Research Article

Concerted dynamics between conserved histidines in the N-extein region regulates *Mycobacterium tuberculosis* SufB intein cleavage

Running Title: Conserved histidines regulate *Mtu* SufB cleavage

Sunita Panda[1†], Ananya Nanda[1†], Nilanjan Sahu[2], Deepak Ojha[1], Biswaranjan Pradhan[3], Anjali Rai[1], Amol R. Suryawanshi[4], Nilesh Banavali[5,6] and Sasmita Nayak[1*]

[1]School of Biotechnology, Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha, India 751024,

[2]School of Biological Sciences, National Institute of Science Education and Research Bhubaneswar, Jatni, Khurda, Odisha, India 752050

[3] S K Dash center of excellence of Biosciences and Engineering & Technology, School of Basic Sciences, Indian Institute of Technology Bhubaneswar, Argul, Khurda, Odisha, India 752050

[4] Clinical Proteomics, Institute of Life Sciences, Bhubaneswar, Odisha 751023 India

[5] Wadsworth Center, 10 New York State Department of Health, Empire State Plaza, P. O. Box 509, Albany, NY 12201

[6] Department of Biomedical Sciences, School of Public Health, University at Albany, 1400 Washington Avenue, Albany, NY 12222

[†] These authors contributed equally: Sunita Panda, Ananya Nanda.

[*]Correspondence: Dr Sasmita Nayak, School of Biotechnology, Kalinga Institute of Industrial Technology, Campus XI, Patia, Bhubaneswar, Odisha, India 751024

Email:-sasmita.n@kiitbiotech.ac.in; Tel: +918763000005

Abbreviations:

Mtu: *Mycobacterium tuberculosis*

WT: Wild type

SI: Splicing inactive

DM: *Mtu* SufB Double mutant

IPTG: Isopropyl β- d-1-thiogalactopyranoside

TCEP: Tris (2-carboxyethyl) phosphine

HA: Hydroxylamine

DTT: Dithiothreitol

HRP: Horseradish peroxidase

ECL: Enhanced chemiluminescence

TBST: Tris-buffered saline, 0.1% tween 20

BSA: Bovine serum albumin

ß-gal: Beta galactosidase

ABC: Ammonium bicarbonate

ACN: Acetonitrile

TFA: Tri-fluoro acetic acid

HPLC: High pressure (or performance) liquid chromatography

FPLC: First protein liquid chromatography

MALDI-TOF/TOF: Matrix assisted laser desorption ionization time-of-flight

SDS: Sodium dodecyl sulphate

PAGE: Polyacrylamide gel electrophoresis

RMSD: Root mean square deviation

SASA: Solvent accessible surface area

RMSF: Root mean square fluctuation

PME: Particle-mesh Ewald

HPCC: High performance computing cluster

MD Simulations: Molecular dynamic simulations

Abstract: Inteins are auto-processing domains that implement a multi-step biochemical reaction termed protein splicing, marked by cleavage and formation of peptide bonds. They excise from a precursor protein, generating a functional protein via covalent bonding of flanking exteins. We report the kinetic study of splicing and cleavage reaction in a [Fe-S] cluster assembly protein SufB from *Mycobacterium tuberculosis*. Although it follows a canonical intein splicing pathway, distinct features are added by extein residues present in the active site. Sequence analysis identified two conserved histidines in the N-extein region; His-5 and His-38. Kinetic analyses of His-5Ala and His-38Ala SufB mutants exhibited significant reductions in splicing and cleavage rates relative to the SufB wild-type precursor protein. Structural analysis and molecular dynamics simulations suggested that *Mtu* SufB displays a unique mechanism where two remote histidines work concurrently to facilitate N- cleavage reaction. His-5, which is exposed outside, because of the random push of water molecules forces His-38 towards the N-cleavage site. Thus, His-5 stabilizes the position of His-38 which in turn activates N-S acyl shift via direct interaction with catalytic Cys1. Understanding intein~extein partnership in an essential mycobacterial protein may diversify into intein-based applied research along with the development of pathogen-specific novel antimicrobials.

Keywords: Splicing regulation; conserved histidines; *Mycobacterium tuberculosis* SufB

1. Introduction

Protein splicing is an auto-catalytic event that generates an active protein by ligating the flanking exteins with a covalent bond. Post-translational auto-excision of intein remains critical for the formation of an active protein (1-4). Splicing and cleavage products (Fig. 1C) are obtained through a series of nucleophilic displacement reactions mediated by the coordinated activity of catalytic residues (5-9). Therefore, inteins interrupting the functional domain of the

host protein, play a regulatory role in protein activation(10, 11). An intein fragment essentially constitutes an N-terminal and a C-terminal end. The sequence upstream to the N-terminal intein end is termed as N-extein and C-extein is downstream to the intein C-terminal end. Intein residues are numbered sequentially from 1 to n. N-extein residues are numbered as -1 to -n from the N-extein-intein junction such that the last residue of the N-extein is numbered -1. C-extein is numbered as +1 to +n from the C-intein-extein junction.

The intein folds along with the precursor protein into a horseshoe-like structure shaping the catalytic cleft. The formation of a catalytic cleft or active site is a crucial conformational change that brings the conserved catalytic residues and the splice junctions (The N- and C-terminal extein~intein junctions) within catalytic distance for initiation of splicing(12). The intein sequence houses these catalytic residues within conserved segments termed as Blocks or motifs that facilitate the splicing and cleavage reaction(s). N-terminal intein comprises of A, N2, B, and N4 structural motifs or Blocks while F- and G-Blocks are part of the C-terminal intein. Typically A-Block contains Cys/Ser or Thr; B-Block includes His and Thr residues; F-Block usually has Asp and His and G-Block bears two conserved residues; a penultimate His and a terminal Asn. Classical or Canonical (Class 1) intein splicing involves 4 sequential acyl rearrangements (Fig. 1C), a first nucleophilic attack by C1 or S1/T1 leading to N-S/N-O acyl shift converting the peptide bond of N-terminal splice junction to a thioester linkage, a second nucleophilic attack by C+1 forms a branched intermediate at C-terminal splice junction through transesterification. The branched intermediate is resolved by terminal Asn cyclization through cleavage of C-terminal splice junction, and an (S-N/O-N) acyl shift to ligate the two extein segments by an amide bond formation. Thus the conserved catalytic residues present within the motifs/Blocks help in rearranging the bonds at splice junctions to carry out the cleavage and

splicing reaction(s), however, non-catalytic intein residues assist in the activation and stabilization of various intermediate structure(s)(13). Block A residues Cys, Ser, or Thr are vital for carrying out the first step of splicing with significant assistance from Block B His and Thr residues. Especially, the highly conserved Block B His destabilizes the scissile peptide bond either by reducing the energy barrier or by loss of resonance or protonation of the Cys1 amide bond via His imidazole ring to catalyze the N-S acyl shift (14-18). Block F Asp plays a pivotal role in driving the thioesterification and stabilizing the tetrahedral intermediate by ground-state destabilization (8, 16, 19). It is also proposed that Block B histidine plays a dual catalytic role; being weakly basic it deprotonates the Cys1 to accelerate the N-S acyl shift and subsequently acts as an acid to stabilize the tetrahedral intermediate (15, 17, 18). Block F Asp also deprotonates C+1 residue that stabilizes the net positive charge on Cys1 that drives the transesterification reaction. F- and G-Block His are critical in the coordination of terminal Asn cyclization, while F-Block His increases nucleophilicity of Asn, G-Block His accelerates the Asn cyclization by increasing the electrophilicity of backbone peptide (20-23). The final step is energetically favorable and does not require any assistance from either intein or extein residues(24). These residues are not always conserved since inteins show polymorphisms in the catalytic residues leading to variation in the splicing mechanism such as Class 2 and Class 3 intein splicing (11, 25-29). Studies have emphasized and engaged on the presence of putative interaction between extein residues and the catalytic intein core in the regulation of splicing reaction- by modulating intein activity, changing the conformation of catalytic cleft, and restraining the activity of catalytic residues (6, 30-34).

Previous literature has embarked on the effect of intein~extein partnership on intein splicing mediated by extein residues remote or near N- and C-terminal splice junctions (7, 30, 31, 33,

35-37). N-extein residue at the first position (-1) is important for the first thioester reaction and shows enhanced N-terminal cleavage rate (>4-fold) or attenuated cleavage by 1,000-fold by replacing the native residue to aspartate and proline respectively (30). Substitution of bulky amino acids at this position can cause local bond distortion and induce N-cleavage reaction (17). The participation of the first C-extein Cys+1 is well documented in the second and third steps of the splicing reaction. Mutation of Cys+1 has been shown to dramatically augment or inhibit splicing and generate off-pathway N-cleavage products (12, 35, 38-40). In *Pho* RadA intein, distal exteins seem to act as an environmental sensor with a role in regulating splicing depending on solution environment and temperature (6, 10, 34). It has been well established that exteins play a regulatory role in intein splicing; however, there is very little precedence for distal extein residues to promote such effect. Instead, this extein effect was assumed to be limited to residues proximal to the intein (12, 31, 41).

We have characterized the effect of extein~intein partnership on N-cleavage and splicing reactions in an essential protein SufB from *Mycobacterium tuberculosis* (*Mtu*). SufB is a critical component of [Fe-S] cluster assembly and repair machinery called SUF (mobilization of sulfur) complex. This is a stress response system that gets upregulated during periods of oxidative stress and Fe starvation (42, 43). Though there are multiple pathways for [Fe-S] cluster biogenesis among the three kingdoms of life such as NIF (nitrogen fixation) and ISC (Iron-sulfur cluster), the SUF complex is unique to mycobacteria. Fe-S cluster containing proteins execute a broad spectrum of cellular functions in organisms such as respiration, gene regulation, RNA modification, DNA repair, and replication (44-46). Previous research has specified the prominence of SufB intein splicing in the formation of functional SUF complex in

mycobacteria (47). Although the SUF system has been well characterized in the *E. coli* system, so far there is limited research on the *Mtu* SUF system (39, 42, 43, 47-49).

Current work reports a systematic investigation on splicing and cleavage reactions in an un-spliced precursor protein, *Mtu* full-length SufB (FL-SufB). However, we need to address few fundamental questions here: a) What are the different structural domains of *Mtu* SufB and whether it follows a canonical or non-canonical intein splicing pathway, b) Which residues from SufB intein and extein sequences take part in the active site or catalytic cleft formation, c) Are these residues conserved in different mycobacterial species, d) If yes, then how do they regulate cleavage and/or splicing reactions, and e) how is the above reaction mechanism(s) similar to and/or different from other intein carrying proteins?

We have performed computational analysis to delineate the structural domains of *Mtu* FL-SufB and found that intein critical residues are highly conserved in different mycobacterial species favoring a classic or canonical splicing mechanism (Fig. 1A and 1B). Additionally, we detected two conserved histidines in the N-extein region, His-5 and His-38 in mycobacteria, archaea, and other microbes where SUF is the exclusive system for Fe-S cluster biogenesis (Fig. 2A and 2B). Kinetic analyses of H-5A SufB mutant demonstrated 3-fold and 1.4-fold reductions in splicing and cleavage rates respectively, relative to wild-type *Mtu* SufB precursor. Likewise, a 3.4-fold and 3-fold reduction in splicing and cleavage rates were observed in the H-38A SufB mutant. A side-by-side *Mtu* SufB structure prediction was done using homology (chimera) modeling, secondary structure prediction through consensus with other protein sequences. Subsequently, molecular dynamics (MD) simulations in aqueous media were carried out in order to find the equilibrated structure. Furthermore, MD simulations clarified the structural features of the SufB intein active site and indicated that N-terminal cleavage reaction is catalyzed by H-38 with the

assistance of H-5. Taken together, our study substantiates a distinct mechanism for N-terminal cleavage reaction shown by *Mtu* SufB. Although H-38 is relatively distal to the N-terminal splice junction when supported by His-5, it can efficiently activate the first step of splicing reaction. Finally, we have proposed a novel mechanism for the N-cleavage reaction mediated via the concerted actions of these conserved histidines in the N-extein region of *Mtu* FL-SufB precursor. Interestingly, attempts to express the SufB double mutant (H-5A/H-38A) protein resulted in a truncated protein. These observations suggest that H-5 and H-38 might have important biological role(s) in the SufB precursor stabilization and perhaps the functionality of *Mtu* SufB protein.

2. MATERIALS AND METHODS

2.1. Synthesis of plasmid constructs

Mtu full length (FL) *sufb* gene and *sufb* intein gene were cloned into low copy expression vector pACYCDuetTM-1 (Novagen) driven by a T7 promoter/lac operator that also carries a chloramphenicol resistance gene for selection. The above genes were cloned into multiple cloning site 1 (MCS1) via EcoRI and HindIII restriction sites. The genes were PCR-amplified from heat-killed *Mtu* genomic DNA. Subsequent gel purification and cohesive end ligation (T4 DNA Ligase, NEB Cat. No. M0202S) into the vector gave rise to specific plasmids. The constructs were screened via colony PCR followed by sequencing (Sequencing Core Facility, SUNY, Albany and Eaton Bioscience Inc. sequencing service) using the ACYCDuetUP1 (Novagen Cat. No.71178-3) and DuetDOWN1 Primers (Novagen Cat. No. 71179-3) as well as the original primers.

Mtu FL-SufB mutants; (H-5A), (H-38A), (C+1A), (N359A) were generated by substituting respective key catalytic residues to alanine via phosphorylated inverse PCR primers. Splicing

inactive (SI) *Mtu* FL SufB double mutant (C1A/N359A) was created via inverse PCR run to add N359A mutation into the C1A cleavage mutant. Similarly, (H-5A/H-38A) *Mtu* SufB double mutant (DM) was created to add H-38A mutation into the H-5A mutant by inverse PCR. The genes were PCR amplified using *Pfu* Ultra High-Fidelity DNA Polymerase (Agilent technologies); subsequent gel purification and blunt-end ligation (T4 DNA Ligase, NEB Cat. No. M0202S) gave rise to specific plasmids. The cloned genes [FL-SufB, H-5A, C1A and (C1A/N359A)] were confirmed via sequencing (Sequencing Core Facility, SUNY, Albany and Eaton Bioscience Inc. sequencing service) using the ACYCDuetUP1 (Novagen Cat. No.71178-3) and DuetDOWN1 Primers (Novagen Cat. No. 71179-3) as well as the original primers. H-38A and (H-5A/H-38A) *Mtu* SufB mutant genes were confirmed via sequencing (Agrogenomics, Odisha) using ACYCDuetUP1 (IDT Cat. No. 103948189) and DuetDOWN1 (IDT Cat. No. 103948190) along with the SufB Primers. All the above primers are listed in Table S1.

2.2. Sequence mapping and computational analyses of *Mtu* full length (FL)-SufB

Protein sequences for Fe-S cluster assembly protein SufB [*Mycobacterium tuberculosis* H37Rv] with Accession number YP_006514844.1, GI: 397673309 and the intein less SufB protein [*Mycobacterium smegmatis* strain MC2 155] a 477 amino acid protein (Accession number YP_887437.1GI: 118472504)were obtained from NCBI protein database. Extein and intein sequences were differentiated by a pair-wise sequence alignment in Clustal W interface (50). The identity of *Mtu* SufB intein sequence was confirmed by doing a blast search against the sequences deposited in Inbase (The Intein Database, www.inteins.com) (51). Different structural domains of intein like homing endonuclease, N- and C-terminal inteins were clearly

demarcated by sequence alignment and structural analysis of SufB intein with homing endonuclease domain (I-CreI) and intein homing endonuclease Ii (PDB: 2CW7).

The FL-SufB sequences for the intein-bearing and intein-less mycobacterial, one archaeal, and other bacterial SufB proteins were collected from NCBI database and intein database(51). Alignment of the sequences was performed by Dialign2 software(52). Then conservation of different intein and extein residues were edited manually and color-coded. Then the phylogenetic tree analysis was performed using the Maximum likelihood method in the MEGA-X program (53, 54). Boot strap values were collected and significance was noted. Phylogenetic tree analysis was performed for both SufB inteins [Fig. 2C (ii)] and SufB precursor sequences [Fig. 2C (i)] from different organisms.

