Understanding the Molecular Mechanism Behind Fragility of *BCL11B* in T-cell Acute Lymphoblastic Leukemia

Introduction

T-cell Acute Lymphoblastic Leukemia (T-ALL) involves the leukemogenic transformation of T cells caused by the accumulation of different types of genomic alterations, which threatens to disrupt the genomic stability in cells (1). At the genetic level, T-ALL is primarily characterized by chromosomal rearrangement in the transcription factors *TLX1 (HOX11), TLX3 (HOX11L2), LYL1, TAL1, MLL,* and activating mutations in the *NOTCH1* gene (1, 2). Genes coding for T-cell transcription factors like GATA-3, c-MYB, members of the E2A/HEB family, BCL11B etc. undergo frequent mutations which leads to abnormal proliferation of immature thymocytes (2, 3). The key transcription factor known to play a role in maintaining identity and survival of T cells is BCL11B, which is highly expressed during double-negative (DN) stages of T cell development (4, 5). It keeps the T cells in a suppressed state by modulating the expression of FOXP3, IL-10, proinflammatory cytokines and promotes the differentiation of mature T cells (4, 6). However, frequent missense mutations or deletion of *BCL11B* in human T-ALL cases confers natural killer-cell like properties to T cells, resulting in increased proliferation and survival of immature T lymphocytes (7, 8). These mutations disrupt the structure of the zinc finger domains and prevent the binding of the transcription factor to DNA (8, 9). In addition to this, *BCL11B* has also been implicated in the development of T-cell lymphomas by its frequent involvement in various translocations, *e.g.,* t(5;14) (q35;32), making it one of the key tumor suppressor genes to be deregulated in T-ALL (10, 11). However, the reason behind the generation of such deleterious driver mutations in *BCL11B* gene, which possibly drives the initiation of T-ALL, is not understood.

The current study explores the molecular mechanism behind the generation of deleterious point mutations in the exon 4 of the *BCL11B* gene, deregulation of which contributes to the pathogenesis of T-ALL. Using various biochemical and biophysical assays, we show that exon 4 of *BCL11B* can fold into multiple non-B DNA structures, such as G-quadruplex DNA, triplex DNA, and hairpin DNA. Several studies suggest that one of the major causes of genomic instability is the prevalence of non-B DNA structures in the human genome (12-16). The presence of such non-canonical DNA structures in the genome have been mapped to regions of chromosomal breakpoints associated with different types of leukemias and lymphomas. Examples include occurrence of G-quadruplex-forming sequences in the *Bcl2* Major breakpoint region *(*Mbr) region, *HOX11*, *c-MYC, c-kit, VEGF, HIF1α, and KRAS* (12, 14, 17, 18). In the present study, we identify the factors responsible for the mutagenesis of the single-stranded region in *BCL11B* loci and delineate the precise mechanism behind the deregulation of *BCL11B* gene in T-cell leukemia.

Objectives

I. To investigate the role of activation-induced cytidine deaminase (AID) behind the fragility of *BCL11B* gene and generation of T cell leukemia.

II. To determine the formation of non-canonical DNA structures in *BCL11B* Exon 4 and binding of AID to non-canonical structures in this region.

Materials and Methods

5'end-labeling of oligomers

5'-end labelling of oligomeric DNA was carried out using T4 polynucleotide kinase in a buffer containing 20 mM Tris-acetate (pH 7.9), 10 mM magnesium acetate, 50 mM potassium acetate, 1 mM dithiothreitol (DTT) and [γ-^{32}P] ATP at 37°C for 1 h. The labelled substrates were purified using a G-25 Sephadex size exclusion column and stored at -20°C until further use.

Cell lines and culture

CEM (acute lymphoblastic leukemia), Nalm6 (pre-B cell leukemia), and Reh (pre-B cell leukemia) cells were from Dr. M. Lieber (USA). K562 (chronic myelogenous leukemia), HeLa (human cervical cancer), Molt4 (T-cell lymphoblastic leukemia), SupT1 (T-cell lymphoblastic lymphoma), Molt3 (T-cell lymphoblastic leukemia), and Jurkat (acute T cell leukemia) cells were purchased from National Center for Cell Science, Pune, India. Jurkat, CEM, SupT1, Molt4, Molt3, Reh, K562, and Nalm6 cell lines were cultured in RPMI 1640 supplemented with 10% or 15% FBS. HeLa was cultured in DMEM supplemented with 10% FBS. The media was supplemented with Penicillin G (100 µg/ml) and streptomycin (100 µg/ml) and incubated at 37ºC in a humidified atmosphere containing 5% CO_2. The cells growing in the log phase were used for transfection.

Plasmids and bacterial strains

Oligomers UR2-39 and UR2-54 were used for amplifying Region I of *BCL11B* exon 4. The amplified product was cloned into a TA vector (pTZ57R/T, Thermo Fisher Scientific, USA) in the physiological orientation to generate pAS2-1 and the anti-physiological orientation to generate pSS4-1. The clones were checked by PvuII digestion, and identity was confirmed by DNA sequencing (Medauxin, India).

The GST-AID expression construct (pGEX-5x-3GST) was a kind gift from Dr. Alberto Martin, Toronto. The plasmid expressing AID protein, pCMV-wtAID-3x FLAG, was a gift from Dr. Tasuku Honjo, Japan. Rosetta *BL21* (DE3) pLysS cells were purchased from Novagen (USA).

