Summary

Repair of DNA breaks is essential for maintenance of genomic integrity. DNA double-strand breaks (DSBs) are considered as the most harmful DNA lesions within the cells which if left unrepaired or are misrepaired, can lead to a wide variety of genetic alterations, chromosomal rearrangements either culminating in oncogenic transformation or apoptosis. Therefore, the cell adopts two major pathways, Homologous recombination (HR) and nonhomologous end joining (NHEJ), in order to repair these DSBs thereby maintaining genome stability. HR mediated DSB repair is error-free and is functional during the S and G2 phases of cell cycle and requires the participation of a sister chromatid to act as a template in the repair process. However, in higher eukaryotes, DSBs are primarily repaired by nonhomologous DNA end joining (NHEJ), which is an error-prone DNA repair mechanism and is active throughout the cell cycle. During NHEJ, KU70/KU80 heterodimer binds to the broken DNA ends and recruits DNA-PKcs, Artemis, and Pol μ or λ to the repair site, resulting in processing of broken DNA ends. Following the end processing, ligation occurs with the help of Ligase IV, XRCC4, XLF and PAXX complex.

In general, NHEJ is considered as a DSB repair pathway that does not have a sequence preference and hence thought to be random. Using an oligomer based, cell-free assay system, we evaluated the role of DNA sequence during the end-to-end joining of broken DNA, both at the overhangs and at its flanking regions. Using radiolabeled double-stranded oligomeric DNA substrates, possessing DSBs with different overhang sequences, we observed that there is a distinct difference in joining efficiency of mammalian extracts (rat testis, rat lungs, rat brain and rat heart) when overhangs with G:C rich sequences were studied. In contrast, we observed an enhanced efficiency in DNA end joining when A:T-rich flanking sequences were examined. Specifically, when flanking sequences were homopolymers of A:T or sequences enriched with A:T (adenines and/or thymines), joining efficiency was higher compared to that of random or G:C-rich flanking sequences.

In order to investigate the factors that contribute towards the observed sequence specificity of NHEJ, we have overexpressed and purified one of the essential factors involved during NHEJ, the Ligase IV/XRCC4 complex. Ligase IV has a multidomain architecture, consisting of a conserved DNA binding domain at N-terminus and a tandem BRCT domain at C-terminus. The central catalytic domain comprises of adenylation (AD) and oligo-binding (OB-fold) domains. The N-terminal DNA binding domain (DBD) of Ligase IV is crucial for its interaction with DNA. Two hypomorphic mutations (A3V and T9I) in the DBD are observed in

Ligase IV syndrome patients. Previous studies suggested that recruitment of Ligase IV/XRCC4 complex is dependent on KU protein complex. Interestingly, we observed that binding of Ligase IV/XRCC4 complex can occur independent of KU70/80, however, in a sequence dependent manner. In the present study, we uncover the sequence preference of Ligase IV/XRCC4 complex for the first time and show that it can bind directly to DNA containing DSBs, when flanked with A:T rich sequence in a KU independent manner irrespective of the sequence of the overhangs. All DNA substrates possessing A:T rich flank sequences investigated in the current study, including the one derived from human genome showed preferential binding to Ligase IV/XRCC4, with a binding constant that was ~10 fold higher than that of G:C rich substrates. Gel mobility shift assays, in conjunction with shift-western blotting assay using various radiolabeled double-stranded oligomers and purified Ligase IV/XRCC4 complex revealed stable DNA-protein complex formation in the absence of KU70/80 heterodimer. Importantly, SCR7, a well-known Ligase IV inhibitor, inhibited the recruitment of Ligase IV/XRCC4 in a concentration dependent manner. The DNA:protein complex formed was resistant to DNase I digestion. In addition, biolayer interferometry (BLI) studies using biotinylated double-stranded oligomers possessing A:T or GC rich sequences flanking the DSBs and purified proteins demonstrated that both Ligase IV/XRCC4 complex and Ligase IV could bind strongly to the immobilized double-stranded oligomer, in a sequence dependent manner (KD values were as low as in nanomolar range for AT rich flanking DNA substrates), unlike XRCC4 (KD= 4.12 ± 0.924 µM). Furthermore, BLI results also revealed that the binding of DNA binding domain (DBD) of Ligase IV was restricted to biotinylated polynucleotide of thymines or AT rich flanking ds DNA substrates. This binding got abrogated when point mutations seen in Ligase IV syndrome patients were introduced to DBD, thereby increasing the KD values many folds. Further, immunodepletion of KU70 protein from mammalian tissue extracts and *ex vivo* knockdown of *Ku70* gene in mammalian cells reduced the joining efficiency of a random DNA substrate, while the NHEJ efficiency of DNA substrates remained unperturbed, when specific sequences were present in the flank region. However, *in vitro* ligation assays with Ligase IV immunodepleted cell free extracts exhibited reduced NHEJ efficiency irrespective of the substrates used. These observations suggest the unique DNA sequence dependence of Ligase IV involved during classical NHEJ repair pathway, which is KU-independent.