2.3. Protein overexpression and purification

Full-length (FL) un-spliced precursor and mutant SufB proteins carrying an N-terminal 6X (His) tag were over-expressed in BL21 (DE3) *E.coli* cells via IPTG (sigma 367-93-1) induction. Cells were resuspended in lysis buffer (20mM sodium phosphate, 0.5M NaCl, pH 7.4) and lysed via tip sonicator (Sonics vibra cell VCX-130). Proteins were over-expressed and isolated from inclusion bodies (IB) via centrifugation. The IB materials were solubilized by 8M urea (Merck, 1084870500) buffer (lysis buffer, 8M urea, 20 mM of imidazole (MP–biochemicals-288-32-4) and centrifuged at 16,500g for 20 min to collect the supernatant. Then 6x(His) tagged Full length (FL) precursor and mutant proteins were purified by Ni-NTA affinity column (Ni-NTA His trap, HP GE healthcare life sciences- 17524802). Prior to the sample application, columns were equilibrated with binding buffer (20mM sodium phosphate, 0.5M NaCl, 40mM imidazole) and after sample loading, columns were washed several times

(15 CV) in binding buffer. Finally, test proteins were eluted as purified fractions in elution buffer (20mM sodium phosphate, 0.5M NaCl, 500mM imidazole) followed by quantification via Bradford's assay.

2.4. *In-vitro* splicing and cleavage assays

2.5µM of purified test proteins were allowed to refold in 1 ml of renaturation buffer (20 mM sodium phosphate, 0.5M NaCl, 0.5 M Arginine,1mM EDTA. pH 7.4) in presence of 2mM TCEP-HCl (sigma- 51805-45-9) at 20˚C for 24 h. 0h sample was retrieved by taking it out before renaturation and splicing was quenched abruptly by addition of loading dye (0.1%bromophenol blue, 50%glycerol, ß-mercaptoethanol, 10%SDS, tris 6.8) followed by rapid freezing at -20^{0}C. Our controls; splicing inactive(SI) SufB double mutant and empty expression vector pACYC Duet-1 were treated similarly for the respective *in-vitro* assays.

For N-cleavage assay, test proteins were refolded in presence of reducing agents and nucleophiles such as 2mM TCEP-HCl, 50 mM DTT (Roche-10708984001), and 0.5 M Hydroxylamine (SRL-66164) with 1mM TCEP in renaturation buffer (55). For splicing and cleavage analysis, samples were taken out at different periods followed by the addition of loading dye to stop the reactions and then boiling at 95˚C for 5 min. Resultant products from various refolding reactions were resolved through 4~10% gradient SDS PAGE. Protein bands were stained with Coomassie blue R-250 and densitometric analysis was performed by using GelQuant.Net biochemical solutions. Percentage(s) of splicing and cleavage products were measured by taking the percentage(s) of the ratio of the total splicing product (LE and I) over total proteins (LE+I+P) and total N-cleavage product (NE+NC) over total proteins (NE+NC+P). 0h splicing value(s) were subtracted at each time point for baseline correction.

Kinetic analyses were performed by fitting the curve in the pseudo-first-order reaction kinetics using graph pad prism as explained in the next section.

2.5. Kinetic analyses

Mtu 6X (His)-tagged WT SufB and SufB mutants (H-5A, H-38A, H-5A/H-38A, and C1A/N359A) were purified and renatured (as shown previously) at different temperatures. After normalizing splicing and cleavage values at different time intervals, the plot was generated by taking the percentage of splicing or cleavage product with respect to time (in min). Next, the curve was fitted in pseudo-first-order kinetics, with an equation Y=Y0 + (Plateau-Y0) $*(1\text{-}\exp^{(-K*x)})$ [Where X=time, Y0 =Y value when time (X)= time0, Plateau= max Y value at time t, K = rate constant, expressed in reciprocal of the X-axis (time units)] in graph pad prism software. The fitted curve was generated by automatic outlier elimination fitting in a nonlinear regression equation. The rate constant (K) and V_{max} were generated by the software. Half-life ($t_{1/2}$) was calculated by graph pad prism using the formula (Ln2/K).

2.6. Western blot

Western blot analysis was performed using an anti-His antibody (Invitrogen, LOT 1902132) to confirm the identity of splicing and cleavage products. Following resolution through SDS PAGE, test proteins were transferred to a nitrocellulose membrane; at 50v, 2h. After a successful transfer, blocking was done with 5% skim milk for 2h at room temperature. Then the blot was incubated with HRP conjugated anti-His antibody (Invitrogen, LOT 1902132) at 1:5000 dilutions for 16h at 4^0C. Then blot was washed with 1X TBST and developed using ECL as the substrate. N-Extein detection was done with 1:2500 antibody dilution.

2.7 MALDI-TOF/TOF MS, HPLC, and FPLC

After renaturation, proteins were resolved through 4~10% SDS PAGE. Then gel was stained and de-stained properly for visualization of splicing products. Resolved splicing and cleavage protein bands were excised from the gel and used for protein identification by mass spectrometry at Central Proteomics Facility, Institute of Life Sciences, Bhubaneswar using a standardized protocol as follows. Gel bands were destained with 25mM ammonium bicarbonate (ABC). Cleared gel bands were reduced with 20mM DTT and then alkylated by 100mM Iodoacetamide solution. Then gel plugs were dehydrated sequentially with 50%, 100% acetonitrile (ACN) and then digested with 0.3µg of trypsin (Absciex 4352157) in 25mM ABC solution at 37˚C for 16 hr. The next day, peptides were extracted sequentially using extraction buffer containing 0.1%, 0.5%, and 5% Tri-fluoro acetic acid in 50% ACN from gel bands. Extracted peptides were desalted by using C18 Zip-Tip (Millipore) and then dried in a centrifugal vacuum concentrator. Peptide samples were reconstituted with 0.1% TFA in 50% ACN solution mixed 1:1 ratio with α-cyano-4-hydroxycinnamic acid solution (C8982; 10mg/ml)and then spotted onto metal target plate in triplicates. MS and MS/MS spectra were acquired using MALDI TOF/TOF (AbSciex TOF/TOF 5800). Acquired spectra were searched against taxonomy *Mycobacterium tuberculosis* from NCBInr database using mascot search engine in Protein Pilot Software for protein identification. Searches were performed allowing trypsin mis-cleavage up to 1. Carbamido-methylation of cysteine, oxidation of methionine was also included as variable and fixed modification respectively. The peptide mass tolerance was set as 100ppm for precursor ion and 0.8 Da for fragment ion with +1 charge. Since all the splicing and cleavage products were identified as *Mtu* FL-SufB, the acquired spectra were researched against a customized database containing probable splicing products and cleavage

products of SufB protein of *Mycobacterium tuberculosis* using mascot search engine in Protein Pilot Software with the same parameter. β-gal and BSA were run as internal calibration. Results were shown with protein score, percent coverage, theoretical molecular weight, and Iso-electric pH value. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [1] partner repository with the dataset identifier PXD015199. The details of the submission are given below.

For HPLC, *Mtu* FL-SufB protein was purified and renatured as aforementioned in the experimental procedures section. Analytical HPLC was carried out on an Agilent 1200 series instrument equipped with a Zorbax gf-450 column (6µm, 9.4x250mm) with a flow rate of 1ml/min. All HPLC runs used the following solvent: 1:1 water and isopropanol (solvent 1) and 0.1% TFA (Trifluoroacetic acid)in water (solvent 2). The column was equilibrated using solvent 1 and solvent 2. A protein sample was inserted into the column and ran for 30mins. The retention time vs protein intensity measured at 280nm was noted. For reference, we ran 6µl of Precision plus protein ladder (Biorad1610374) diluted to 30µl using Sodium phosphate buffer. The retention time vs molecular weight of the known protein standards was measured and plotted to make a standard curve. The unknown protein peaks from the test samples were compared with the standard curve to find out the relative molecular weights. The molecular weights of the expected fragments from the MALDI-TOF/TOF MS (Table S3) data were compared with the standard curve and molecular weights of the unknown peaks were determined.

Molecular weight determination of the SufB DM (H-5A/H-38A) by FPLC: H215A/H248A SufB mutant protein (DM) was isolated and purified under denaturing conditions as described earlier in the manuscript. The relative molecular weight of DM was confirmed by Gel filtration

chromatography using Hi prep 16/60 Sephacryl S200 GPC column (HiPrep 16/60 GE Healthcare, Cat. no. 10255928). 8M Urea buffer was removed from the protein sample by dialysis in 20mM sodium phosphate and 0.15M sodium chloride buffer with pH 7.4. Then the sample was concentrated using Protein concentrator (Thermo Scientific 88515).

For FPLC analysis the column was equilibrated with 2 column volume (cv) of sodium phosphate buffer (PH 7.4) followed by loading of molecular weight standard Bovine Serum Albumin (BSA) at a concentration of 1mg/ml. Then 2mg of purified DM protein was loaded onto the GPC column. The run was performed at a flow rate of 1ml/min. keeping the temperature at 20°C. The fractions eluted at an absorbance of 280nm were collected and analyzed by SDS-PAGE under reducing and non-reducing conditions.

2.8 Molecular Dynamics (MD) Simulations

2.8.1. Homology modeling and Validation of Mtu SufB

Chimera model of *Mtu* FL-SufB precursor with consensus secondary structure with predicted loop was built on 5AWF_A, 1VDE_A, 2CW7_A using bioluminate 2.7 from Schrödinger by advanced homology modeling. After completion of modeling, the structure was validated by analyzing the structures via different algorithms like in Ramachandran plot, PROSA(Z-score). RMSD values were calculated for the analysis of the validation of modeled structure(56).

2.8.2. Simulation of protein

Molecular dynamic (MD) simulations were performed using the Gromacs (V5.1.4) simulation package with double precision for better reproducibility (57). The interactions in the system were described by the OPLS-AA force field (58) combined with the SPC/E water model (59,

60). *Mtu* SufB (846 aa) was solvated in a cubic box with a periodic boundary condition at least 1 nm distance from box-edge and neutralized with NaCl. The system was minimized with target F_{max} of no greater than 1000 kJ mol^{-1} nm^{-1} with steepest descent minimization (61) with a spherical cut-off at 1 nm was imposed on all intermolecular interactions with verlet cut-off scheme (62).

The integration of the dynamic equations was done using the leap-frog algorithm with a time step of 2fs (63). The simulations were done under canonical NVT ensemble (constant number (N), volume (V), and temperature (T) isomolar, isochoric, isothermal) conditions for 100 ns, with the temperature maintained at 300K through velocity-rescale coupling and no temperature coupling. A spherical cut-off at 1 nm was imposed on all intermolecular interactions with verlet cut-off scheme(62). H-bonds are constrained using lincs with the order of 4 (64). The efficient particle-mesh Ewald (PME) algorithm was used for implementing long-range electrostatic interactions with the grid dimension of 0.16nm and interpolation order of 4 (65). Histidine at 248 positions was replaced with Alanine by homology modelling and MD simulations were done using GROMACS with similar parameters mentioned above. Trajectory files were analyzed with Gromacs, pymol and plotted using Origin 8.0.

3. RESULTS

3.1 Structural domains of *Mtu* full-length (FL) SufB precursor

We obtained the protein sequences for Fe-S cluster assembly protein SufB [*Mycobacterium tuberculosis* H37Rv], an 846 amino acid protein (Accession number WP_003407484.1, GI: 397673309) and the intein-less SufB protein [*Mycobacterium smegmatis* MC2 155], a 477 amino acid protein (Accession number YP_887437.1, GI: 118472504) from NCBI protein

database. A pair-wise sequence alignment using the ClustalW server very clearly delineated the intein sequence as well as identified ~95% sequence similarities between the extein sequences. A blast search against the sequences deposited in Inbase also identified the intein sequence as 359 amino acid *Mtu* SufB (*Mtu Pps1*) intein From *M. tuberculosis* strains H37Rv. Based on sequence alignment and structural analyses of SufB intein and homing endonuclease domain I-CreI and intein homing endonuclease Ii (PDB: 2CW7), the final demarcation of different structural domains of *M. tuberculosis* and their respective sequences appeared to be as shown in Table S2.

Residues encompassing N- & C-terminal extein and intein sequences from *Mtu* SufB precursor are illustrated in Figure 1A. Multiple sequence alignment of *Mtu* SufB precursor and other intein bearing proteins highlighted conservation of catalytic cysteines (C1 and C+1), Block B His67, Block F Asp77, penultimate His358, and terminal Asn359 in the intein and extein regions (Fig. 1B). Hence, SufB precursor protein (846 aa) is likely to follow a canonical intein splicing pathway to generate the following products (Fig. 1C); ligated N-(252aa) and C-(235 aa) exteins to form native SufB protein (487aa), SufB intein (359aa) with splicing domain (N- and C-terminal inteins; 155aa) and endonuclease domain (204aa). Off-pathway products may give rise to N-terminal cleavage products [NC(594aa) and N-extein (NE)] and C-terminal cleavage products [CC(611aa) and C-extein (CE)] (51, 66).

3.2 Conservation of His-5 and His-38 in different bacterial species

Our sequence analysis results indicate that SufB from different mycobacterial and archeal species exhibits a high degree of sequence similarities (Fig. 2A and B). This is further supported by phylogenetic analysis to show that these mycobacterial proteins possibly have a

common ancestral origin (67). Analysis cladogram of SufB intein sequences from different mycobacterial species shows significant similarities, except in *M. leprae*, *M. lepromatosis*, *M.triplex* and *M. Xenopi*, possibly due to different intein insertion sites (39). Divergence is also noticed in *Ferroplasma* for intein [Fig. 2C (ii)] and precursor sequence [Fig. 2C (i)] perhaps due to differences in the kingdoms . The discrepancies in the phylogenetic tree analyses of FL-SufB precursor proteins suggest that there are variations in the extein sequences and intein insertion sites. This suggests intein evolution and independent intein transfer in different species and kingdoms.

Further, two conserved histidines were identified in the N-extein region of different mycobacterial and archaeal SufB precursor proteins, His-5 and His-38 (Fig. 2A). His-5 is also conserved in intein-less SufB proteins from *Staphylococcus aureus*, *Bacillus subtilis*, and certain *mycobacteria* that use SUF complex as the sole pathway for [Fe-S] cluster generation (Fig. 2B) (68, 69). Although His-38 is conserved in both intein-bearing and intein-less mycobacterial SufB proteins, conservation is missing in intein-less bacterial systems like *Staphylococcus aureus* and *Bacillus subtilis* where SUF system is essential to synthesize [Fe-S] clusters (Fig. 2B).

Although His-5 is proximal to the N-terminal splice junction in the protein sequence our MD simulations results (explained later in this manuscript) detected the proximity of H-38 to catalytic Cys1 in the predicted structural model for *Mtu* SufB. Identification of two highly conserved metal-chelating residues proximal to N-terminal extein~intein junction raises the possibility of their regulatory roles on cleavage and/or splicing as well as in the functionality of [Fe-S] cluster assembly protein SufB. Despite their remote positions these two histidines possibly come closer to the active site and participate in protein splicing either via direct or

indirect interaction with the catalytic residues near cleavage site(s). A comprehensive kinetic study of *in-vitro* splicing and N-cleavage reactions are presented in the next section. Further mechanistic understanding of the roles of highly conserved histidines was explored via molecular dynamics (MD) simulations as shown later in this manuscript.

3.3 Kinetic study to evaluate the roles of conserved His-5 and His-38 on *Mtu* SufB intein splicing

Plasmid constructs for the current work were engineered as mentioned in the materials and methods section. Detection of two highly conserved His-5 and His-38 in the N-extein region raises the possibility of their regulatory effect on *Mtu* SufB intein cleavage and/or splicing either by interaction with catalytic residues directly or indirectly via other active site residues. To test this hypothesis, alanine substitution was made in place of His-5 and His-38 to generate single mutants (H-5A, H-38A) and double mutant (H-5A/H-38A, DM) (Materials and Methods). Alanine substitutions for active site residues have resulted in a complete blockage of cleavage and splicing reactions as shown by earlier studies (30). The negative control used for the above study was splicing inactive (SI) SufB double mutant. Substitution of both C1 and N359 to alanine (C1A/N359A) is expected to abolish intein splicing completely (70).