Bioinformatic analysis for Non-B DNA structure

A specialized database, non-B DB database ver 2.0 was used to analyze the formation of different types of non-B DNA structures. A Non-B DNA motif search tool (nBMST) was used wherein the non-B DNA motifs of interest were selected, their sequence

added, and the results obtained were used for further analysis. Additionally, various control genes of similar length and GC content were investigated for the presence of non-B DNA motifs in comparison to the *BCL11B* gene.

***In silico* analysis of AID expression in T-ALL patients**

T-ALL patient data in the form of Read counts was downloaded from the GEO dataset GSE110637. Normalized count (RPKM) was calculated for the AID gene using the following formula; RPKM= (Read counts)/ ([gene length / 10^3] *[total number of reads mapped in the sample /10^6]). The patients were arranged in the increasing order of RPKM values and a bar graph was plotted.

Genomic DNA extraction

Jurkat, SupT1, CEM, Reh, Molt4, and Molt3 cells were pelleted down and washed with 1X PBS. Genomic DNA was extracted by lysing the cells in TE (10 mM Tris, pH 8.0 and 1 mM EDTA) and 0.5% SDS, followed by proteinase K digestion (20 µg/ml) at 37^{o}C overnight. The genomic DNA was purified by phenol: chloroform extraction and precipitated with absolute ethanol. DNA was rinsed in 70% ethanol and air dried. The pellet was resuspended in TE and stored at 4^{o}C.

Detection of single-stranded regions by sodium bisulphite modification assay

Approximately, 2 µg of genomic DNA from Molt4 and Molt3 or 1 µg of plasmid DNA was resuspended in 30 µl of distilled water, mixed with 12.5 µl of 20 mM hydroquinone and 457.5 µl of 2.5 M sodium bisulphite (pH 5.2) and incubated for 14 to 16 h at 55^{o}C in the dark. The bisulfite treated DNA was purified using the Wizard DNA cleanup kit (Promega, Madison, WI) according to the manufacturer's instructions. The purified bisulphite-modified DNA was desulphonated with 0.3 M NaOH at 37°C for 15 min; ethanol precipitated with 96 µl of 7.5 M NH_4OAc and 768 µl of chilled ethanol at -20^{o}C for >2 h, washed with 70% ethanol and resuspended in 30 µl TE buffer. The *BCL11B* Region I was PCR amplified from the bisulphite-treated genomic DNA or plasmid DNA, the amplified fragment was resolved in an agarose gel, purified from the gel, TA cloned, and sequenced. The sequencing of the clone was carried out at Medauxin, India, and analyzed.

Primer extension

The replication blocks due to the formation of non-B DNA structure at *BCL11B* exon 4 regions were studied in plasmids pAS2-1 and pSS4-1 by primer extension. Plasmids were heat denatured in 100 mM KCl and allowed to cool slowly. The reaction was carried out by mixing various concentration of these DNA sample in 1X thermo polymerase buffer [10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl (pH 8.8), 4 mM $MgSO_4$ and 0.1% Triton X-100], 4 mM $MgSO_4$, 200 µM deoxynucleotide triphosphates (dNTPs), 0.5 µM end-labelled primers and 1 U Vent (exo-) polymerase. Linear amplification by primer extensions using UR2-39 was carried out (35 cycles) under the following conditions: heat denaturation at 95^{o}C,

annealing at 61^{o}C, extension by Taq polymerase at 72^{o}C, and final extension for 4 min. Similarly, for UR2-56 and UR2-57, linear amplification of primers was carried out (35 cycles) under the following conditions: heat denaturation at 95^{o}C, annealing at 59^{o}C, extension by Taq polymerase at 72^{o}C, and final extension for 4 min. For the bottom strand primer, UR2-55, the annealing temperature used was 56^{o}C. The reactions were terminated by adding a dye containing formamide, products were resolved on a 10% polyacrylamide gel, dried, and signals were detected using phosphorImager FLA9000 (Fuji, Japan).

Western blotting

For immunoblotting analysis, 30 µg protein extracts from Reh, CEM, Jurkat, SupT1, Molt4 and Nalm6 cells were resolved on 10% SDS-PAGE and performed as described before. Following electrophoresis, the protein was transferred to PVDF membrane (Millipore, USA), Ponceau S stained (5% glacial acetic acid and 0.1% Ponceau S, Sigma), blocked using 5% skimmed milk powder, probed with anti-AID antibody (Santa Cruz Biotechnology, USA). Blots were washed in PBST (1X PBS and 0.1% Tween 20), incubated with anti-mouse secondary antibody (Santa Cruz Biotechnology, USA) at room temperature for 2 h followed by incubation with HRP (horseradish peroxidase) at RT for 30 min. The blots were rinsed in PBST, developed using the chemiluminescent solution (Immobilon ^{TM} western, Millipore, USA), and scanned by gel documentation system (LAS 3000, Fuji, Japan). Each western blot was repeated minimum of three times with different batches of extracts.

Analysis of mutation frequency in Region I of *BCL11B* exon 4 after AID overexpression

The plasmid expressing AID, pCMV-wtAID-3x FLAG was transfected in Reh and HeLa cells. Overexpression of AID was performed using polyethylimine (PEI) method of transfection. In brief, ~4 lakh HeLa and Reh cells were transfected with increasing concentrations of the plasmid, mixed with PEI (1:1) and OptiMEM medium (GIBCO, Gaithersburg, MD, USA). The mixture was then added to the cells and incubated at 37^{o}C for 48 h. First, cells were harvested, and protein was extracted using the RIPA extraction method. AID overexpression was confirmed by western blotting using anti-FLAG antibody. Parallelly, another fraction of cells from the same batch was used to isolate genomic DNA followed by PCR amplification of Region I of *BCL11B* gene using specific primers.