The requirement of process of ligation catalysed by any of the three mammalian ligases (Ligase I, III or IV) is absolute, during physiological processes such as DNA replication, recombination and in almost all DNA repair pathways. This makes DNA ligases as attractive therapeutic targets to treat cancer. All three DNA ligases share a high degree of homology both

at structural and functional levels (Ligase IV shares 13% and 15% sequence identity to Ligase I and Ligase III, respectively). They possess a well conserved catalytic domain but differ particularly in the DNA recognizing well conserved DNA Binding domain (DBD), that if bound by small molecules, the joining activity of ligases might get severely impaired.

Based on homology modelling, and considering the anti-tumor properties of a well-known Ligase IV inhibitor, SCR7, we describe its first water soluble, auto cyclized and oxidized form, known as Sodium salt of SCR7-Pyrazine (Na-SCR7-P). Na-SCR7-P exhibited enhanced bioavailability, unlike the other DMSO-soluble Ligase inhibitors. In the present study, we found that like its parental compound (SCR7), Na-SCR7-P, also inhibited NHEJ in a Ligase IV dependent manner. However, unlike SCR7, it blocked joining catalysed by all three ligases *in vitro*, making it as an ideal tool for cancer therapeutic studies, as it may target multiple DNA transaction processes within the cancer cells. In depth studies revealed that Na-SCR7-P treatment resulted in reduction of mitochondrial membrane potential and activation of apoptosis culminating in cell death in various cancer cell lines. Importantly, administration of Na-SCR7-P led to significant reduction in tumor growth from 12th day of treatment and its impact was significantly higher than previously described SCR7, which predominantly targets Ligase IV within cells. Antitumor activity of Na-SCR7-P in mice resulted in enhanced lifespan, with minimal side effects. In addition, *in ovo* chorio-allantoic membrane assay revealed the potent anti-angiogenic property of Na-SCR7-P. Thus, we successfully identified another potent DNA ligase inhibitor, Na-SCR7-P that can potentially be used as a strategy for cancer treatment, owing to its water solubility.

In addition, based on various biochemical and biophysical screening approaches, we identified two prospective DNA Ligase I inhibitors, SCR17 and SCR21. Considering the indispensable role of DNA Ligase I during physiological processes such as DNA replication, repair and recombination, DNA Ligase I is considered as an important target for cancer therapy as it can impede proliferation of cancer cells upon treatment with specific small molecule inhibitors. Both the inhibitors blocked the ligation of nicks on DNA catalysed by cell-free extracts or purified Ligase I in a concentration-dependent manner. Docking studies in conjunction with biolayer interferometry and gel shift assays revealed that both SCR17 and SCR21 can bind to Ligase I, particularly to the DNA Binding Domain of Ligase I with KD values in nanomolar range (39 nM ± 8.88 nM and 42 nM ± 12.49 nM, respectively). The inhibitors did not show significant affinities towards DNA Ligase III/XRCC1 and DNA Ligase IV/XRCC4. Further, addition of DNA Ligase I could restore the joining, when the inhibitors were treated with testicular cell-free extracts. *Ex vivo* studies using multiple assays showed that even though cell death was limited

in the presence of inhibitors in cancer cells, their proliferation was certainly compromised. Hence, we identify two promising DNA Ligase I inhibitors, which can be used in biochemical and cellular assays, and could be further modified and optimized to target cancer cells.

In summary, the present study describes a sequence specific mechanism of recruitment of Ligase IV/XRCC4 to the broken DNA ends, which will have implications in understanding the sequence preference of NHEJ, in mammalian cells. In addition, we successfully developed novel DNA Ligase I inhibitors (SCR 17 and SCR21) and, a novel, water-soluble inhibitor (Na-SCR7-P) of NHEJ pathway which can further be improved as novel therapeutic strategies for sensitizing cancer cells to DSBs.