The above proteins were purified under denaturing conditions and refolded at 20°Cas described in the materials and methods. FL-SufB precursor, H-5A and H-38A SufB mutants gave rise to the precursor (P, 95.98KDa), N-terminal cleavage (NC, 66 KDa) product, N-extein (NE, 29.9KDa), ligated exteins (LE, 55.7 KDa), and intein (I, 40.2 KDa) upon *in vitro* refolding (Fig. 3). As predicted none of these products were seen in the case of SI double mutant [Fig. S2E (i)] and transformants expressing empty vector pACYC Duet-1 [Fig. S2E(ii)]. C-terminal cleavage

product (CC, 70.12 KDa) and C-extein (CE, 25.76 KDa) were not detected by SDS PAGE possibly due to protein degradation. Consequently, we were unable to analyze the effects of His-5Ala and His-38Ala mutation on the C-terminal cleavage reaction of *Mtu* SufB. However, we did confirm the presence of P, LE, NE, and CC products via western blot (Fig. 3B).

3.3.1 His-5Ala and His-38Ala SufB mutants exhibit attenuated splicing reaction

Splicing efficiency was determined following *in vitro* renaturation of FL- and mutant SufB proteins (H-5A and H-38A) over a period for 24h. Precursor (P) proteins detected at time 0h were quite comparable for FL- and mutant proteins (Fig. 3A). Ligated extein fractions were visualized at 0h as a possible soluble contaminant and accumulated product of protein purification. Densitometric analysis of protein products was performed via Gelquant.NET biochemical solutions. Splicing efficiency was calculated as a percentage of splicing [(I+LE/P+I+LE) x 100].

Splicing efficiency was plotted over time as pseudo-first-order reaction kinetics in GraphPad Prism version 9.0 or 8.0 (Fig. 3C). At 20°C, WT FL-SufB shows a splicing efficiency (Y_{max}) of (53.7±1.31) % with a rate constant (K) value of (9.92±0.001 x 10^{-3} min^{-1}) and a half-life ($t_{1/2}$) of 1.1h (Table 1). Under the same experimental conditions, H-5A SufB mutant splices out with maximum efficiency (Y_{max}) of (37.8±0.68)%, a rate constant (K) value of (3.3±0.0001 x10^{-3} min^{-1}) and a half-life ($t_{1/2}$) of 3.4 h. H-38A SufB mutant splices out with an efficiency (Y_{max}) of (43.35±1.7)%, the rate constant (K) (2.9±0.0003x10^{-3}), and half-life ($t_{1/2}$) 3.8h. This clearly demonstrates a 3 fold and 3.4 fold reduction ($p<0.0001$; one-way ANOVA) in splicing efficiency for H-5A and H-38A mutants, respectively, relative to FL-SufB precursor. These

observations further strengthen our hypothesis that histidines at -5 and -38 positions may have regulatory role on *Mtu* SufB intein splicing.

Further analysis was done by the kinetic study of H-5A/H-38A SufB DM under similar experimental conditions. Surprisingly, H-5A/H-38A DM expression gave rise to a faint precursor (P) band and a truncated product of 28KDa size detected by SDS PAGE analysis whereas western blot detected both the P and the truncated protein (Fig. 4 and Fig. S2). The 28KDa product was confirmed by FPLC as well (Fig. S3). The possibility of a frameshift during cloning was ruled out by repeat sequencing. A comparative analysis of splicing and cleavage products for FL-SufB, N359A SufB (cleavage mutant with isolated NC products), C+1A SufB (cleavage mutant with isolated NC products), and H-5A/H-38A SufB DM provided further clarification (Fig. 4). The 28KDa (H-5A/H-38A SufB DM) truncated protein was identified as a possible splicing intermediate or a product of protein degradation, but different from the N-extein (29.9KDa).

3.4 Kinetic study to assess the roles of His-5 and H-38 in the first amide and thioester equilibrium

Next, we examined TCEP [Tris(2-carboxyethyl)phosphine], DTT (Dithiothreitol), and HA (Hydroxylamine) induced N-cleavage reactions in FL- and mutant SufB proteins. DTT can act both as a nucleophile and reducing agent. Hydroxylamine acts as a nucleophile and like DTT targets the linear thioester intermediate in the first step of splicing (71). TCEP, a reducing agent was used as a control for DTT and HA-induced N-cleavage reactions (72).

Since the optimum temperature for FL- SufB splicing is found to be 20°C, N-cleavage reactions for the test proteins were also conducted at the same temperature. Purified proteins were

renatured in presence of different reducing agents, DTT (50mM), HA(0.5M), and TCEP(2mM) for different durations of time (materials and methods). Products of N-cleavage reactions were visualized by 4%~10% gradient SDS PAGE (Fig. 5, 6, and Fig. S4) and densitometric analysis of the cleavage products was performed by Gelquant.NET biochemical solutions. The percentage of N-cleavage reaction product was calculated as [(NE+NC/P+NE+NC) x 100].

3.4.1 H-5A and H-38A SufB mutants display reduced TCEP mediated N-cleavage reaction

TCEP mediated reaction was analyzed for 4h in FL- SufB and mutant SufB proteins. The mutant proteins clearly exhibited diminished production of N-cleavage (NC) products relative to FL-SufB [Fig. 5, Table 2(i)]. Reaction mediated by TCEP did not give a proper curve that could be fitted in linear or non-linear regression. In FL- SufB, % N-cleavage was accelerated until 40 min followed by a decline as splicing became distinct. H-5A and H-38A SufB mutants presented a sluggish course with the N-cleavage product peak at 60-80 min and then tapered off. At 40 min and 60 min, H-5A displayed 2 fold and 1.5 fold diminution in % N-cleavage respectively, relative to FL-SufB. Likewise, at 40 min and 60 min, H-38A mutant displayed a 3.8 fold and 3 fold reduction in % of N-cleavage respectively relative to FL-SufB (Fig. 5).

3.4.2 H-5A and H-38A SufB mutants exhibit reduced Hydroxylamine (HA) induced N-cleavage reaction

HA-induced N-cleavage reaction was monitored over a 24h period and then kinetics analysis was performed. The extracted data for HA-induced reactions were plotted over different time frames and the curve was fitted to a pseudo-first-order reaction with the equation, $Y=Y0 + (Plateau-Y0) *(1-exp^{(-K*x)})$, (Fig. 6). H-5A and H-38A mutants exhibited about 1.4 fold and 1.5 fold reduction ($p=0.0025$; one-way ANOVA) in % of N-cleavage, respectively, relative to FL-

SufB although the K and $t_{1/2}$ values were comparable [Fig. 6C, Table 2 (ii)]. HA is an alpha nucleophile that intercepts between amide-ester equilibrium and induces N-cleavage. It enhances the nucleophilicity of residues that leads to N-cleavage (73, 74). The mutation effect on HA-induced N-cleavage reaction was similar to what is observed in presence of TCEP. However, no significant differences in N-cleavage efficiency were noticed between H-5A, H-38A mutants, and FL- SufB in presence of DTT (Fig. S4 and Table S4). Similar to HA, DTT is also a nucleophile that enhances the nucleophilicity of residues and facilitates N- terminal cleavage reaction (72, 75). Besides, DTT interacts directly with the residues that may increase the flexibility of the active site structure (76). Possibly, this abolishes the effect of H-5A and H-38A mutation on N-terminal cleavage. Supported by our current results we hypothesize that histidines at -5 and -38 positions in the N-extein region may have a role in Cys1 activation directly or via structural alignment of other active site residues. This likely explains why catalytic cysteines Cys1 and Cys+1 are interacting differently in His mutants, hindering the N-cleavage reaction. Affirmation of the above hypothesis was done by MD simulations (upcoming section).

3.5 Confirmation of protein products

The identity of splicing and cleavage products (P, CC, LE, and NE) for the N-terminal 6X (His) tagged test proteins were confirmed via western blot using anti-His antibodies (Fig. 3B, 4B, 6B and S4B). SI (C1A/N359A) double mutant SufB was used as a control for the study and gave rise to unspliced precursor protein (P) [Fig. 3B (iv)]. Interestingly, one additional protein band was noticed above LE and below CC for FL- SufB, H-5A, and H-38A mutants possibly due to the generation of a splicing intermediate as an off-pathway product. We also observed a product

just below the P protein for FL-SufB, SI-SufB, H-5A, and H-38A mutants (Fig.3B, 6B, and S4B). However, this protein was missing in cells expressing SufB intein or empty vector pACYC Duet-1 [Fig. 3B (iv)]. Hereby, we conclude that we are getting a degradation product of SufB precursor in the cells over-expressing active and inactive *Mtu* full-length (FL) SufB precursor protein.

Next, MALDI-TOF/TOF mass spectrometry was performed for further confirmation of different splicing and cleavage products (Table S3). Different protein products were cut from SDS PAGE gel and subjected to protein identification by mass spectrometry. First, acquired MS and MS/MS spectra were searched against the taxonomy *Mycobacterium tuberculosis* from NCBInr (77) database using mascot search engine in Protein Pilot Software. It is found that all the splicing products were identified as SufB protein of *Mycobacterium tuberculosis complex* with a statistically significant score. As all the splicing products were derivatives of a single protein, the protein identification results confirmed the same (Table S3). Further, acquired MS and MS/MS spectra were checked against a customized database containing probable splicing and cleavage products of SufB protein of *Mycobacterium tuberculosis* using mascot search engine in Protein Pilot Software with same parameters.

The individual protein band spectra resulted in the identification of protein splicing products matching to it. Individual protein bands identification result is represented with protein score, percent coverage, and theoretical molecular weight and Isoelectric pH value in Table S3. Significance is measured from the expectancy-value (with a p-value ≤ 0.05).

We performed HPLC analysis to find out fragment weight against a known protein ladder as mentioned in materials and methods. Protein fragments of *Mtu* FL-SufB were found to be within the expected mass range as per retention time (RT) from HPLC (Fig. S1).

The molecular weight of SufB DM (H-5A/H-38A) protein was determined by FPLC, as mentioned in the experimental section. The eluted fractions of both, BSA standard and DM protein were analyzed by SDS gel electrophoresis under both reducing and non-reducing conditions, yielding molecular weights of 66KDa and 28KDa, respectively (Fig. S3). No significant difference was observed in their molecular weights under reducing and non-reducing conditions, thereby concluding the existence of both the proteins in their monomeric form.

3.6 Proposed mechanism for the cleavage of N- terminal intein~extein peptide bond

The current work investigates the mechanism of *Mtu* SufB intein cleavage at the N-terminal cleavage site between Gly252 (G-1) and Cys253 (C1). We have identified two highly conserved histidines at -5 (248, full-length protein) and -38 (215, full-length protein) positions of the N-extein sequence from different intein-carrying and intein-less bacterial and archeal SufB proteins. Our *in vitro* experiments suggested that His-5 (248) and His-38 (215) are likely to exert a regulatory role on SufB splicing and/or N-cleavage reactions.

To ascertain this hypothesis, we performed MD simulations to predict a 3-dimensional (3D) model for *Mtu* full-length (FL)-SufB precursor protein (Fig. 7A). This also included identification of possible critical residues at the SufB intein active site (N- and C-terminal splice junctions) (Fig. 7B and Fig. S8), given their proximity to catalytic residues Cys253 (C1), Cys612 (C+1), and Asn611 (N359).

A chimera structure of *Mtu* FL-SufB protein was built in Schrodinger Bioluminate 2.7 using different templates based on their sequence identity with *Mtu* FL-SufB (Fig. 7A). The sequence identity of *Mtu* FL-SufB to SufBCD complex from *Escherichia coli* (5AWF_A), PI-SceI; a homing endonuclease with protein splicing activity (1VDE_A) and intein homing endonuclease II (2CW7_A) was found to be 40%, 45%, and 22% respectively. The resultant structure of *Mtu* FL-SufB was found to have many loops in the intein region because of low sequence identity to the templates. In addition, nearly 13% of amino acids are found to be present in the disallowed region in the Ramachandran plot (Fig. S6A). Therefore, we carried out molecular dynamics (MD) simulations of the aforementioned sub-optimized *Mtu* SufB structure in presence of water and NaCl to mimic *in-vivo* conditions in GROMACS for proper folding of the protein. The structure was solvated in water using the SPC/E water model (59, 60) and interactions were described through the OPLS-AA force field (58). The system was minimized by steepest descent minimization with a target maximum force not greater than 1000 kJ mol^{-1}nm. The simulations were done using GROMACS with time steps of 2fs for a duration of 100ns.

After the 100ns simulation various parameters such as solvent accessible surface area (SASA) (Fig. S5B), root mean square deviation (RMSD) was computed and the structure was found to be stabilized after 80ns, suggesting that the dynamics have reached equilibrium (Fig. S5A). Models of *Mtu* FL-SufB and H-5A SufB mutant were simulated for 20ns. Their RMSD and Root Mean Square Fluctuation (RMSF) were compared (Fig. S7A and B). The equilibrated *Mtu* SufB structure obtained from the above MD simulations was found to have only ~1% amino acids in the disallowed region of the Ramachandran plot (Fig. S6B). We found that the long side chains of the amino acids that were present in the disallowed region of the Ramachandran

plot were exposed outside. The model was further validated by ProSA web before and after 100ns of MD simulation (Fig. S6C, D).

Next, the trajectory analysis of the *Mtu* FL-SufB structure was done to find out the possible role of critical residues including His248 (-5) present near the N- and C- cleavage sites. In particular, we followed the dynamics of His248 (-5) with respect to Gly252 (-1). Our MD simulations showed an interesting dynamical result. It was observed that His248 (-5) pushes another N-extein residue His215 (-38) towards the active site Gly252 (-1). In other words, the MD result suggests that His248 (-5) and His215 (-38) act in a concerted manner to facilitate *Mtu* SufB N-cleavage reaction. We should point out here, although His248 (-5) is closer to the active site compared to His215 (-38) in the primary sequence, our 3-dimensional structure finds the opposite i. e. His215 (-38) is closer to the active site relative to His248 (-5). Specifically, the shortest dynamical distance between His215 (-38) and Gly252 (-1) was observed to be 6.83Å, similar to catalytic distances observed between active residues in other intein systems (12, 34). Therefore, we propose that His248 (-5) and His215 (-38) act concordantly to facilitate the *Mtu* SufB N-cleavage reaction as explained in the discussion section.

Based on the above MD results, a schematic cleavage reaction is shown in (Fig. 7C) along the line of canonical intein splicing mechanism (78). Thus, we propose that conjoint effects of active site (N-terminal intein~extein splice junction) residues such as Gly252 (-1), Cys253 (1), and His215 (-38) supported by structural and conformational changes due to His248 (-5) contribute towards *Mtu* SufB N-terminal cleavage reaction. In order to provide further confidence to our structural and dynamical data, quantum mechanical (QM) calculations were required to obtain energetics of N-cleavage reaction and compared with earlier results. All the quantum chemical calculations were performed with the Gaussian16 package (32, 79, 80). All

geometry optimization of the molecules was computed using the B3LYP method of DFT theory at 6-31G (d) level with charge-neutral singlet (32, 79, 81). Similar techniques were used in the study of intein cleavage of *Mtu* Rec Intein (17). After the optimized structures of the reactants and the products for each step were obtained, they could be used as the initial and final states respectively in the TS module to find the corresponding energy barriers. Frequency analyses were performed at the same level to identify all of the stationary points of the intermediates (only real frequencies) and the saddle points of the transition states (one imaginary frequency). The reaction energy barrier was obtained from the energy difference between the transition state and reactant of the reaction with the zero-point correction. All stationary states were confirmed by means of vibrational frequency analysis, and the Gibbs free energy at physiological temperature was calculated for all stationary points, including the zero-point energy, the entropy, and the thermal energy (all at 298.15K) at the B3LYP/6-31G (d) (Fig. 7D, 7E). After exploring the reaction energy landscape and obtaining the optimized geometries for intermediates, transition states, and products, we hereby propose a detailed reaction mechanism for *Mtu* full-length SufB N-cleavage reaction as elucidated in the discussion section (Fig. 7C).