The PCR products were subsequently cloned at the EcoRV site of the pBSK(+) vector. The presence of insert was confirmed by restriction digestion analysis and clones of interest were sequenced and analyzed. Additionally, two control regions which did not show any potential non-B DNA forming motif were investigated for AID induced mutation.

Gel mobility shift assay for analysis of G-quadruplex DNA

The radiolabeled oligomers with G4 motifs and their corresponding mutants were incubated either in the presence or absence of 100 mM KCl in TE buffer (pH 8.0) at 37°C for

1 h. The oligomers were then resolved on 15% native polyacrylamide gels in the presence or absence of 100 mM KCl, both in the gel and the buffer, at 150 V at room temperature. The gels were dried and exposed to a screen, and the signal was detected using phosphorImager FLA9000 (Fuji, Japan).

Electromobility shift assay for triplex DNA structures

The radiolabeled oligomer NK1 was incubated with increasing concentrations (15, 30,120, 300, and 600 nM) of the complementary strands, NK2 in the presence of 10 mM $MgCl_2$ and 100 mM NaCl in Tris buffer (pH 8.0) at room temperature for 4 h after heat denaturation at 95°C for (10 min). The products were then resolved on 15% native polyacrylamide gels in the presence of 10 mM $MgCl_2$, both in the gel and the buffer (Tris-borate, pH 7.8), at 150 V at 4°C (14, 19-21). The gels were dried and exposed to a screen, and the signal was detected using phosphorImager FLA9000 (Fuji, Japan).

Gel mobility shift assay for hairpin DNA structures

The radiolabeled oligomers forming hairpin DNA structures and the respective mutants were incubated in the presence of 100 mM NaCl in Tris-EDTA buffer (pH 8.0) at room temperature for 4 h after heat denaturation at 95°C (10 min). The oligomers were then resolved on 15% native and 12% denaturing polyacrylamide gels in 1X Tris-EDTA buffer at room temperature as described. The gels were dried and exposed to a screen, and the signal was detected.

DMS protection assay

The radiolabeled oligomer forming G-quadruplex structure was incubated in TE in the presence of either 100 mM KCl or 100 mM LiCl at RT overnight after heat denaturation at 95°C. Dimethyl sulfate (DMS) was added to the reaction mixture (1/200 dilution) and incubated (15 min) at room temperature. An equal volume of piperidine (10%) was added to each tube, and the reaction mixture was incubated at 90°C for 30 min. The reaction was stopped by adding double the volume of TE and vacuum dried. The pellet was further washed with water thrice and dried using a speed vac concentrator. The reaction products were resolved on a 18% denaturing polyacrylamide gel, which was further dried and visualized as described earlier.

Circular dichroism

The G-quadruplex forming oligomers, its mutants, and complementary C-strand oligomer were incubated in TE either in the presence or absence of 100 mM KCl, at 37°C for 1 h. The CD spectra were recorded at room temperature from 210 to 300 nm, and 10 cycles were accumulated for every sample, using a JASCO J-810 spectropolarimeter at a scan speed of 50 nm/min. A separate spectrum was measured for the buffer alone, and this was subtracted from all the experimental spectra. The ellipticity was calculated using the Spectra Manager software and plotted as a function of wavelength.

CD spectra were also taken for the triplex forming oligomers (NK1, NK2) by incubating them in 10 mM Tris (pH 7.0) in the presence of 140 mM KCl and 5 mM $MgCl_2$ at 22°C for 1 h with a scan range of 210-300 nm. The hairpin forming oligomers were incubated in TE with 100 mM of sodium cacodylate buffer at 37°C for 1 h, and the reading was taken as described earlier.

P1 nuclease assay

The 5'-end labeled hairpin oligomers, NK4, NK5, NK6, and NK7, were heat denatured and slowly annealed in the presence of 10 mM $MgCl_2$ overnight in a water bath. The oligomeric DNA were tested for P1 nuclease sensitivity by incubating with increasing concentrations of P1 nuclease (0.000001, 0.00001, 0.0001 and 0.001 units) in a buffer containing 10 mM Tris-HCl (pH 7.9), 10 mM $MgCl_2$, 50 mM NaCl, and 1 mM DTT at 37°C for 30 min. The reaction was stopped by the addition of 2X formamide dye, and the products were resolved on 18% denaturing PAGE and analyzed.

Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitation was performed as described with slight modifications. Briefly, Jurkat ($4x10^6$) or Molt4 ($8x10^6$) cells were grown for 24 h at 5% CO_2 in a 37°C incubator. Cells were then harvested and crosslinked with 1% formaldehyde. Crosslinking was quenched by the addition of 100 µl of 1.375 M glycine per ml of culture. Cells were pelleted down, washed with 1X PBS, and lysed in cell lysis buffer (5 mM PIPES, 85 mM KCl, 0.5% NP40), and the nuclei were collected by centrifugation. The nuclei were then incubated in nuclei lysis buffer (50 mM Tris [pH 8.0], 10 mM EDTA, 1% SDS). Chromatin was sheared by sonication (Diagenode biorupter) with 75 sec bursts (30 sec pulse and 45 sec break) at low output, 30 cycles. After removal of precipitated proteins, concentration and purity of DNA was measured using Nanodrop 2000, ThermoFisher (Massachusetts, USA). 10% chromatin was saved as input control, and the rest was divided equally into two fractions, one for secondary antibody control, incubated with anti-rabbit secondary antibody, and the rest incubated with anti-AID (2 µg/µl) at 4°C. The samples were incubated further with Protein A/G-agarose beads (Sigma) for 2 h at 4°C with mild rotation. The beads were pelleted down and washed with high salt buffer (50 mM HEPES (pH 7.9), 500 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% Triton X-100, 0.1% deoxycholate) five times and further washed twice with TE. The beads were then incubated in elution buffer (50 mM Tris [pH 8.0], 10 mM EDTA, 1% SDS) and Proteinase K (20 µg/µl) at 55°C for 2 h and cross-linking was reversed by incubating overnight at 65°C. Finally, DNA was purified by phenol:chloroform extraction and ethanol precipitated. PCR amplification for *BCL11B* Region I, and control regions was carried out. When needed, purified PCR fragment was cloned in pTZ57R/T vector and sequenced to confirm the identity.

