNA substrates tested in monkey cells revealed efficient end joining in the absence of C-NHEJ key factors and this provided evidence for A-NHEJ pathway (Roth DB et al., 1985). It is shown that Ku protein strongly represses alternate end-joining pathway (Bennardo N et al., 2008 and Fattah F et al., 2010). However, lower eukaryotes like *E. coli* use ligase A-dependent alternate end joining method for repairing double strand breaks and for horizontal gene transfer. *In E. coli*, HR is the major repair mechanism but A-NHEJ is seen in a very low proportion (Chayot R et. al., 2010). Aniukwu J et al., 2008 showed the fidelity of NHEJ mediated repair of the two broken DNA ends in *Mycobacterium* by DNA sequencing. Hela cell extracts depleted of Ku and DNA-PKcs rejoined blunt ends and homologous ends with 3' or 5' single stranded protruding ends with similar efficiency but addition of Ku suppressed joining of blunt ends and homologous ends with 3’ overhang (Wang H et al., 2003). This study presented the biochemical evidence for the presence of Ku-independent backup NHEJ pathway. Audebert M et al., 2004 found an end joining activity in mammalian cells independent of the DNA-PK/XRCC4-ligase IV complex but that actually required the synapsis activity of PARP1 and ligation activity of XRCC1-ligase III. From the human cell extracts, polynucleotide kinase (PNK) was identified as the 5’-DNA kinase associated with the poly (ADP-ribose) polymerase-1 (PARP1) dependent end-joining pathway (Audebert M et al., 2006). Wang H et al., 2005 identified DNA ligase III as the main component of backup NHEJ pathway in extracts of mammalian cells. Liang L et al., 2008 showed that human DNA ligase I or
III are required in micro-homology mediated end joining (MHEJ), and in contrast DNA ligase IV is not required for MHEJ. Their data indicated that the MHEJ and NHEJ pathways require different ligases. Ligase III functions with XRCC1 as a complex and is regulated by PARP1 (reviewed in Mladenov E and Iliakis G, 2011). Cheng Q et al., 2011 established in human cell lines that Ku is the main factor that counteracts the mobilization of PARP-1 and MRN to the damaged chromatin and also limits the synthesis of poly (ADP-ribose) (PAR) and ssDNA formation in response to DSBs. They showed that there is a shift to B-NHEJ in case of C-NHEJ defective condition. PARP-1 binds to DNA DSBs in direct competition with Ku and PARP-1 function is essential in backup NHEJ (Wang M et al., 2006).

Couëdel C et al., 2004 studies in mice revealed that the double knockout Rad54−/− ku80−/− was severely compromised in survival than single knockouts. They were very sensitive to DNA damaging agents and ionizing radiation. It implied that cooperation between two DNA DSB repair pathways was required for survival and genomic integrity in the animal.

**C2) Non-homologous end joining in apicomplexa**

NHEJ apparently seems to be absent in *Plasmodium falciparum*. The parasitic genome does not give any evidence for the existence of the homologues of NHEJ proteins.

Fox BA et al., 2009 first reported that a significant KU-dependent NHEJ pathway is operational in *Toxoplasma gondii*. Knockout at *KU80* locus disrupts the NHEJ DSB repair pathway. *Toxoplasma* genome revealed the genes encoding *KU70* (50.m03211), *KU80* (583.m05492), DNA ligase IV (TGGT1_073840) and DNA-dependent protein kinase (57.m01765) components of eukaryotic NHEJ (www.toxodb.org). *KU80* knockouts retained normal tachyzoite growth and were
viable in murine infection. They exhibited increased sensitivity to DNA double strand breaks induced by phleomycin or γ-irradiation.

**D) DNA double strand break repair in other protozoan parasites**

*Leishmania major* Rad51 exhibited ATPase activity and showed magnesium dependent binding to both single stranded and double stranded DNA. When the *Leishmania* promastigotes were treated with phleomycin, elevated expression of LmRad51 at mRNA and protein level was observed (McKean PG et al., 2001). Kinetoplastids, like *Trypanosoma* and *Leishmania* species, naturally exhibit efficient HR of exogenous DNA (Cruz A et al., 1990; Eid J et al., 1991). *Trypanosoma brucei* Ku70/80 functions in telomere maintenance and regulation of telomerase (Janzen CJ et al., 2004) and Ku dependent NHEJ was not observed in *T. brucei* cell extract (Burton P et al., 2007; Glover L et al., 2008). Burton P et al., 2007 showed from *T. brucei* cell extracts that Ku independent end joining of linear DNA molecules occurs based on sequence microhomology. Further based on the bioinformatic analysis, it was found that XRCC4 and DNA ligase IV homologues may not be encoded by the genome of *T. brucei*, raising the possibility that diverged ligase factors may be present.

### 1.6 Gene targeting in different organisms

The chinese hamster cells defective in *ku80* were shown to exhibit gene targeting frequencies comparable to that of wildtype cells (Liang F et al., 1996). Further mouse embryonic stem cells defective in *ku70, Xrcc4* or *DNA-PKcs* did not exhibit increased gene targeting efficiencies (Pierce AJ et al., 2001; Dominguez-Bendala J et al., 2006). Iiizumi S et al., 2008 examined the frequencies of random and targeted integration in
chicken DT-40 and human Nalm-6 cell lines. Loss of NHEJ led to increased targeted integration in chicken DT-40 cell lines. However, in human Nalm-6 cell lines, the results demonstrated that NHEJ is not the sole mechanism of random integration. Nevertheless, the authors presented evidence that NHEJ inactivation can lead to enhanced gene targeting through a reduction of random integration.