Further, the amino acids near the C-cleavage site in the 3D model of *Mtu* SufB are identified as His610 (358), Asn 611 (359), and Cys612 (+1) (Fig. S8). The distance between Asn611 (N359) and Cys612 (C+1) was found to be 5.52± 0.11Å and the distance between Asn611 (N359) and His610 (H358) was 4.33±0.25Å (Table S5). Due to their proximity, the C-cleavage reaction may be facilitated by interactions between these three amino acids. Future research could shed light on the possible roles of these residues in the mechanism of C-cleavage reaction for *Mtu* FL- SufB.

4. DISCUSSION

4.1. Mtu SufB harbors a classic intein splicing system with two conserved histidine residues in the N-extein region

SufB is the central and essential component of the mycobacterial SUF system. This work reports that SufB protein is well conserved in mycobacteria, bacteria, and archaea during the evolutionary course. Intein insertion points may vary in some species, but the conservation of critical residues in intein and extein regions does exist. Intein-carrying SufB proteins are found in different mycobacterial species (Fig. 2) (51). Simple sequence analysis depicted the conservation of SufB intein in different organisms and phylogenetic analysis of intein and whole sequence inferred the possibility of common ancestor through evolution (Fig. 2). Cladogram analysis suggested a trivial diversion in *M. leprae, M. lepromatosis, M. xenopi, M. triplex* and *Ferroplasma*. The discrepancies in the phylogenetic tree analyses of un-spliced SufB precursors indicated that there are diversity in the extein sequences and intein insertion sites. This suggests intein evolution and independent intein transfer in different species and kingdoms.

Domain analysis of *Mtu* SufB precursor; an 846 aa residue protein clearly demarcated the intein and extein structural regions (Fig. 1, Table S2). We found that *Mtu* SufB intein spanning over 359 residues has an intact endonuclease domain. Catalytic residues critical for classic intein splicing pathways like Cys1 (Block A), Cys+1 (Block G), penultimate His (Block G), and terminal Asn (Block G) are conserved in different mycobacterial species including *Mtu*. Besides, *Mtu* SufB intein also contains conserved Block B His67 that is known to catalyze the first N-S acyl shift by destabilizing the scissile peptide bond due to loss of resonance and

reduction in energy barrier. *In vitro* refolding of *Mtu* SufB precursor gave rise to splicing and cleavage products confirming the catalytic roles of the aforementioned conserved residues. Thus, we conclude that *Mtu* SufB carries a canonical (classic) cis- splicing intein system (Fig. 1). Furthermore, sequence analysis also detected two highly conserved His residues in the N-extein sequence of *Mtu* SufB protein; His-38 and His-5 located in all SufB proteins where SUF constitutes the exclusive pathway for [Fe-S] cluster generation irrespective of genus and kingdoms (Fig. 2A, B). Later, these histidines were identified to play key regulatory roles during *Mtu* SufB N-cleavage reaction.

4.2. A distinct extein~intein partnership mechanism guided by highly conserved H-5 and H-38 regulates Mtu SufB N-cleavage reaction

Catalytic residues such as Cys1, Cys+1, and terminal Asn directly participate in intein splicing via promoting sequential nucleophilic displacement reactions or by rearrangement of bonds near splice junctions. Non-catalytic residues assist indirectly via activation of active site residues and stabilization of various intermediate structure(s). Conserved His residues within intein sequence are known to play important roles during protein splicing as well. Block B His accelerates N-S acyl shift and cleavage of N-terminal intein~extein peptide bond whereas F- and G-Block His are crucial in the coordination of terminal Asn cyclization and cleavage of C-terminal splice site. We found that *Mtu* SufB displays a unique mechanism where two remote histidines located in the N-extein sequence work together to facilitate N-cleavage reaction via activation of catalytic Cys1.

To ascertain the roles of H-5 and H-38, a detailed kinetic analysis was performed on *Mtu* SufB full-length (FL) precursor and mutant proteins (H-5A and H-38A) [Fig. 3, 4, and 5]. We found

that both H-5A and H-38A mutants exhibit a sluggish splicing reaction relative to FL- SufB protein under optimum experimental conditions (Table 1). N-cleavage reaction kinetics of the aforesaid histidine mutants were examined in presence of TCEP and nucleophiles such as Hydroxylamine and DTT. H-5A and H-38A mutants demonstrated a significant reduction in N-cleavage efficiency in comparison to FL- SufB protein in presence of TCEP and nucleophile HA (Table 2 (i) and (ii), Fig.5 and 6). But DTT-induced thiolysis reversed the effect of alanine mutation in the histidine mutants. DTT is a thiol protectant reducing agent and exhibits a more efficient thiolysis and di-sulfide reduction. A strong nucleophile such as DTT abolishes the inhibitory effect of alanine mutation on SufB N-cleavage reaction. Herein, we may conclude that H-5 and H-38 possibly have an assistive role in *Mtu* SufB N-cleavage reaction along with other active site residues. Consequently, the effect of H-5A and H-38A mutation can be reversed in presence of strong nucleophile(s) like DTT (Fig. S4).

To further ascertain the roles of conserved histidines, a 3D model of the *Mtu* FL-SufB protein was built by homology modeling. Molecular dynamics (MD) simulations were performed to equilibrate the aforementioned sub-optimized *Mtu* SufB structure. We obtained an interesting result during the trajectory analysis of the dynamical data. Although H-38 is relatively distal to the N-terminal splice junction in the protein sequence, in the 3-dimensional structure, it comes closer to the active site. MD results suggest that His248 (-5) and His215 (-38) act in a concerted manner to facilitate *Mtu* SufB N-cleavage reaction. Thus, we propose that concomitant effects of N-terminal active site residues such as Gly252 (-1), Cys253 (1), and His215 (-38) supported by His248 (-5) contribute towards *Mtu* SufB N-terminal cleavage reaction. Next, QM/MM calculations were done to determine the energetics of the proposed N-cleavage reaction. Taken together, we hereby propose a detailed mechanism for *Mtu* SufB N-cleavage reaction. A

schematic for the same is depicted in (Fig. 7C) along the line of canonical intein splicing mechanism.

In the process to form thioester intermediate from peptide bond of Gly-Cys dipeptide capped with a methyl group, the steps are as follows. The thiol group of Cys253 (1) comes closer to peptidyl C=O by a conformational change to a rotational transition state. This state is enthalpy favored. Then the thiol H is polarized towards the peptidyl O atom. The C=O and S-H experiences change in charge density in the system. The C, S, H atoms in the system experience decrease in electron density, and the O atom experiences an increase in electron density. In the next transition state, C and S come closer as at this state the imaginary vibrational frequency displacement vector of C=O is towards thiol. Thus, a C-S bond is formed. The proton migration from S to O is highly favored energetically and entropically with a free energy change of -9.2 kcal/mol. The computed value agrees well with earlier results (82, 83). After this step sp^2 hybridized peptide carbon becomes sp^3 hybridized forming a tetrahedral transition state, a 2-hydroxy thiazolidine ring. This particular state annihilates the peptide resonance, therefore free energy change of +3.7 kcal/mol in contrast to the ground state of Gly252 (-1) and Cys253 (1) system. To again regain the sp^2 hybridization and decrease the energy of the system from resonance stabilization, the 2-hydroxy group expels proton to the environment. This process is energetically advantageous to the system elucidating negative free energy. In a tetrahedral state, S and N atoms have a Mulliken charge of +0.05 and -0.55. Thus, the N atom of the thiazolidine ring accepts a proton from the environment to the system for charge neutrality. This process is also entropically preferred to form thioester intermediate (82, 84). The conversion of a peptide bond to a thioester bond is energetically equivalent at the start as well as the end of the reaction which suggests this process is entropy-driven. Further thioester is hydrolyzed by the imidazole

side chain of His215 (-38) mediated by water. Base catalyzed thioester hydrolysis in an aqueous environment is well known (85). It is also important to point out that the least distance between His215 (-38) and His248 (-5) obtained in MD simulations after equilibrium is about 6.83Å suggesting catalytic distance for an interaction between these two residues (12, 13). Asp217 (-36) prepares protonated histidine His215 (-38) for catalyzing reactions by neutralizing charge through proton transfer. Furthermore, His248 (-5) which is exposed to the solvent, experiences random collisions from water molecules. This action forces His215 (-38) towards the N-cleavage site. This is corroborated by the dynamical distance pattern between His215 (-38) and His248 (-5). Thus, a congruent effort by His248 (-5) facilitates the thioester hydrolysis for cleavage by attacking species i.e. hydroxide ion generation catalyzed by His215 (-38).These results provide confidence in the proposed role of His248 (-5) and His215 (-38). Since histidines are known as metal chelating centers for various metalloproteins, conserved histidines at -5 and -38 positions might coordinate to [Fe-S] clusters contributing towards *Mtu* SufB functionality as well.

4.3. Biological significance of conserved histidines in the N-extein sequence of Mtu SufB

[Fe-S] cluster-bearing proteins have important physiological roles in electron transfer, redox regulation, metabolic pathways, cellular responses to external stimuli, and as regulators of gene expression (42). The role of the [Fe-S] cluster is closely associated with the functionality of their bound protein framework. In mycobacteria, the SUF system is the sole pathway for [Fe-S] cluster assembly and repair, especially in response to oxidative stress and iron limiting conditions inside macrophages (43, 47). During such an event, the intracellular Fe supply is either from siderophore chelation or via the metabolism of [Fe-S] clusters associated with specific proteins (42). Thus, iron homeostasis plays a major role in mycobacterial survival and

virulence. SufB is a [Fe-S] cluster scaffold protein and a vital component of the functional SUF system, and indirectly promotes mycobacterial persistence under stress (47). Further, SufB has been implicated in mycobacterial iron metabolism (42).

Earlier studies have specified the bonding of Fe in [Fe-S] clusters to mostly cysteines from the protein backbone although there is increasing evidence for other ligands such as histidine, aspartate, arginine, threonine, and tyrosine. The most common alternative ligand for [Fe-S] cluster coordination is histidine that is highly conserved with a role in redox tuning and proton-coupled electron transfer (86). His433 (SufB) and His360 (SufD) are identified as key protein-ligands for the *de novo* [Fe-S] cluster assembly in *E. coli* (49). It has been shown that these non-cysteine ligands can influence the stability and reactivity of [Fe-S] clusters.

We have identified two highly conserved His-5 and His-38 residues in the N-extein sequence of *Mtu* SufB. These were shown to interact concordantly to initiate N-cleavage reaction via catalytic Cys1 activation. Mutational analysis of alanine-substituted His-5 and His-38 residues revealed significantly attenuated cleavage and splicing rates. Attempts to express and analyze the *Mtu* SufB double mutant (His-5A/ His-38A) resulted in a truncated protein product. One of the possible explanations could be the concomitant role of His-5 and His-38 in SufB precursor stabilization. Lack of the histidine residues at the native position perhaps leads to loss of protein stability forming off-pathway splicing intermediate and/or protein degradation product. In summary, the two histidines (His-5 and His-38) work concurrently to activate/initiate N-cleavage reaction leading to the generation of functional SufB protein via protein splicing. As proposed by earlier studies, apart from Cys, His is also an important ligand for metal coordination (49, 86). Thus, it is plausible that the highly conserved His-5 and His-38 possibly coordinate to Fe^{+2}/ Fe^{+3} during [Fe-S] cluster biogenesis. Lastly, His-5 and His-38 together

might also play a role in *Mtu* SufB precursor stabilization prior to cleavage and splicing as we were unable to express a full-length SufB double mutant (His-5A and His-38A), unlike the single mutants. However, whether the SufB precursor stabilization is executed by coordination of conserved histidines to [Fe-S] cluster or bonding to Fe^{+2}/ Fe^{+3}, it needs extensive experimental and computational confirmation.

Suf is an exclusive system for biogenesis of Fe-S clusters in many pathogenic organisms such as *Staphylococcus aureus*, *Mycobacterium tuberculosis*, *Plasmodium* sp. and *Toxoplasma*, making it an attractive drug target. For instance, D-Cycloserine is a clinical second-line drug currently used against *M. tuberculosis* and inhibits SufS. Investigation on *S. aureus* revealed that SUF is the target system for a polycyclic molecule 882 that has direct interaction with SufC (69). Likewise, the current work may have an implication towards the development of novel anti-TB drugs targeting SufB destabilization, as a way to fight tuberculosis.

Data Availability: All the data are available in the manuscript and supplemental data.

The accession code for the *Mycobacterium tuberculosis* SufB protein used in this study is *P9WFP7 (UniProtKB/Swiss-Prot) and* WP_003407484.1, GI: 397673309 (NCBI protein database).

Confirmation of the identity of different splicing and cleavage products of Mtu FL-SufB protein by mass spectrometric analysis: The data has been deposited to the ProteomeXchange Consortium via PRIDE partner repository with data set identifier PXD015199. The details of the submission are given below.

Project Name: Identification of full length, splicing and cleavage products of SufB protein of SUF-complex of *Mycobacterium tuberculosis*.

Project accession: PXD015199

Reviewer account details:

Username- reviewer73867@ebi.ac.uk

Password: bRxRcbvP

Confirmation of the identity of H-38A (215)/H-5A (248) Mtu SufB double mutant product by mass spectrometric analysis: The data has been deposited to the ProteomeXchange Consortium via PRIDE partner repository with data set identifier PXD023785. The details of the project submission are given below.

Project Name: Identification of splicing and cleavage products of H-38A/H-5A SufB double mutant protein of SUF-complex of *Mycobacterium tuberculosis.*

Project accession: PXD023785

Reviewer account details:

Username- reviewer_pxd023785@ebi.ac.uk

Password: 1i6COWkr

The homology model for the full length Mtu SufB protein has been submitted to Model Archive [Project: ma-x807d]:

Full length *Mtu* SufB protein of SUF complex of *Mycobacterium tuberculosis*; [https://www.modelarchive.org/doi/10.5452/ma-x807d] with the access code: 6pmXRNkvwR.

Funding: Current work and Ananya Nanda are supported by UGC-DAE consortium for scientific research(UGC-DAE-CSR-KC/CRS/15/IOP/08/0562), Kolkata, India. Dr. Sunita Panda (201500000557/IF 140155) was supported by INSPIRE fellowship; INSPIRE Division, DST, Government of India.

Acknowledgements: Our sincere thanks to Prof. Marlene Belfort, Department of Biological Sciences and RNA Institute, University at Albany, Albany, NY, USA, for her valuable comments on the manuscript. Plasmids used in this work were engineered by Dr. Sasmita Nayak during her PhD work under the guidance of Prof. Marlene Belfort. We also thank Mr. Sourya

Subhra Nasker for providing critical comments on the manuscript. We are thankful to Mr. Rajendra Reddy for technical support at Central Proteomics Facility, Institute of Life sciences, and Bhubaneswar. We also thank Mr. A. Ananda Raman for his timely help in HPC (High Performance computing cluster) to perform computational work at National Institute of Science Education and Research (NISER) Bhubaneswar. Our sincere gratitude to Mrs. Anuradha Das for her help in conducting the HPLC experiment at School of Chemical Sciences, NISER Bhubaneswar and to Dr. Rahul Modak for his help in conducting FPLC at School of Biotechnology, KIIT Deemed to be University.

Conflict of Interest: Authors declare no conflict of interest.

Author contributions: Dr. Sunita Panda: Methodology; Validation; Formal analysis; Investigation; Resources; Data curation; Writing - original draft; Visualization; Funding acquisition. **Ananya Nanda:** Methodology; Validation; Formal analysis; Investigation; Resources; Data curation; Writing - original draft; Writing- Reviewing and Editing; Visualization. **Nilanjan Sahu:** Methodology; Software; Validation; Formal analysis; Writing - original draft. **Deepak Kumar Ojha:** Validation; Formal analysis; Investigation; Resources. **Dr. Biswaranjan Pradhan:** Formal analysis; Resources. **Anjali Rai**: Resources. **Dr. Amol R. Suryawanshi:** Methodology; Validation. **Dr. Nilesh Banavali**: Methodology; Software; Validation. **Dr. Sasmita Nayak:** Conceptualization; Methodology; Validation; Investigation; Data curation; Writing - original draft; Writing- Reviewing and Editing; Visualization; Supervision; Funding acquisition; Project administration.