Overexpression and purification of GST-AID

E. coli BL21 (DE3) cells were transformed with the GST-AID construct (pGEX-5x-3GST) and AID purification was carried out as described earlier (22). Briefly, the bacteria were grown to a log-phase culture, and protein expression was induced by addition of 1 mM IPTG followed by 16 h of incubation at 16°C. The cells were harvested and lysed in lysis buffer (1X PBS, 1% Triton X-100, 1 mM PMSF). The lysate was loaded on the glutathione-sepharose column (GE), and the purified protein was eluted with 10 mM reduced glutathione. The purity of protein was checked on 10% SDS-PAGE, and the identity was confirmed by immunoblotting.

AID deamination assay

Radiolabeled DNA substrate containing the WRC motif (ASM4) was incubated with different purified fractions of AID for 60 min at 37°C in a buffer containing 50 mM Tris (pH 7.5), 100 mM NaCl, and 2 µM $MgCl_2$ in a volume of 10 µl. The protein was then heat inactivated for 15 min at 75°C. 1 U of uracil DNA glycosylase, was added to this in 1X base-excision buffer (20 mM Tris (pH 8), 1 mM DTT and 1 mM EDTA) and RNase Inhibitor were added, followed by incubation at 37°C for 90 min. Finally, 10 mM NaOH was added, and the samples were heated to 95°C for 8 min to cleave the abasic site generated by removal of uracil. The reaction products were resolved on a 12% Denaturing PAGE and visualized using phosphorImager, FLA9000 (Fuji, Japan).

Electrophoretic mobility shift assay

Radiolabeled oligomeric DNA substrates (corresponding to wild-type G strand, UR2-60 and its mutant, UR2-61) were incubated with purified GST-AID for 1 h on ice in 1X deamination buffer (50 mM Tris (pH 7.5), 100 mM NaCl, 2 µM $MgCl_2$). In the control reaction, only buffer was added. The bound complexes were resolved on 4% native PAGE containing 100 mM KCl. The gels were dried and exposed to a screen, and the signal was detected using phosphorImager FLA9000 (Fuji, Japan). For quantification of the bound-complexes, multi gauge software was used. The same assay was also performed by incubating G-strand oligomer with increasing concentration of AID (0, 0.5, 1, 1.5 µg) followed by steps described above.

For the determination of AID binding to triplex DNA, radiolabelled oligomeric DNA substrate, NK1 was allowed to form triple helical structure with its complementary strand, NK2 by following the reaction conditions described above. The reaction products were incubated with increasing concentration of purified GST-AID (0, 1, 1.5 and 2 µg) for 1 h on ice in 1 X deamination buffer (50 mM Tris (pH 7.5), 100 mM NaCl, 2 µM $MgCl_2$). Parallelly, NK1 alone was incubated with increasing concentration of purified GST-AID (0, 1, 1.5 and 2 µg) and reaction was performed as described above. In the control reaction, only buffer was added. The bound complexes were resolved on 4% native PAGE containing 100 mM KCl.

Biolayer interferometry

Forte Bio Octet RED 96 (Forte Bio, Menlo Park, CA) and Super Streptavidin (SSA) sensors (Forte Bio, USA) were used for studying the binding of AID to biotinylated G-strand (DR71) containing oligomer as described previously (23-26). As a control, a biotinylated T strand (MS3) was used. Briefly, all the SSA sensor tips were initially hydrated in the buffer (1X PBS) for 10 min. The programme consisted of a series of steps, which includes the establishment of a stable baseline with 1X PBS (60 sec), loading of sensors with biotinylated oligo (1 µM) for 5 min followed by addition of increasing concentrations of purified AID protein (66 and 100 nM) to 96-microwell plate filled with 200 µl of assay buffer (0.1% BSA, Tween 20 and 1X PBS). The following steps were then carried out: Association (5 min), dissociation (5 min), regeneration with 1 M NaCl (30 sec*3), baseline (5 min). A reference sensor without bound biotinylated DNA was subjected to the same procedure as the sample sensors and used for subtraction of the background signals. The K_D values were calculated using the curve fit (1:1, association: dissociation) model.