Ninomiya Y et al., 2004 research group identified and disrupted *Neurospora crassa* Ku70 and Ku80 homologues mus-51 and mus-52 genes and studied the frequency of the integration of the exogenous DNA into the genome of this disruption strains. They found that the suppression of NHEJ increases the frequency of homologous recombination. In order to investigate the relationship between the targeting frequency and the length of the homologous sequence, 50, 100, 500 and 1000 base pairs homologous to 5′ and 3′ flanking ends of the mtr gene was replaced by bar gene. The authors concluded that more than 1000 bp homology was required for complete gene replacement in *Neurospora crassa*.

Gene targeting studies in *Plasmodium falciparum* show that the targeted gene reaches right locus. This is in accordance as HR is the main DNA DSB repair mechanism in *P. falciparum*. Owing to very high NHEJ, it is mostly random gene integration in *Toxoplasma gondii* (Fox BA et al., 2009; Huynh MH et al., 2009). However, the efficiency of gene targeting in these organisms is very low. Table 1 shows the accuracy and efficiency of gene targeting in *Plasmodium falciparum* and *Toxoplasma gondii*. 
Table 1) The gene integration and the efficiency of gene targeting in apicomplexan parasites:

<table>
<thead>
<tr>
<th>Organism</th>
<th>Efficiency of gene targeting</th>
<th>Gene integration</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Plasmodium falciparum</em></td>
<td>Very low</td>
<td>accurate</td>
</tr>
<tr>
<td><em>Toxoplasma gondii</em></td>
<td>Very low</td>
<td>random</td>
</tr>
</tbody>
</table>
1.7 Significance of studying homologous recombination mechanism in *T. gondii*:

Understanding DNA break repair mechanisms can help in eradication of the deadly diseases caused by the pathogens. In an effort to curb Toxoplasmosis in humans, we tried to understand the DNA double strand break repair mechanisms in *Toxoplasma gondii*. To address why *Toxoplasma gondii* chooses NHEJ, we identified and cloned *MRE11, KU80* and *RAD51* genes of *T. gondii* and in order to understand the interplay between these three proteins during DSB repair. Understanding the functional role of TgMre11 would give insights on pathway choice as it binds to the broken DNA ends much before the decision is made between HR and NHEJ. To understand NHEJ mechanism, we choose the central player TgKu80 for characterization. Further to understand HR pathway and the targeted gene disruption in this parasite, we choose the key player TgRad51 for our study. We have cloned, purified and characterized TgRad51 biochemically as well as genetically. *T. gondii* and *P. falciparum* show surprisingly opposite choice in DNA repair pathways. *Plasmodium* genome apparently lacks the key components of NHEJ pathway and exhibits high frequency of homologous recombination. Here we report mechanistic insights for differential repair choices between two closely related lower eukaryotes. Though, TgRad51 and PfRad51 have 82% identity in their catalytic domain, they differ markedly in ssDNA dependent ATP hydrolysis activity. Since it is extremely difficult to perform the functional studies in *Toxoplasma*, we choose yeast as a surrogate model system. We hypothesize that the compromised ATPase activity of TgRAD51 leads to inefficient gene targeting and poor gene conversion efficiency in *T. gondii*.

The interplay between genomic investigations and those targeted at chemotherapy is essential for new drug designing to treat Toxoplasmosis. The overall goal of my PhD
research was to understand the DSB repair pathways in T. gondii by comparative analysis and gain evolutionary insights into the choice of pathways during DNA DSB repair.
1.8 Objectives of the study

Toxoplasma gondii is a lower eukaryote which exceptionally shows preference to non-homologous end joining during DNA double strand break repair. In contrast, in all other known eukaryotes, homologous recombination is the major DSB repair pathway. Plasmodium falciparum, a closely related apicomplexan parasite to T. gondii and Saccharomyces cerevisiae are the best examples of lower eukaryotes which prefer HR. Recent work done by two research groups (Huynh MH et al., 2009; Fox BA et al., 2009) revealed that the non-homologous end joining pathway is the predominant pathway in T. gondii. The authors had shown that only in the Ku80 (an important protein recruited very early during classical NHEJ repair) knockout background, homologous mediated gene targeting studies were successful and increased by several fold. However, it does not increase the integration efficiency at the correct locus (Fox BA et al., 2009). We wanted to know why Toxoplasma gondii, despite being a lower eukaryote chooses NHEJ. In order to investigate this, we choose TgMre11 protein for our study which is recruited within minutes of DNA damage and it binds to the broken ends much before the decision is made between HR and NHEJ. To understand NHEJ mechanism, we choose the central player TgKu80 for characterization. Further to understand HR pathway in this parasite, we choose the key player TgRad51 for our study. Inefficient gene targeting can be due to weak ATPase activity which results in inefficient motor activity and poor homology searching. We speculated that the ATPase activity of Toxoplasma gondii Rad51 was weak which led to inefficient gene targeting and poor gene conversion efficiency in Toxoplasma gondii.
Questions addressed

1) Why *Toxoplasma gondii*, despite being a lower eukaryote chooses NHEJ over HR?
2) What determines the pathway choice in this organism?
3) How does the cell enable the chosen pathway once the decision is made?

Aims of the Ph.D work

i. Why *Toxoplasma gondii*, despite being a lower eukaryote chooses NHEJ over HR? In this regard, we have chosen three proteins TgMre11, TgKu80 and TgRad51 for characterization. We wanted to understand the interplay between these three proteins.

ii. Why is the gene targeting less efficient in *Toxoplasma gondii*? Is it because of less efficient TgRad51 protein. Our approach was to clone, express and purify TgRad51 protein and we tried to look at its ATPase activity.