References

1. Derbyshire, V., and Belfort, M. (1998) Lightning strikes twice: intron–intein coincidence. *Proceedings of the National Academy of Sciences* **95**, 1356-1357

2. Lennon, C. W., and Belfort, M. (2017) Inteins. *Current Biology* **27**, R204-R206
3. Starokadomskyy, P. (2007) Protein splicing. *Molecular Biology* **41**, 278-293
4. Novikova, O., Topilina, N., and Belfort, M. (2014) Enigmatic distribution, evolution, and function of inteins. *Journal of Biological Chemistry* **289**, 14490-14497
5. Elleuche, S., and Pöggeler, S. (2010) Inteins, valuable genetic elements in molecular biology and biotechnology. *Applied Microbiology and Biotechnology* **87**, 479-489
6. Lennon, C. W., Stanger, M., Banavali, N. K., and Belfort, M. (2018) Conditional Protein Splicing Switch in Hyperthermophiles through an Intein-Extein Partnership. *mBio* **9**, e02304-02317
7. Lennon, C. W., Stanger, M. J., and Belfort, M. (2019) Mechanism of single-stranded DNA activation of recombinase intein splicing. *Biochemistry* **58**, 3335-3339
8. Nanda, A., Nasker, S. S., Mehra, A., Panda, S., and Nayak, S. (2020) Inteins in Science: Evolution to Application. *Microorganisms* **8**, 2004
9. Panda, S., Nanda, A., Nasker, S. S., Sen, D., Mehra, A., and Nayak, S. (2021) Metal effect on intein splicing: A review. *Biochimie*
10. Lennon, C. W., Stanger, M., and Belfort, M. (2016) Protein splicing of a recombinase intein induced by ssDNA and DNA damage. *Genes & development* **30**, 2663-2668
11. Kelley, D. S., Lennon, C. W., Li, Z., Miller, M. R., Banavali, N. K., Li, H., and Belfort, M. (2018) Mycobacterial DnaB helicase intein as oxidative stress sensor. *Nature communications* **9**, 1-15
12. Eryilmaz, E., Shah, N. H., Muir, T. W., and Cowburn, D. (2014) Structural and dynamical features of inteins and implications on protein splicing. *Journal of Biological Chemistry* **289**, 14506-14511
13. Hiraga, K., Soga, I., Dansereau, J. T., Pereira, B., Derbyshire, V., Du, Z., Wang, C., Van Roey, P., Belfort, G., and Belfort, M. (2009) Selection and structure of hyperactive inteins: peripheral changes relayed to the catalytic center. *Journal of molecular biology* **393**, 1106-1117
14. Binschik, J., and Mootz, H. D. (2013) Chemical Bypass of Intein-Catalyzed N–S Acyl Shift in Protein Splicing. *Angewandte Chemie International Edition* **52**, 4260-4264
15. Du, Z., Shemella, P. T., Liu, Y., McCallum, S. A., Pereira, B., Nayak, S. K., Belfort, G., Belfort, M., and Wang, C. (2009) Highly Conserved Histidine Plays a Dual Catalytic Role in Protein Splicing: A p K a Shift Mechanism. *Journal of the American Chemical Society* **131**, 11581-11589
16. Du, Z., Zheng, Y., Patterson, M., Liu, Y., and Wang, C. (2011) p K a Coupling at the Intein Active Site: Implications for the Coordination Mechanism of Protein Splicing with a Conserved Aspartate. *Journal of the American Chemical Society* **133**, 10275-10282
17. Friedel, K., Popp, M. A., Matern, J. C., Gazdag, E. M., Thiel, I. V., Volkmann, G., Blankenfeldt, W., and Mootz, H. D. (2019) A functional interplay between intein and extein sequences in protein splicing compensates for the essential block B histidine. *Chemical science* **10**, 239-251
18. Mujika, J., and Lopez, X. (2017) Unveiling the Catalytic Role of B-Block Histidine in the N–S Acyl Shift Step of Protein Splicing. *The Journal of Physical Chemistry B* **121**, 7786-7796
19. Pereira, B., Shemella, P. T., Amitai, G., Belfort, G., Nayak, S. K., and Belfort, M. (2011) Spontaneous proton transfer to a conserved intein residue determines on-pathway protein splicing. *Journal of molecular biology* **406**, 430-442

20. Ding, Y., Xu, M.-Q., Ghosh, I., Chen, X., Ferrandon, S., Lesage, G., and Rao, Z. (2003) Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing. *Journal of Biological Chemistry* **278**, 39133-39142
21. Klabunde, T., Sharma, S., Telenti, A., Jacobs, W. R., and Sacchettini, J. C. (1998) Crystal structure of GyrA intein from Mycobacterium xenopi reveals structural basis of protein splicing. *Nature structural biology* **5**, 31-36
22. Frutos, S., Goger, M., Giovani, B., Cowburn, D., and Muir, T. W. (2010) Branched intermediate formation stimulates peptide bond cleavage in protein splicing. *Nature chemical biology* **6**, 527-533
23. Mujika, J. I., Lopez, X., and Mulholland, A. J. (2009) Modeling protein splicing: reaction pathway for C-terminal splice and intein scission. *The journal of physical chemistry B* **113**, 5607-5616
24. Shao, Y., and PAULUS, E. (1997) Protein splicing: estimation of the rate of O-N and S-N acyl rearrangements, the last step of the splicing process. *The Journal of peptide research* **50**, 193-198
25. Johnson, M. A., Southworth, M. W., Herrmann, T., Brace, L., Perler, F. B., and Wüthrich, K. (2007) NMR structure of a KlbA intein precursor from Methanococcus jannaschii. *Protein Science* **16**, 1316-1328
26. Southworth, M. W., Benner, J., and Perler, F. B. (2000) An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. *The EMBO Journal* **19**, 5019-5026
27. Brace, L. E., Southworth, M. W., Tori, K., Cushing, M. L., and Perler, F. (2010) The Deinococcus radiodurans Snf2 intein caught in the act: detection of the Class 3 intein signature Block F branched intermediate. *Protein Science* **19**, 1525-1533
28. Tori, K., Dassa, B., Johnson, M. A., Southworth, M. W., Brace, L. E., Ishino, Y., Pietrokovski, S., and Perler, F. B. (2010) Splicing of the mycobacteriophage Bethlehem DnaB intein: identification of a new mechanistic class of inteins that contain an obligate block F nucleophile. *Journal of Biological Chemistry* **285**, 2515-2526
29. Tori, K., and Perler, F. B. (2011) Expanding the definition of class 3 inteins and their proposed phage origin. *Journal of bacteriology* **193**, 2035-2041
30. Amitai, G., Callahan, B. P., Stanger, M. J., Belfort, G., and Belfort, M. (2009) Modulation of intein activity by its neighboring extein substrates. *Proceedings of the National Academy of Sciences* **106**, 11005-11010
31. Shah, N. H., Eryilmaz, E., Cowburn, D., and Muir, T. W. (2013) Extein residues play an intimate role in the rate-limiting step of protein trans-splicing. *Journal of the American Chemical Society* **135**, 5839-5847
32. Shemella, P., Pereira, B., Zhang, Y., Van Roey, P., Belfort, G., Garde, S., and Nayak, S. K. (2007) Mechanism for intein C-terminal cleavage: a proposal from quantum mechanical calculations. *Biophysical journal* **92**, 847-853
33. Sun, P., Ye, S., Ferrandon, S., Evans, T. C., Xu, M.-Q., and Rao, Z. (2005) Crystal structures of an intein from the split dnaE gene of Synechocystis sp. PCC6803 reveal the catalytic model without the penultimate histidine and the mechanism of zinc ion inhibition of protein splicing. *Journal of molecular biology* **353**, 1093-1105

34. Topilina, N. I., Novikova, O., Stanger, M., Banavali, N. K., and Belfort, M. (2015) Post-translational environmental switch of RadA activity by extein–intein interactions in protein splicing. *Nucleic acids research* **43**, 6631-6648
35. Cheriyan, M., Pedamallu, C. S., Tori, K., and Perler, F. (2013) Faster protein splicing with the Nostoc punctiforme DnaE intein using non-native extein residues. *Journal of Biological Chemistry* **288**, 6202-6211
36. Chong, S., Williams, K. S., Wotkowicz, C., and Xu, M.-Q. (1998) Modulation of protein splicing of the Saccharomyces cerevisiae vacuolar membrane ATPase intein. *Journal of Biological Chemistry* **273**, 10567-10577
37. Oeemig, J. S., Zhou, D., Kajander, T., Wlodawer, A., and Iwaï, H. (2012) NMR and crystal structures of the Pyrococcus horikoshii RadA intein guide a strategy for engineering a highly efficient and promiscuous intein. *Journal of molecular biology* **421**, 85-99
38. Shemella, P. T., Topilina, N. I., Soga, I., Pereira, B., Belfort, G., Belfort, M., and Nayak, S. K. (2011) Electronic structure of neighboring extein residue modulates intein C-terminal cleavage activity. *Biophysical journal* **100**, 2217-2225
39. Topilina, N. I., Green, C. M., Jayachandran, P., Kelley, D. S., Stanger, M. J., Piazza, C. L., Nayak, S., and Belfort, M. (2015) SufB intein of Mycobacterium tuberculosis as a sensor for oxidative and nitrosative stresses. *Proceedings of the National Academy of Sciences* **112**, 10348-10353
40. Mills, K. V., Johnson, M. A., and Perler, F. B. (2014) Protein splicing: how inteins escape from precursor proteins. *Journal of Biological Chemistry* **289**, 14498-14505
41. Cheriyan, M., Chan, S.-H., and Perler, F. (2014) Traceless splicing enabled by substrate-induced activation of the Nostoc punctiforme Npu DnaE intein after mutation of a catalytic cysteine to serine. *Journal of molecular biology* **426**, 4018-4029
42. Huet, G., Daffé, M., and Saves, I. (2005) Identification of the Mycobacterium tuberculosis SUF machinery as the exclusive mycobacterial system of [Fe-S] cluster assembly: evidence for its implication in the pathogen's survival. *Journal of bacteriology* **187**, 6137-6146
43. Outten, F. W. (2015) Recent advances in the Suf Fe–S cluster biogenesis pathway: beyond the Proteobacteria. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1853**, 1464-1469
44. Lin, C.-W., McCabe, J. W., Russell, D. H., and Barondeau, D. P. (2020) Molecular Mechanism of ISC Iron–Sulfur Cluster Biogenesis Revealed by High-Resolution Native Mass Spectrometry. *Journal of the American Chemical Society* **142**, 6018-6029
45. Mettert, E. L., and Kiley, P. J. (2015) Fe–S proteins that regulate gene expression. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1853**, 1284-1293
46. Novikova, O., Jayachandran, P., Kelley, D. S., Morton, Z., Merwin, S., Topilina, N. I., and Belfort, M. (2016) Intein clustering suggests functional importance in different domains of life. *Molecular biology and evolution* **33**, 783-799
47. Huet, G., Castaing, J.-P., Fournier, D., Daffé, M., and Saves, I. (2006) Protein splicing of SufB is crucial for the functionality of the Mycobacterium tuberculosis SUF machinery. *Journal of bacteriology* **188**, 3412-3414
48. Outten, F. W., Djaman, O., and Storz, G. (2004) A suf operon requirement for Fe–S cluster assembly during iron starvation in Escherichia coli. *Molecular microbiology* **52**, 861-872

49. Yuda, E., Tanaka, N., Fujishiro, T., Yokoyama, N., Hirabayashi, K., Fukuyama, K., Wada, K., and Takahashi, Y. (2017) Mapping the key residues of SufB and SufD essential for biosynthesis of iron-sulfur clusters. *Scientific reports* **7**, 9387
50. Larking, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, G., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., and Lopez, R. (2007) ClustalW and ClustalX version 2. *Bioinformatics* **23**, 2947-2948
51. Perler, F. B. (2002) InBase: the intein database. *Nucleic acids research* **30**, 383-384
52. Morgenstern, B. (1999) DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics (Oxford, England)* **15**, 211-218
53. Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution* **35**, 1547-1549
54. Truszkowski, J., and Goldman, N. (2016) Maximum likelihood phylogenetic inference is consistent on multiple sequence alignments, with or without gaps. *Systematic biology* **65**, 328-333
55. Mills, K. V., and Paulus, H. (2001) Reversible inhibition of protein splicing by zinc ion. *Journal of Biological Chemistry* **276**, 10832-10838
56. Wiederstein, M., and Sippl, M. J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research* **35**, W407-W410
57. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19-25
58. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B* **105**, 6474-6487
59. Berendsen, H., Grigera, J., and Straatsma, T. (1987) The missing term in effective pair potentials. *Journal of Physical Chemistry* **91**, 6269-6271
60. Kusalik, P. G., and Svishchev, I. M. (1994) The spatial structure in liquid water. *Science* **265**, 1219-1221
61. Fletcher, R., and Powell, M. J. (1963) A rapidly convergent descent method for minimization. *The computer journal* **6**, 163-168
62. Grubmüller, H., Heller, H., Windemuth, A., and Schulten, K. (1991) Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Molecular Simulation* **6**, 121-142
63. Barnes, J., and Hut, P. (1986) A hierarchical O (N log N) force-calculation algorithm. *nature* **324**, 446
64. Hess, B., Bekker, H., Berendsen, H. J., and Fraaije, J. G. (1997) LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* **18**, 1463-1472
65. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *The Journal of chemical physics* **98**, 10089-10092
66. Perler, F. B. (1998) Protein splicing of inteins and hedgehog autoproteolysis: structure, function, and evolution. *Cell* **92**, 1-4

67. Pearson, W. R. (2013) An introduction to sequence similarity ("homology") searching. *Current protocols in bioinformatics* **42**, 3.1. 1-3.1. 8
68. Albrecht, A. G., Netz, D. J., Miethke, M., Pierik, A. J., Burghaus, O., Peuckert, F., Lill, R., and Marahiel, M. A. (2010) SufU is an essential iron-sulfur cluster scaffold protein in Bacillus subtilis. *Journal of bacteriology* **192**, 1643-1651
69. Pérard, J., and de Choudens, S. O. (2018) Iron–sulfur clusters biogenesis by the SUF machinery: close to the molecular mechanism understanding. *JBIC Journal of Biological Inorganic Chemistry* **23**, 581-596
70. Zhang, B., Rapolu, M., Liang, Z., Han, Z., Williams, P. G., and Su, W. W. (2015) A dual-intein autoprocessing domain that directs synchronized protein co-expression in both prokaryotes and eukaryotes. *Scientific reports* **5**, 8541
71. Kirby, A. J., Souza, B. S., Medeiros, M., Priebe, J. P., Manfredi, A. M., and Nome, F. (2008) Hydroxylamine as an oxygen nucleophile. Chemical evidence from its reaction with a phosphate triester. *Chemical Communications*, 4428-4429
72. Bachmann, A.-L., and Mootz, H. D. (2015) An unprecedented combination of serine and cysteine nucleophiles in a split intein with an atypical split site. *Journal of Biological Chemistry* **290**, 28792-28804
73. Simanenko, Y. S., Popov, A., Prokop'eva, T., Savelova, V., and Belousova, I. (1994) Hydroxylamine anion—an effective α-nucleophile in acyl transfer reactions. *Theoretical and Experimental Chemistry* **30**, 61-64
74. Xu, M.-Q., and Perler, F. B. (1996) The mechanism of protein splicing and its modulation by mutation. *The EMBO journal* **15**, 5146-5153
75. Volkmann, G., Sun, W., and Liu, X. Q. (2009) Controllable protein cleavages through intein fragment complementation. *Protein Science* **18**, 2393-2402
76. Alliegro, M. C. (2000) Effects of Dithiothreitol on Protein Activity Unrelated to Thiol–Disulfide Exchange: For Consideration in the Analysis of Protein Function with Cleland's Reagent. *Analytical biochemistry* **282**, 102-106
77. Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research* **33**, D501-D504
78. Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, F. B., and Xu, M.-Q. (1996) Protein Splicing Involving the Saccharomyces cerevisiae VMA Intein The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system. *Journal of Biological Chemistry* **271**, 22159-22168
79. Dearden, A. K., Callahan, B., Van Roey, P., Li, Z., Kumar, U., Belfort, M., and Nayak, S. K. (2013) A conserved threonine spring-loads precursor for intein splicing. *Protein Science* **22**, 557-563
80. Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., Montgomery Jr, J., Vreven, T., Kudin, K., and Burant, J. (2004) Gaussian 03, Rev. C. 02, Gaussian Inc. *Wallingford, CT*
81. Becke, A. D. (1993) Density-functional thermochemistry. III. The role of exact exchange. *The Journal of chemical physics* **98**, 5648-5652
82. Burke, H. M., McSweeney, L., and Scanlan, E. M. (2017) Exploring chemoselective S-to-N acyl transfer reactions in synthesis and chemical biology. *Nature communications* **8**, 1-16

83. Devaraj, N. K., and Perrin, C. L. (2018) Approach control. Stereoelectronic origin of geometric constraints on N-to-S and N-to-O acyl shifts in peptides. *Chemical science* **9**, 1789-1794
84. Monbaliu, J.-C. M., Dive, G., Stevens, C. V., and Katritzky, A. R. (2013) Governing parameters of long-range intramolecular S-to-N acyl transfers within (S)-acyl isopeptides. *Journal of chemical theory and computation* **9**, 927-934
85. Kirsch, J. F., and Jencks, W. P. (1964) Base catalysis of imidazole catalysis of ester hydrolysis. *Journal of the American Chemical Society* **86**, 833-837
86. Bak, D. W., and Elliott, S. J. (2014) Alternative FeS cluster ligands: tuning redox potentials and chemistry. *Current opinion in chemical biology* **19**, 50-58

TABLES:

Table 1. Comparative analysis of different kinetic parameters for splicing in FL-SufB precursor, H-5A and H-38A SufB mutant proteins, at 20˚C temperature. These data were extracted from Fig. 3C.