Results

In the present study, we have identified the molecular mechanism behind the deregulation of the T-cell transcription factor, BCL11B, which undergoes frequent mutations in T-cell acute lymphoblastic leukemia patients. The mutation rate is significantly higher in exon 4 of *BCL11B* compared to the other three exons. Mutation in the exon 4 of *BCL11B* abrogates its DNA-binding ability and leads to proliferation and differentiation of immature thymocytes. Further characterization of the type of mutations reveal that they are mostly cytosine to thymine (C>T) or guanine to adenine (G>A) conversion and ~85% of missense mutations in T-ALL patients occur at or adjacent to the AID-hotspot motifs (DGWY/WRCH/WRC, D=A/G/T, Y=C/T, R=A/G, H=T/C/A and W=A/T). Moreover, the cytosines which are part of these motifs or adjacent to it undergo deamination at ~11-fold higher rate than cytosines randomly distributed in *BCL11B* exons. This observation prompted us to investigate the role of AID in the fragility of *BCL11B* gene. Here, we report that AID is expressed aberrantly in T-ALL patients and cell lines. Using chromatin immunoprecipitation assay, we show that AID can bind to *BCL11B* fragile Region I in exon 4. The Region I of exon 4 corresponds to the region which harbours the missense mutations reported in T-ALL patients. The binding of AID to *BCL11B* Region I was specific since AID-specific enrichment was not found when random control regions (*DHR, EPR1,* and *PCBD2*) were analyzed.

To investigate the role of AID behind the mutagenesis of the *BCL11B* gene, we overexpressed AID using the vector, pCMV-wtAID-3X FLAG and confirmed it using immunoblotting. An enhanced mutation frequency in Region I at the AID-hotspot motifs observed upon overexpression of AID reinforces its binding to exon 4 and generation of

mutations upon aberrant repair. To rule out that the observed nucleotide conversions were not random, several control regions were analyzed for nucleotide changes. Interestingly, we did not find any nucleotide conversion in the control regions studied under AID-overexpressed condition. We further determined the mutation frequency due to endogenous expression of AID in different T-ALL cell lines and observed the enrichment of AID-induced mutations at the AID preferred substrates in *BCL11B* Region I, further strengthening the role of AID in hypermutating a non-Ig locus, *BCL11B* Region I.

Several studies have shown that the formation of non-B DNA structures is known to cause genomic fragility in leukemia and lymphoma (12, 13, 27). *In silico* analysis using Non-B DB database v2.0 (28) revealed the presence of triplex motifs, G-quadruplex forming motifs, cruciform motifs, short tandem repeats *etc.*, in the *BCL11B* exon 4. Using various biochemical and biophysical assays like electrophoretic mobility shift assay, circular dichroism, biolayer interferometry, primer extension, P1 nuclease assay etc. we determined the formation of non-canonical structures in *BCL11B* fragile Region I. To our surprise, we also found the specific binding of GST-tagged purified AID protein to the G-quadruplex and triplex DNA formed in this region.

The enzyme AID prefers single-stranded regions for deamination of cytosine to uracil. Since we observed the binding of AID to our region of interest and subsequent generation of mutation, we were curious to determine the single-strandedness of Region I, both at the chromosomal and plasmid DNA level, for which we resorted to sodium bisulphite modification assay. We observed the single-strandedness of region corresponding to G4, triplex and hairpin-forming motif, further indicating the propensity of these sequences to fold into non-B DNA structures and subsequent action of AID in the complementary C-rich strand. Taken together, our results reveal that AID binds to *BCL11B* exon 4 due to the formation of non-B DNA, leading to a U:G mismatch, which when repaired erroneously generates deleterious mutations, resulting in loss of functionality of *BCL11B*, and T-ALL.

Statistical analysis

Each experiment was repeated a minimum of three times. Statistical significance was determined by GraphPad Prism ver 5 using the Student's t-test, and the obtained values were considered significant if the p-value was less than 0.05. Statistical significance was considered as ns: not significant, *$p < 0.05$, **$p < 0.005$, ***$p < 0.0001$.

Discussion

In the current study, we have investigated the plausible mechanism behind the generation of mutations in the exon 4 of the *BCL11B* gene during T-ALL. The frequent involvement of *BCL11B* in various translocations, e.g., t(5;14) (q35;32) and mutagenesis events makes it one of the key tumor suppressor gene to be deregulated in T-ALL (8, 10).

Here, we show the oncogenic role of AID in causing deleterious point mutations in *BCL11B* gene using various experimental assays. The single-strandedness of this region due to the formation of various non-B DNA structures makes it an ideal region in the genome for AID binding and deamination. The preferential binding of AID to non-B DNA like G4 over double-stranded DNA explains its strong affinity towards the *BCL11B* gene (29).

Formation of non-B DNA structures at *BCL11B* Region I

Using extensive *in vitro* assays, we show that G-quadruplex formation, both on oligomeric and plasmid DNA, could lead to a significant block in replication. EMSA and CD studies suggested parallel intramolecular G-quadruplex formation, with a minimum of three stretches of guanine required for its stabilization. On the other hand, the folding of mirror and inverted-repeat sequences into triplex and hairpin structures respectively in the template strand of exon 4 was determined using various experimental assays.

Non-B DNA sequences are widely interspersed in the genome of bacteria, yeast, and human. The formation of non-canonical structures in the genome, can affect processes like replication, transcription, recombination, etc., which turns on mutagenic events causing large-scale deletions, insertions, chromosomal rearrangements etc. (30, 31). Formation of cruciform, triplex, and G4 structures *in vivo* has been reported to associate with several translocation breakpoint regions, e.g., t(14;18), t(11;22), t(10;14) etc. observed in different types of leukemia and lymphoma (12, 13, 17, 32). Thus, our findings along with several other published literatures implicate the generation of fragile regions in the genome due to frequent occurrence of non-B DNA structures.