Kinetic parameters	**FL-SufB**	**H-5A**	**H-38A**
Y_{max}	53.81±1.03	37.82±1.68	43.35±1.7
Rate constant (K) (min^{-1})	$9.92 \pm 0.007 \times 10^{-3}$	$3.3 \pm 0.0003 \times 10^{-3}$	$2.9 \pm 0.0003 \times 10^{-3}$
Half-life($t_{1/2}$)	1.1h	3.4h	3.8h

Table 2. (i) Comparative analysis of TCEP mediated N-terminal cleavage reaction at 20^0C over different time up to 240 mins. The data were extracted from Fig.5B. (ii) Different kinetic parameters (Y_{max}, rate constant, and half-life) for HA-induced N-terminal cleavage in FL-SufB precursor, H-5A, and H-38A SufB mutants. These data were extracted from Fig. 6C.

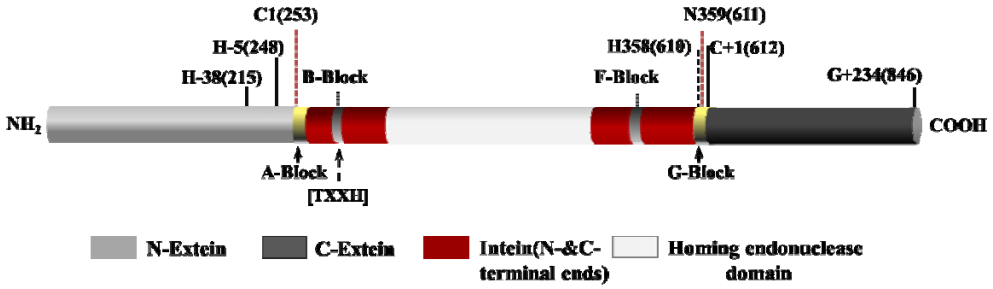
(i)

Time (min)	FL-SufB (% of N-cleavage)	H-5A (% of N-cleavage)	H-38A (% of N-cleavage)
20	28.2±5.5	15.1±0.97	7.42±2
40	36±4.3	17.89±2.7	9.3±2.7
60	30.63±5.1	20.76±3.7	10.2±1.7
80	25.7±5.7	14.9±2.5	12.42±1.9
120	19.8±1.6	8.4±0.7	10.5±2.7
240	16.78±1.5	6.2±0.89	7.07±1.9

(ii)

Kinetic parameters	FL-SufB	H-5A	H-38A
Y_{max}	50.67±1.6	35.72±1.4	32.45±0.94
Rate Constant(K) (min^{-1})	3.7±0.01 x10^{-2}	3.9±0.01x10^{-2}	3.7±0.009x10^{-2}
Half-life($t_{1/2}$)	18.5 min	17.3 min	18.2 min

A



B

	Block A	Block B	Block F	Block G
Mtu SufB	CLPAGELITTADG	ANAFSVTAEHPLLA	PVYNLDVENPDSYL	YGFAVHNC
Mbo SufB	CLPAGELITTADG	ANAFSVTAEHPLLA	PVYNLDVENPDSYL	YGFAVHNC
Maf SufB	CLPAGELITTADG	ANAFSVTAEHPLLA	PVYNLDVENPDSYL	YGFAVHNC
Mtu RecA	CLAEGTRIFDPVT	GAIVWATPDHKVLT	RTFDLEVEELHTLV	EGVVVHNC
Npu DnaE	CLSYETEILTVEY	GSVIRATSDHRFLT	NVYDIGVERDHNFA	NGFIASNC
Npu DnaB	CLAGDSLVTLVDS	GRKIRATGNHKFLT	EVFDLTVPGLHNFV	NNIIVHNS
Ssp DnaE	CLSFGTEILTVEY	GSVIRATSDHRFLT	RIFDiGLPQDHNFL	NGAIAANC
Sce VMA	CFAKGTNVLMADG	LLKFTCNATHELVV	YGITLSDDSDHQFL	NQVVVHNC

C

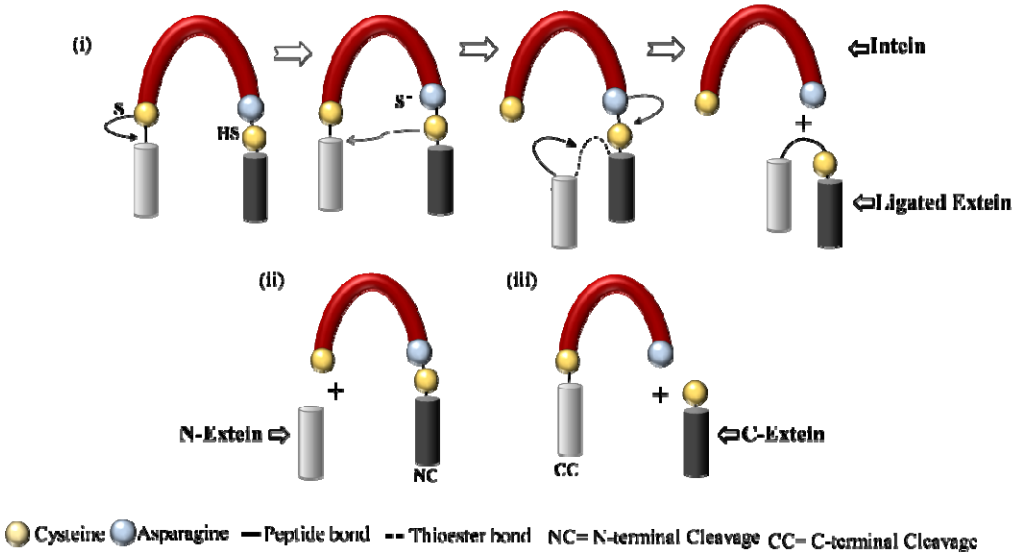


Fig. 1. (A) Schematic representation of different structural domains of *Mtu* full-length SufB precursor. The conserved and the catalytic residues in the FL-SufB protein sequence are shown at the top. Further, numbering for different structural domains are also displayed; intein residues (1, 2……,n), N-extein residues (-1, -2,…..-n) and C-extein residues (+1, +2,……+n). Catalytic cysteines responsible for sequential nucleophilic displacement reactions are C1 and C+1, terminal intein residue responsible for branched intermediate resolution is N359, conserved histidines in the N-extein region are H-5 (248, full-length protein) and H-38 (215, full-length protein). G+234 is the last residues of *Mtu* SufB, **(B)** Multiple sequence alignment of *Mtu* SufB with other intein bearing species shows conservation of critical residues (highlighted in different colors) that participate in a canonical splicing pathway. **(C)** Schematic diagram showing the mechanism of intein splicing. (i) There are four major reaction steps; first two steps of sequential nucleophilic attack by C1 and C+1 generate branched intermediate by trans-esterification. Next, cyclization of terminal Asn resolves the branched intermediate, ligates exteins and cleaves off SufB intein, (ii) and (iii) represent off-pathway products from protein splicing; N- and C-terminal cleavage products respectively.

A

```
M.tuberculosis (206)   ---------- ---------- ----SFIYVP PGVHVDIPLQ AYFRINTENM (231)
M.microti      (206)   ---------- ---------- ----SFIYVP PGVHVDIPLQ AYFRINTENM (231)
M.africanum    (206)                              SFIYVP PGVHVDIPLQ AYFRINTENM (231)
M.bovis        (206)   ---------- ---------- ----SFIYVP PGVHVDIPLQ AYFRINTENM (231)
M.Bourgelatii  (199)   ---------- ---------- ----SFIYVP PGVHVDIPLQ AYFRINTENM (224)
M.Kansasii     (199)   ---------- ---------- ----SFIYVP PGVHVDIPLQ AYFRINTENM (224)
M.orygis       (206)   ---------- ---------- ----SFIYVP PGVHVDIPLQ AYFRINTENM (231)
M.gastri_W     (199)   ---------- ---------- ----SFIYVP PGVHVDIPLQ AYFRINTENM (224)
M.gastri_p     (453)   ---------- ---------- ----SFIYVP PGVHVDIPLQ AYFRINXENM (224)
M.canettii     (211)   ---------- ---------- ----SFIYVP PGVHVDIPLQ AYFRINTENM (236)
M.caprae       (206)   ---------- ---------- ----SFIYVP PGVHVDIPLQ AYFRINTENM (231)
M.Pinnipeddii  (206)                              SFIYVP PGVHVDIPLQ AYFRINTENM (231)
M.xenopi       (578)   ---------- ---------- ----SFIYVP PGVHVEIPLQ AYFRINTENM (603)
M.triplex      (577)   ---------- ---------- ----SFIYVP PGVHVDIPLQ AYFRINTENM (602)
M.leprae       (564)   EKPTYDTQVV GLENFVANGI VAHNSFIYVP PGVHVDIPLQ AYFRINTENM (613)
M.lepromatosis (563)   EKPTYDTQVV GLENFVANGI VAHNSFIYVP PGVHVDIPLQ AYFRINTENM (612)
Ferroplasma    (196)   ---------- ---------- ----SFLYVP KNVTIDMPLQ TYFRMNGEQS (221)

M.tuberculosis (232)   GQFERTLIIA DEGSYVHYVE GGLPAGELIT TADGDLRPIE SIRVGDFVTG (281)
M.microti      (232)   GQFERTLIIA DEGSYVHYVE GGLPAGELIT TADGDLRPIE SIRVGDFVTG (281)
M.africanum    (232)   GQFERTLIIA DEGSYVHYVE GGLPAGELIT TADGDLRPIE SIRVGDFVTG (281)
M.bovis        (232)   GQFERTLIIA DEGSYVHYVE GGLPAGELIT TADGDLRPIE SIRVGDFVTG (281)
M.Bourgelatii  (225)   GQFERTLIIA DEGSYVHYVE GGLPAGELIT TADGDLRPIE SIQVGDYVTG (274)
M.Kansasii     (225)   GQFERTLIIA DEGSYVHYVE GGVPAGELIT TADGDLRPIE SIRVGDYVSS (274)
M.orygis       (232)   GQFERTLIIA DEGSYVHYVE GGLPAGELIT TADGDLRPIE SIRVGDFVTG (281)
M.gastri_W     (225)   GQFERTLIIA DEGSYVHYVE GGLPAGELIT TADGDLRPIE SIRVGDYVTG (274)
M.gastri_p     (479)   GQFERTLIIV GENAYVTYVE G--------- ---------- ---------- (499)
M.canettii     (237)   GQFERTLIIA DEGSYVHYVE GGLPAGELIT TADGDLRPIE SIRVGDFVTG (286)
M.caprae       (232)   GQFERTLIIA DEGSYVHYVE GGLPAGELIT TADGDLRPIE SIRVGDFVTG (281)
M.Pinnipeddi   (232)   GQFERTLIIA DEGSYVHYVE GGLPAGELIT TADGDLRPIE SIRVGDFVTG (281)
M.xenopi       (604)   GQFERTLIIV DENAYVHYVE G--------- ---------- ---------- (624)
M.triplex      (603)   GQFERTLIIV DEGAYVHYVE G--------- ---------- ---------- (623)
M.leprae       (614)   GQFERTLIIA DTGSYVHYVE G--------- ---------- ---------- (634)
M.lepromatosis (613)   GQFERTLIIA DEGSYVHYVE G--------- ---------- ---------- (633)
Ferroplasma    (222)   GQFEITIVVA DEGAKVHYIE GGLPEDELIS QGDkfv-PIN ELLTDDSVVS (270)
```

B

```
M.tuberculosis    (181) EYFCTVIPAGDNKFSALNTAVWSCGSFIYVPPGVHVDIPLQAYFRINTENMGQFERTLII(240)
M.marinum         (174) QYFGTVIPAGDNKFSALNTAVWSGGSFIYVPPGVHVDIPLQAYFRINTENMGQFERTLII(245)
M.fortuitum       (181) KYFCTVIPAGDNKFSALNTAVWSCGSFIYVPPGVHVDIPLQAYFRINTENMGQFERTLII(251)
M.smegmatis       (171) QYFGTVIPAGDNKFSALNTAVWSGGSFIYVPPGVHVDIPLQAYFRINTENMGQFERTLII(241)
M.shigaense       (175) QYFCSVIPAGDNKFSALNTAVWSCGSFIYVPPGVHVDIPLQAYFRINTENMGQFERTLII(245)
S.aureus          (159) KYFASVVPAADNKFAALNSAVWSGGSFIYVPKNTKLDTPLQAYFRINSENMGQFERTLII(229)
B.subtilis        (159) EHWAKVIPPTDNKFAALNSAVWSCGSFIYVPKGVKVETPLQAYFRINSENMGQFERTLII(229)

M.tuberculosis    (241) ADECSYVHYVECCLPAGELITTADGDLRPIESIRVGDFVTGHDGRPHRVTAVQVRDLDGE(290)
M.marinum         (235) ADEGSYVHYVEG--------------------------------------------(246)
M.fortuitum       (241) ADECSYVHYVEG--------------------------------------------(252)
M.smegmatis       (231) VDEGAYVHYVEG--------------------------------------------(242)
M.shigaense       (235) VDEGAYVHYVEG--------------------------------------------(246)
S.aureus          (219) ADEGASVHYVEG--------------------------------------------(230)
B.subtilis        (219) VDEEASVHYVEG--------------------------------------------(230)
```