AID induced-mutation at *BCL11B* exon 4 in T-ALL

Analysis of the mutational signature in *BCL11B* exon 4 indicated 3.1-fold higher incidence of G>A and 5-fold of C>T over any random mutation, specifically enriched at the AID-hotspot motifs. The expression of AID in activated B cells is well established due to its role in somatic hypermutation and class-switch recombination of the Ig genes to generate antibody diversity (33-35). Overexpression of AID results in various cancers including T and B cell lymphomas/ leukemias (36, 37). In our study, we report for the first time that the aberrant expression of AID in T-ALL cells can induce signature mutations in *BCL11B* Region I, which is augmented on overexpression of AID. Notably, our study shows that the mutations generated either due to endogenous or overexpression of AID in cell lines or chemical probing are mostly enriched at the AID-targeting substrates. Induction of mutation specifically in *BCL11B* Region I and not the control regions reiterate the mutagenic property of AID in the *BCL11B* gene. However, we cannot completely rule out the fact that at least some of these mutations may also arise due to aberrant replication or DNA repair of non-B DNA regions in *BCL11B*.

Plausible mechanism of fragility of *BCL11B* gene

Based on our findings, we propose a molecular mechanism of the AID-induced mutations in the *BCL11B* gene (Fig. 1). The transient opening of DNA strands during processes like transcription, replication etc. paves the way for the formation of different non-B DNA structures like G4, cruciform, triplex etc. inside the cell (12, 18, 32, 38). The cytosines on the complementary single-stranded regions are acted upon by the cytidine deaminase, AID, creating a U:G mismatch (Fig. 1). Replication of these strands leads to C>T or G>A alterations. Activation of base-excision repair pathway excises out the uracil by an Uracil-DNA glycosylase, forming abasic sites. AP endonuclease then creates a single-strand break to ensure the incorporation of the correct nucleotide. However, in the event of repair via the error-prone pathway extensive nucleotide mismatches are generated, leading to different type of nucleotide alterations or single and double-strand breaks (39). Interestingly, a recent report by Sawai *et al.* showed that C: G>T: A transitions are frequently observed in tumor-related genes like *KRAS*, *c-MYC*, and *SMAD4* due to ectopic expression of AID in the pancreatic tumor (40). A close association was found between aberrant expression of AID and increased rate of *TP53* mutations in lung cancer (41). Furthermore, AID transgenic mice spontaneously develop T cell lymphomas and micro-adenomas in the lung and harbour frequent point mutations in the TCR and the *c-MYC* genes (42, 43).

Erroneous repair system can sometimes result in double-strand breaks, which act as a translocation hotspot e.g., breaks at the *c-MYC* gene during t(8;14) translocation triggers the development of Burkitt's lymphoma. The propensity of the *c-MYC* fragile region to fold into G-loops facilitates the binding of AID to this region (44, 45). Thus, deregulated or constitutive AID expression might be a driving factor responsible in part for haematological and non-haematological malignancies.

Overall, our study delineates the mechanisms by which AID induces mutation in *BCL11B* exon 4 and explains how that makes the region fragile and prone to frequent mutations, deletions, and translocation events. A major future challenge will be to determine whether the presence of AID alone is sufficient to support mutagenic events inside the cell or additional factors are needed to facilitate its recruitment and subsequent oncogenic activity.

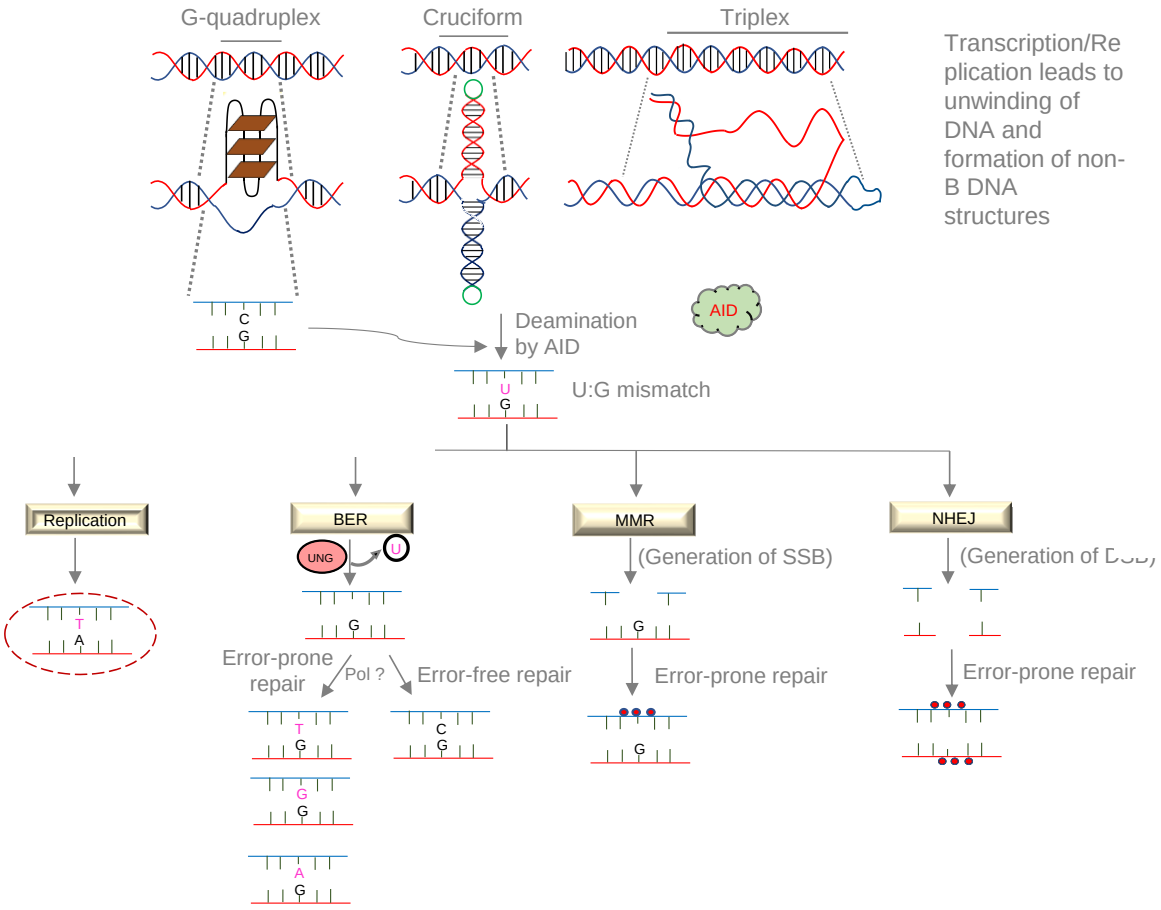

Figure 1: Plausible mechanism of fragility of *BCL11B* gene induced by oncogenic action of AID.

Impact of the research in the advancement of knowledge or benefit to mankind

1) Our study delineates the role of AID as the primary mediator of oncogenesis in *BCL11B* which could possibly lead to T-cell leukemia. It opens up new avenues to study the alteration in the genetic and epigenetic landscape of T-cell leukemia due to aberrant expression of AID.
2) Our study highlights the non-canonical role of AID in inducing hypermutation in non-Ig loci.
3) The observation made during this study could pave the way for the therapeutic targeting of AID in clinics to improve the overall survival and prognosis of cancer patients.

References

1. Roy U, Raghavan SC. Deleterious point mutations in T-cell acute lymphoblastic leukemia: Mechanistic insights into leukemogenesis. International journal of cancer. 2021;149(6):1210-20.
2. Neumann M, Vosberg S, Schlee C, Heesch S, Schwartz S, Gökbuget N, et al. Mutational spectrum of adult T-ALL. Oncotarget. 2015;6(5):2754-66.
3. Rothenberg EV, Taghon T. Molecular genetics of T cell development. Annual review of immunology. 2005;23:601-49.
4. Liu P, Li P, Burke S. Critical roles of Bcl11b in T-cell development and maintenance of T-cell identity. Immunol Rev. 2010;238(1):138-49.

5. Kominami R. Role of the transcription factor Bcl11b in development and lymphomagenesis. Proceedings of the Japan Academy Series B, Physical and biological sciences. 2012;88(3):72-87.
6. Vanvalkenburgh J, Albu DI, Bapanpally C, Casanova S, Califano D, Jones DM, et al. Critical role of Bcl11b in suppressor function of T regulatory cells and prevention of inflammatory bowel disease. The Journal of experimental medicine. 2011;208(10):2069-81.
7. Li P, Burke S, Wang J, Chen X, Ortiz M, Lee SC, et al. Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. Science (New York, NY). 2010;329(5987):85-9.
8. Gutierrez A, Kentsis A, Sanda T, Holmfeldt L, Chen SC, Zhang J, et al. The BCL11B tumor suppressor is mutated across the major molecular subtypes of T-cell acute lymphoblastic leukemia. Blood. 2011;118(15):4169-73.
9. Bartram I, Gökbuget N, Schlee C, Heesch S, Fransecky L, Schwartz S, et al. Low expression of T-cell transcription factor BCL11b predicts inferior survival in adult standard risk T-cell acute lymphoblastic leukemia patients. Journal of hematology & oncology. 2014;7:51.
10. Huang X, Du X, Li Y. The role of BCL11B in hematological malignancy. Experimental hematology & oncology. 2012;1(1):22.
11. MacLeod RA, Nagel S, Drexler HG. BCL11B rearrangements probably target T-cell neoplasia rather than acute myelocytic leukemia. Cancer genetics and cytogenetics. 2004;153(1):88-9.
12. Nambiar M, Goldsmith G, Moorthy BT, Lieber MR, Joshi MV, Choudhary B, et al. Formation of a G-quadruplex at the BCL2 major breakpoint region of the t(14;18) translocation in follicular lymphoma. Nucleic acids research. 2011;39(3):936-48.
13. Nambiar M, Srivastava M, Gopalakrishnan V, Sankaran SK, Raghavan SC. G-quadruplex structures formed at the HOX11 breakpoint region contribute to its fragility during t(10;14) translocation in T-cell leukemia. Molecular and cellular biology. 2013;33(21):4266-81.
14. Raghavan SC, Swanson PC, Wu X, Hsieh CL, Lieber MR. A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. Nature. 2004;428(6978):88-93.
15. Raghavan SC, Lieber MR. Chromosomal translocations and non-B DNA structures in the human genome. Cell cycle (Georgetown, Tex). 2004;3(6):762-8.
16. Mirkin SM. Discovery of alternative DNA structures: a heroic decade (1979-1989). Frontiers in bioscience : a journal and virtual library. 2008;13:1064-71.
17. Katapadi VK, Nambiar M, Raghavan SC. Potential G-quadruplex formation at breakpoint regions of chromosomal translocations in cancer may explain their fragility. Genomics. 2012;100(2):72-80.
18. Nambiar M, Raghavan SC. Mechanism of fragility at BCL2 gene minor breakpoint cluster region during t(14;18) chromosomal translocation. The Journal of biological chemistry. 2012;287(12):8688-701.
19. Raghavan SC, Swanson PC, Ma Y, Lieber MR. Double-strand break formation by the RAG complex at the bcl-2 major breakpoint region and at other non-B DNA structures in vitro. Molecular and cellular biology. 2005;25(14):5904-19.
20. Kumari N, Vartak SV, Dahal S, Kumari S, Desai SS, Gopalakrishnan V, et al. G-quadruplex Structures Contribute to Differential Radiosensitivity of the Human Genome. iScience. 2019;21:288-307.
21. Raghavan SC, Kirsch IR, Lieber MR. Analysis of the V(D)J recombination efficiency at lymphoid chromosomal translocation breakpoints. The Journal of biological chemistry. 2001;276(31):29126-33.
22. Larijani M, Petrov AP, Kolenchenko O, Berru M, Krylov SN, Martin A. AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner. Molecular and cellular biology. 2007;27(1):20-30.
23. Sultana A, Lee JE. Measuring protein-protein and protein-nucleic Acid interactions by biolayer interferometry. Current protocols in protein science. 2015;79:19.25.1-19.25.6.
24. Pandey M, Kumar S, Goldsmith G, Srivastava M, Elango S, Shameem M, et al. Identification and characterization of novel ligase I inhibitors. Molecular carcinogenesis. 2017;56(2):550-66.
25. Nilavar NM, Paranjape AM, Raghavan SC. Biochemical activity of RAGs is impeded by Dolutegravir, an HIV integrase inhibitor. Cell death discovery. 2020;6:50.
26. Vartak SV, Iyer D, Santhoshkumar TR, Sharma S, Mishra A, Goldsmith G, et al. Novel BCL2 inhibitor, Disarib induces apoptosis by disruption of BCL2-BAK interaction. Biochem Pharmacol. 2017;131:16-28.
27. Tsai AG, Lu H, Raghavan SC, Muschen M, Hsieh CL, Lieber MR. Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. Cell. 2008;135(6):1130-42.