C

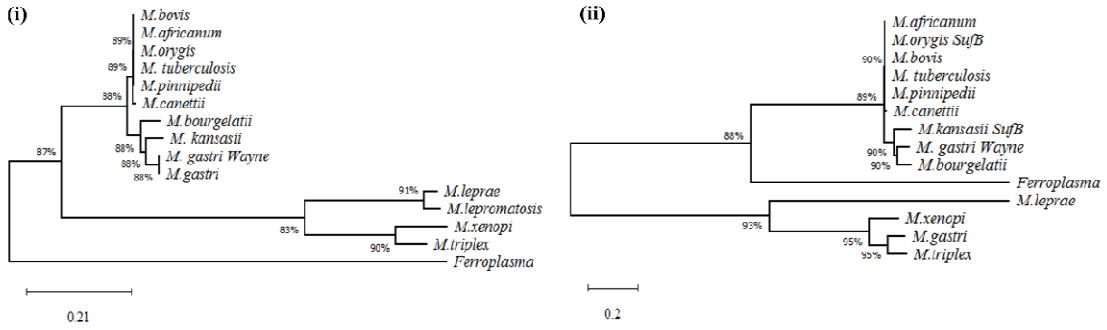
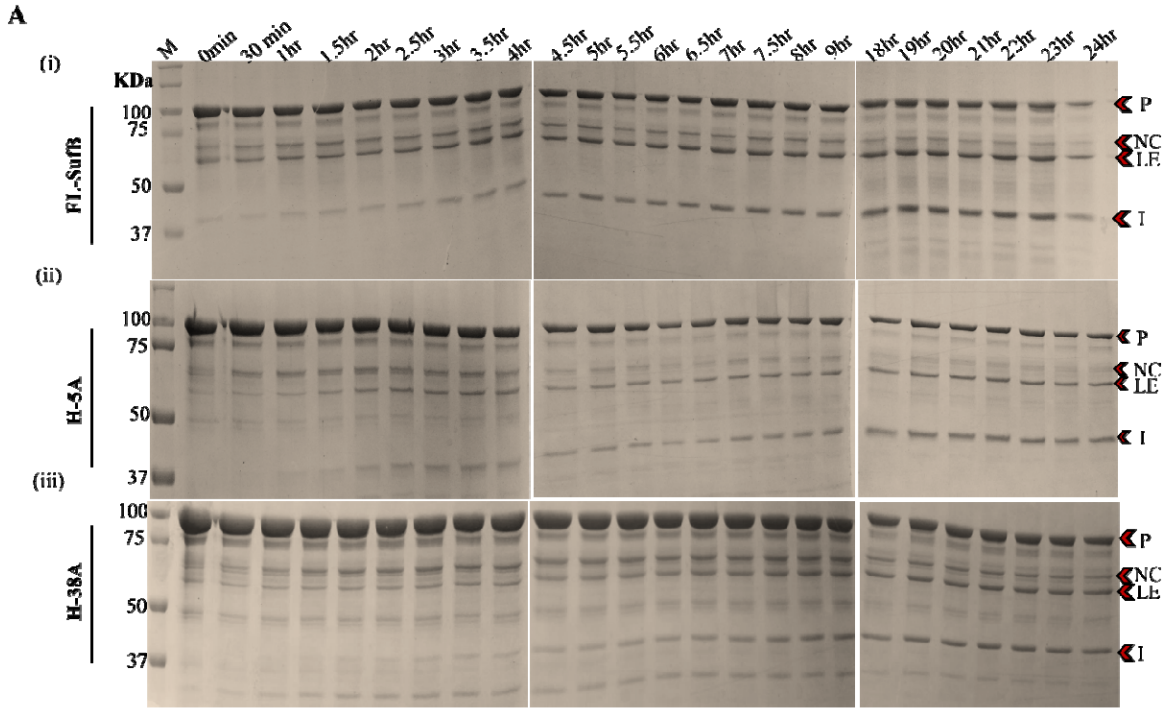


Fig. 2. Histidine conservation in the N-extein region of *Mtu* SufB protein (**A**) Multiple sequence alignment of SufB, focusing on N-extein~intein junction in mycobacterial and archeal species. Conserved histidine residues at -5 and -38 positions in the N-extein region are highlighted in yellow and green respectively. Catalytic C1 of inteins are marked in red. N-extein and intein residues are delineated by shading in grey and cyan blue, (**B**) Multiple sequence alignment of SufB, focusing on N-extein~intein junction in *Mtu* and other intein-less mycobacteria and in organisms where SUF is the exclusive system for Fe-S cluster assembly, (**C**) (i) & (ii) phylogenetic analyses of full-length (FL) SufB protein and SufB intein sequences in mycobacteria and *Ferroplasma acidarmans* respectively.


A
(i)
(ii)
(iii)
KDa
M
0min
30 min
1hr
1.5hr
2hr
2.5hr
3hr
3.5hr
4hr
4.5hr
5hr
5.5hr
6hr
6.5hr
7hr
7.5hr
8hr
9hr
18hr
19hr
20hr
21hr
22hr
23hr
24hr
100
75
50
37
FL-SufB
H-5A
H-38A
P
NC
LE
I

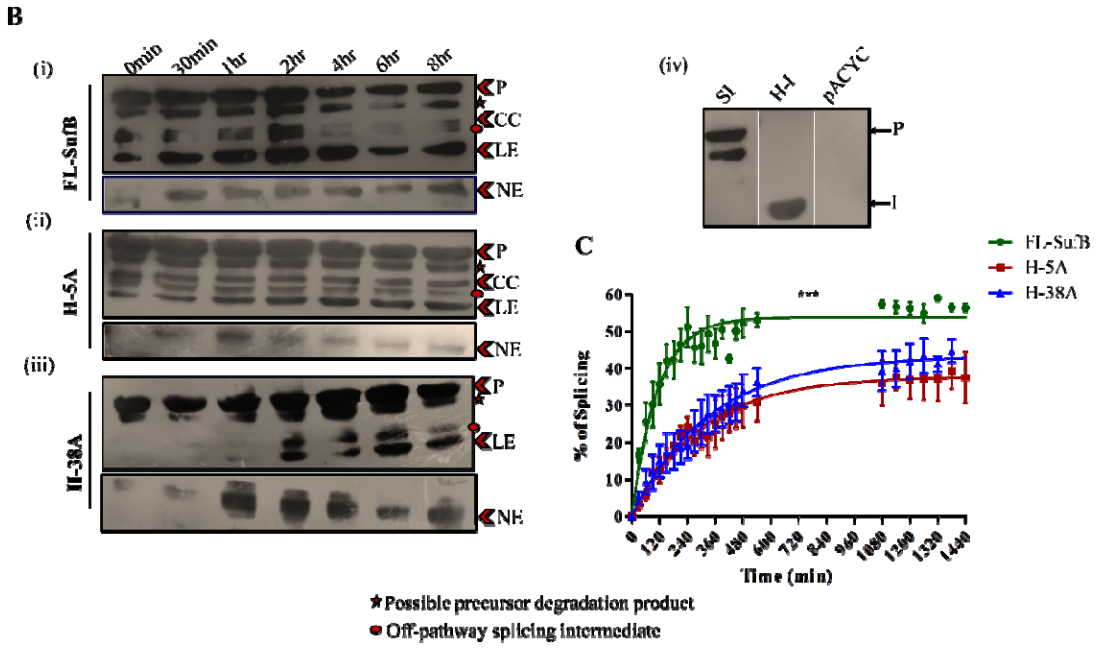

B
(i)
(ii)
(iii)
(iv)
0min
30min
1hr
2hr
4hr
6hr
8hr
FL-SufB
H-5A
H-38A
P
CC
LE
NE
St
H-I
pACYC
I
C
FL-SufB
H-5A
H-38A
% of Splicing
Time (min)
0
10
20
30
40
50
60
120
240
360
480
600
720
840
960
1080
1200
1320
1440
★Possible precursor degradation product
● Off-pathway splicing intermediate

Fig. 3. Effect of H-5 and H-38 mutation on *Mtu* SufB splicing. (**A**) Products from *in-vitro* refolding reactions were resolved through 4~10% gradient SDS PAGE as shown in, (i) FL-SufB precursor, (ii) H-5A and (iii) H-38A SufB mutant proteins exhibiting splicing over different time periods, (**B**) Western blot to confirm the identity of splicing and cleavage products for (i) FL-SufB precursor (ii) H-5A iii) H-38A SufB mutants and iv) controls; SI (splicing inactive double mutant SufB), H-I[6x(His)- tagged SufB intein], pACYC Duet-1(cell lysates expressing empty expression vector). Anti-His antibodies detected presence of 6x(His)-tagged P, CC, LE and NE. NE is blotted separately with higher concentration of primary antibody. (**C**) Kinetic study of protein splicing in *Mtu* FL-SufB precursor, H-5A and H-38A SufB mutant proteins. Splicing products were quantified and plotted over different time periods and the curve was fitted to a pseudo first order reaction; $Y=Y0 + (Plateau-Y0) *(1-exp^{(-K*x)})$. All the experiments were performed in triplicates and error bars represents (±1) SEM from 3 independent sets of experiments. A comparative analysis demonstrates statistically significant difference in splicing efficiency ($p<0.0001$) in FL-SufB precursor, H-5A and H-38A mutant SufB. P=Precursor, CC=C-cleavage, NC=N-cleavage, LE=Ligated Extein, I=Intein, NE=N-Extein, M= Protein marker.

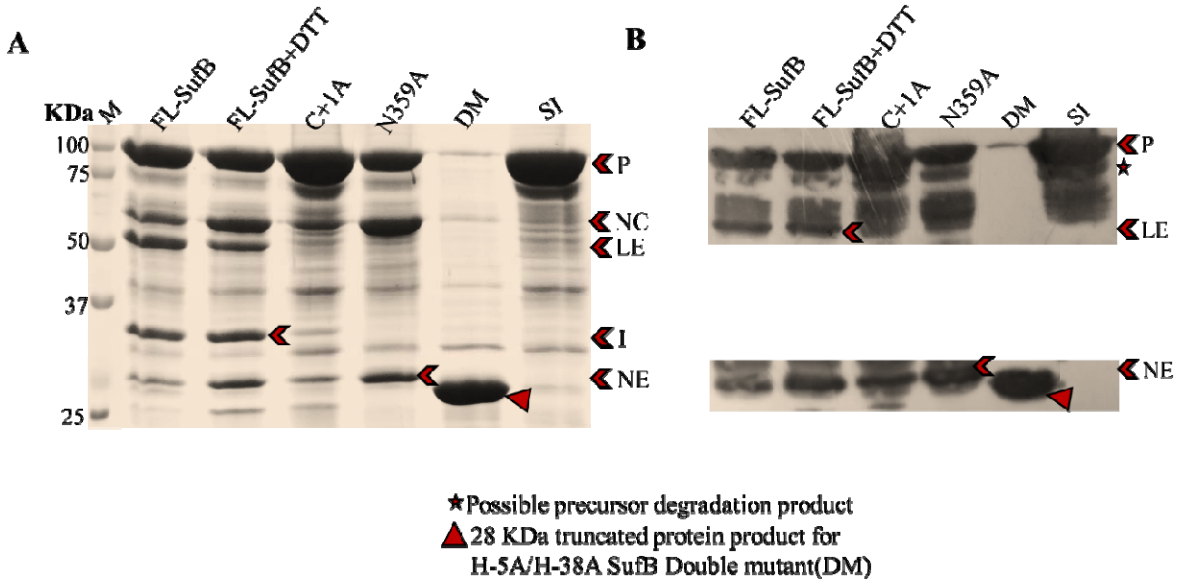


Fig. 4. Identification of *Mtu* H-5A/H-38A SufB double mutant(DM): (A) SDS Gel showing splicing and N-cleavage reactions in *Mtu* FL-SufB precursor, C+1A, N359A, DM (H-5A/H-38A) and SI (C1A/N359A) SufB mutant **(B)** Confirmation of protein products for FL-SufB, C+1A, N359A, DM (H-5A/H-38A) and SI (C1A/N359A) SufB mutant proteins via immuno- blotting assay. Anti-His antibodies detected presence of 6X(His) tagged P and LE. NE and SufB DM truncated protein product are blotted separately with higher concentration of primary antibody. P=Precursor, NC=N-cleavage, LE=Ligated Extein, I=Intein, NE=N-Extein.

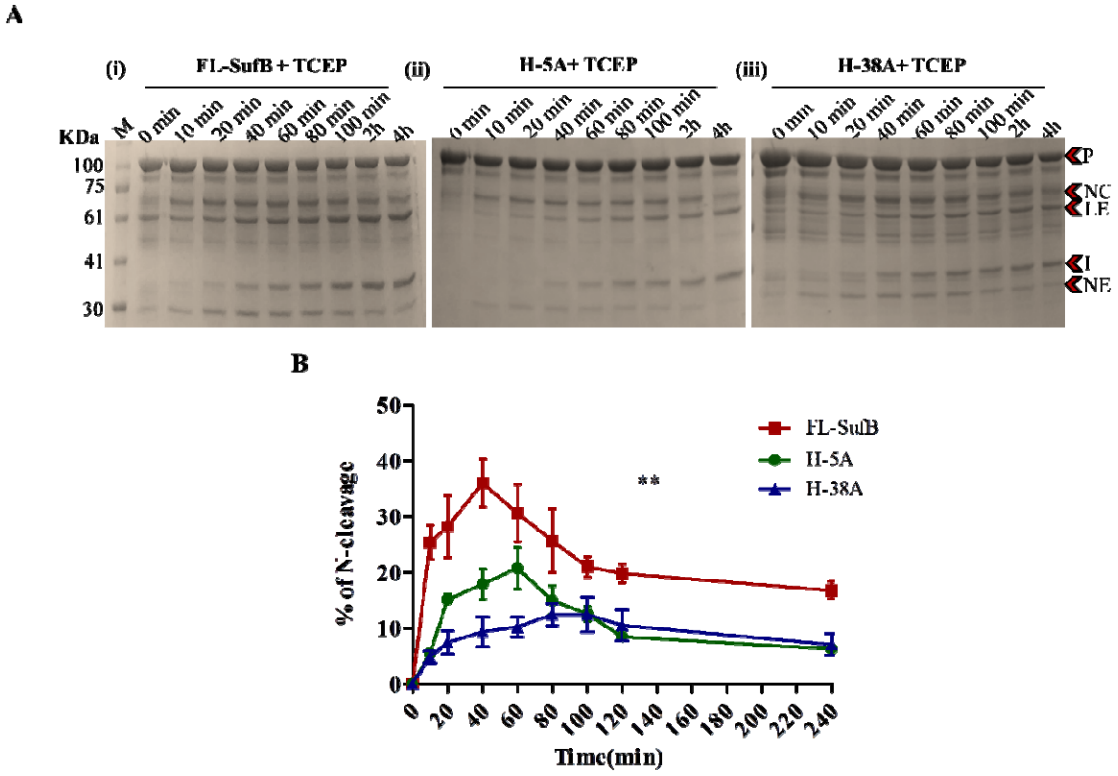


Fig. 5. Effect of H-5A and H-38A mutation on TCEP mediated N-cleavage reaction in *Mtu* SufB: (A) SDS Gel image of (i) FL-SufB precursor (ii) H-5A and (iii) H-38A SufB mutant proteins displaying N-cleavage reactions over different time periods; **(B)** (i) Comparative analysis of TCEP mediated reaction in FL-SufB, H-5A and H-38A SufB mutant proteins, indicates statistically significant differences (P=0.0011, one way ANOVA) in % of N cleavage over different time periods. Reaction mediated by TCEP did not give a proper curve which could be fitted in linear or non-linear regression. All the experiments were performed in triplicates and error bars represent (±1) SEM. P=Precursor, NC=N-cleavage, LE=Ligated Extein, I=Intein, NE=N-Extein, CC= C-Cleavage.

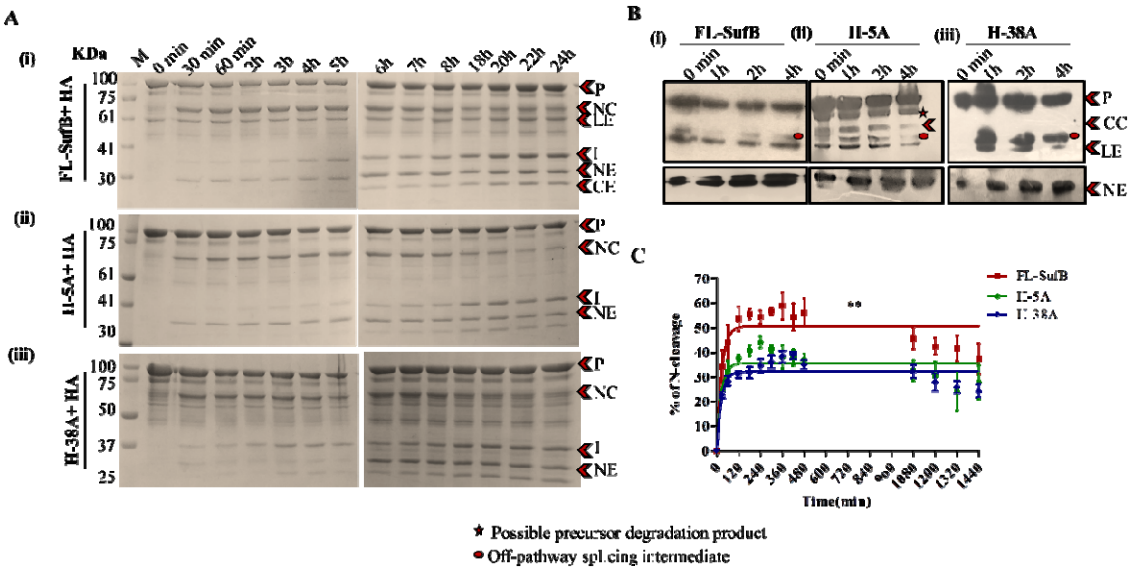
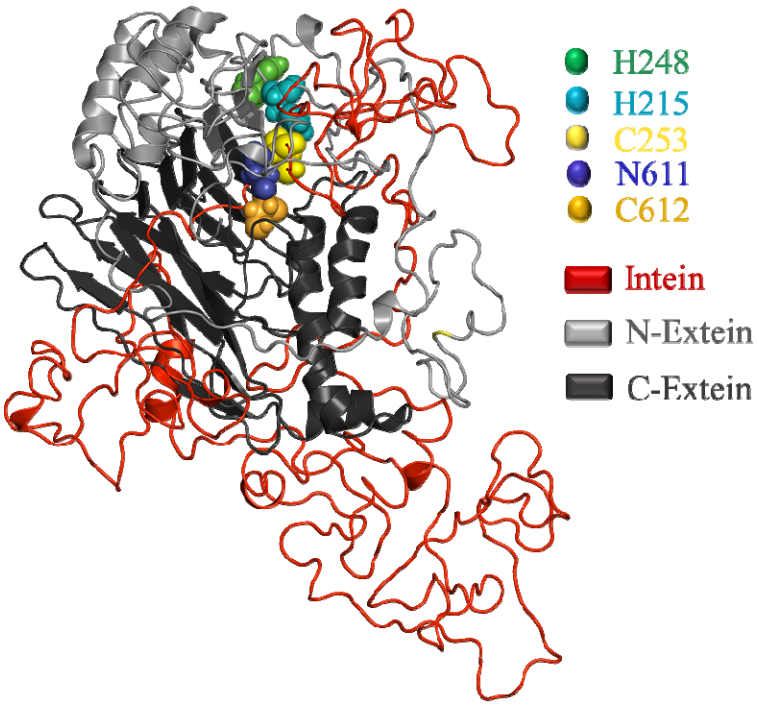
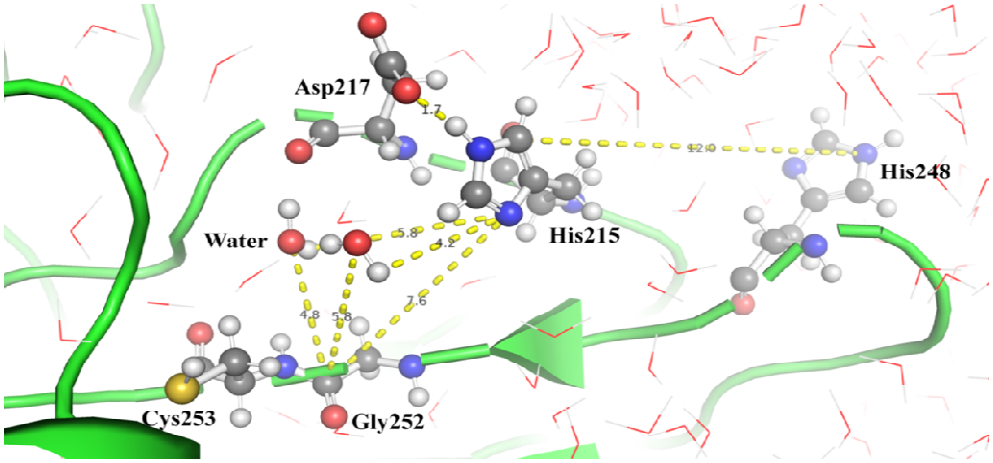


Fig. 6. Effect of H-5A and H-38A mutation on HA mediated N-cleavage reaction in *Mtu* SufB : (A) SDS Gel showing HA induced N-cleavage reactions in (i) FL- SufB precursor (ii) H-5A and (iii) H-38A SufB mutant; **(B)** Confirmation of splicing and cleavage products for HA induced N-cleavage reaction in (i) FL- SufB precursor, (ii) H-5A and (iii) H-38A mutant proteins via western blot. Anti-His antibodies detected presence of 6X (His) tagged P, CC and LE. NE is blotted separately with higher concentration of primary antibody; **(C)** (ii) Kinetic analysis of HA induced N-cleavage in WT and H-5A mutant proteins suggests statistically significant differences (P= 0.0025; one way ANOVA) in % of N-cleavage over different time periods. These curves were fitted into a pseudo first order reaction with equation Y=Y0 + (Plateau-Y0) *[1-exp$^{(-K*x)}$]. All the experiments were performed in triplicates and error bars represent (±1) SEM. P=Precursor, NC=N-cleavage, LE=Ligated Extein, I=Intein, NE=N-Extein, CC= C-Cleavage.

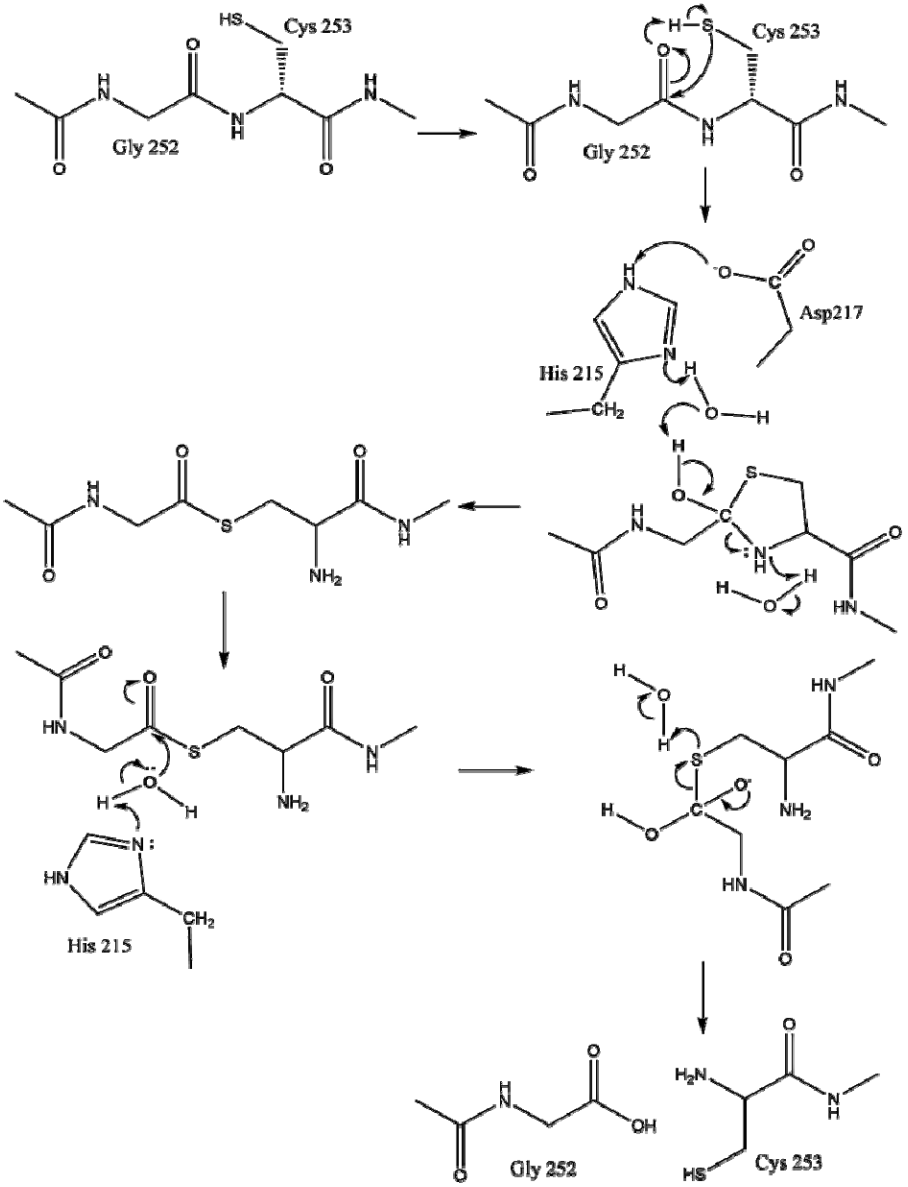
A


H248
H215
C253
N611
C612
Intein
N-Extein
C-Extein

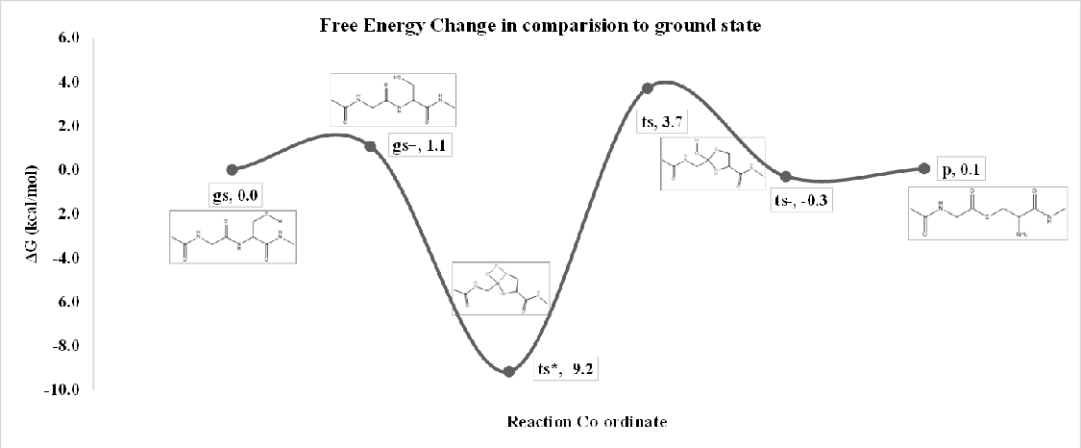
B


Asp217
His248
Water
His215
Cys253
Gly252

C


Cys 253
Gly 252
Cys 253
Gly 252
Asp217
His 215
His 215
Gly 252
Cys 253

D



E

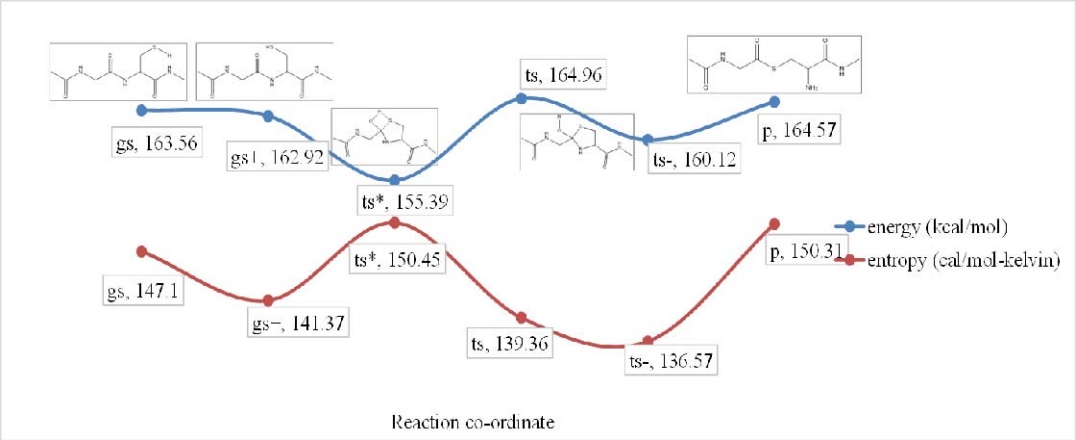
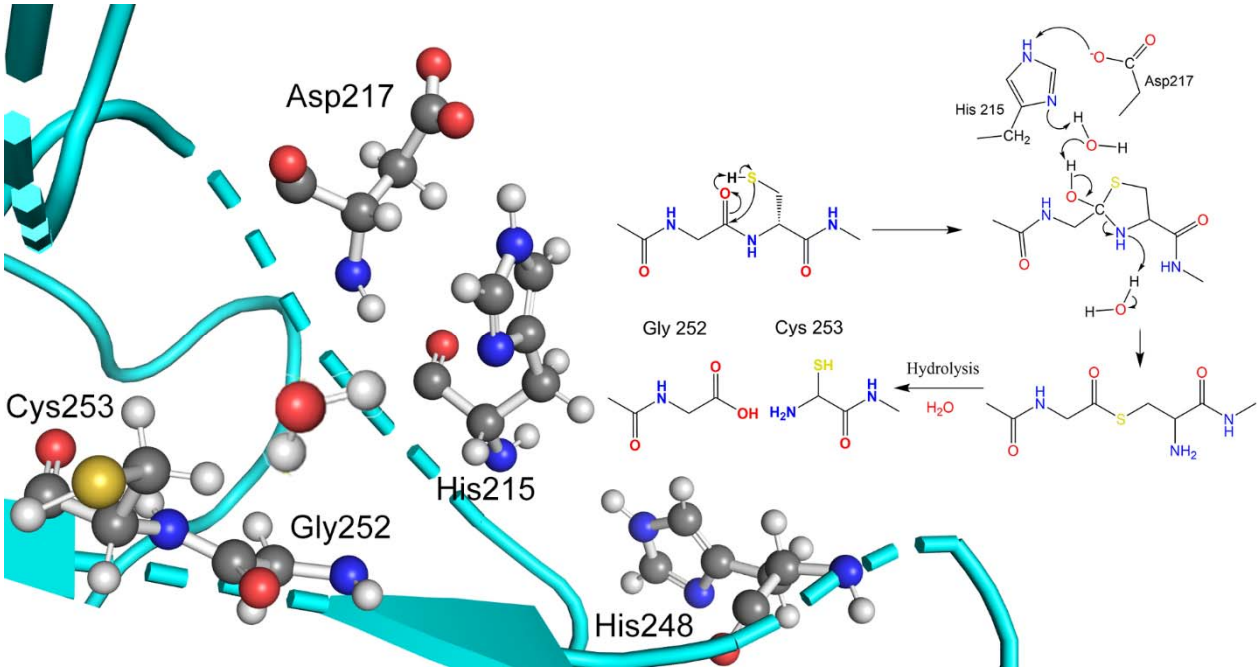


Fig. 7. Structural model and proposed mechanism for N-cleavage reaction in *Mtu* FL-SufB precursor. (**A**) The 3D model of *Mtu* FL-SufB precursor obtained in presence of water and NaCl. Different domains and conserved catalytic residues are color coded as shown in the legend, (**B**) Spatial arrangement of the N-terminal active site residues; Cys253 (C1) and Gly252 (G-1) along with His248 (H-5), His215 (H-38) and Asp217 (D-36) at the N-terminal cleavage site, (**C**) Proposed mechanism for N- terminal cleavage reaction at intein~exetin junction**,** (**D**) Free energy change, and (**E**) Energy and entropy of the system during QM calculation of different reaction states of N-cleavage junction from peptide bond to thio-ester bond of *Mtu* FL-SufB precursor protein. gs = ground state, gs+ = conformational transition state, ts*= transition state of proton transfer, ts = tetrahedral intermediate, ts = transition state of proton transfer, p = product.

Graphical Abstract