28. Cer RZ, Donohue DE, Mudunuri US, Temiz NA, Loss MA, Starner NJ, et al. Non-B DB v2.0: a database of predicted non-B DNA-forming motifs and its associated tools. Nucleic acids research. 2013;41(Database issue):D94-D100.
29. Qiao Q, Wang L, Meng FL, Hwang JK, Alt FW, Wu H. AID Recognizes Structured DNA for Class Switch Recombination. Molecular cell. 2017;67(3):361-73.e4.
30. Bacolla A, Wells RD. Non-B DNA conformations, genomic rearrangements, and human disease. The Journal of biological chemistry. 2004;279(46):47411-4.
31. Maizels N. Dynamic roles for G4 DNA in the biology of eukaryotic cells. Nature structural & molecular biology. 2006;13(12):1055-9.
32. Nambiar M, Kari V, Raghavan SC. Chromosomal translocations in cancer. Biochim Biophys Acta. 2008;1786(2):139-52.
33. Honjo T, Kinoshita K, Muramatsu M. Molecular mechanism of class switch recombination: linkage with somatic hypermutation. Annual review of immunology. 2002;20:165-96.
34. Cantaert T, Schickel JN, Bannock JM, Ng YS, Massad C, Oe T, et al. Activation-Induced Cytidine Deaminase Expression in Human B Cell Precursors Is Essential for Central B Cell Tolerance. Immunity. 2015;43(5):884-95.
35. Kuraoka M, Holl TM, Liao D, Womble M, Cain DW, Reynolds AE, et al. Activation-induced cytidine deaminase mediates central tolerance in B cells. Proceedings of the National Academy of Sciences. 2011;108(28):11560-5.
36. Komeno Y, Kitaura J, Watanabe-Okochi N, Kato N, Oki T, Nakahara F, et al. AID-induced T-lymphoma or B-leukemia/lymphoma in a mouse BMT model. Leukemia. 2010;24(5):1018-24.
37. Rebhandl S, Geisberger R. AIDing cancer treatment: Reducing AID activity via HSP90 inhibition. European journal of immunology. 2015;45(8):2208-11.
38. Roy U, Raghavan SC. Deleterious Point Mutations in T-cell Acute Lymphoblastic Leukemia: Mechanistic Insights into Leukemogenesis. 2021.
39. Wallace SS. Base excision repair: a critical player in many games. DNA Repair (Amst). 2014;19:14-26.
40. Sawai Y, Kodama Y, Shimizu T, Ota Y, Maruno T, Eso Y, et al. Activation-Induced Cytidine Deaminase Contributes to Pancreatic Tumorigenesis by Inducing Tumor-Related Gene Mutations. Cancer Research. 2015;75(16):3292-301.
41. Shinmura K, Igarashi H, Goto M, Tao H, Yamada H, Matsuura S, et al. Aberrant Expression and Mutation-Inducing Activity of AID in Human Lung Cancer. Annals of Surgical Oncology. 2011;18(7):2084-92.
42. Okazaki IM, Hiai H, Kakazu N, Yamada S, Muramatsu M, Kinoshita K, et al. Constitutive expression of AID leads to tumorigenesis. The Journal of experimental medicine. 2003;197(9):1173-81.
43. Pettersen HS, Galashevskaya A, Doseth B, Sousa MML, Sarno A, Visnes T, et al. AID expression in B-cell lymphomas causes accumulation of genomic uracil and a distinct AID mutational signature. DNA Repair. 2015;25:60-71.
44. Duquette ML, Pham P, Goodman MF, Maizels N. AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation. Oncogene. 2005;24(38):5791-8.
45. Nambiar M, Raghavan SC. How does DNA break during chromosomal translocations? Nucleic acids research. 2011;39(14):5813-25.

Urbi Roy
Graduate Student
C/o Prof. Sathees C. Raghavan
Department of Biochemistry
Indian Institute of Science
Bangalore-56001