Research Article

SufB intein splicing in *Mycobacterium tuberculosis* is influenced by two remote conserved N-extein Histidines

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] Laboratory of Cellular and Molecular Basis of Diseases, Division of Transitional Medicine, Wadsworth Center, New York State Department of Health, Empire State Plaza, Albany New York 12237

[6] Department of Biomedical Sciences, School of Public Health, 1 University Place, Rensselaer, New York 12144

[†] 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

28 IPTG: Isopropyl β- d-1-thiogalactopyranoside

29 TCEP: Tris (2-carboxyethyl) phosphine

30 HA: Hydroxylamine

31 DTT: Dithiothreitol

32 HRP: Horseradish peroxidase

33 ECL: Enhanced chemiluminescence

34 TBST: Tris-buffered saline, 0.1% tween 20

35 BSA: Bovine serum albumin

36 ß-gal: Beta galactosidase

37 ABC: Ammonium bicarbonate

38 ACN: Acetonitrile

39 TFA: Tri-fluoro acetic acid

40 HPLC: High pressure (or performance) liquid chromatography

41 FPLC: First protein liquid chromatography

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

43 SDS: Sodium dodecyl sulphate

44 PAGE: Polyacrylamide gel electrophoresis

45 RMSD: Root mean square deviation

46 SASA: Solvent accessible surface area

47 RMSF: Root mean square fluctuation

48 PME: Particle-mesh Ewald

49 HPCC: High performance computing cluster

50 MD Simulations: Molecular dynamic simulations

51

52

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 [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- terminal cleavage reaction. His-38 is stabilized by the solvent-exposed His-5, and can impact N-S acyl shift by direct interaction with the catalytic Cys1. Development of inteins as biotechnological tools or as pathogen specific novel antimicrobial targets requires a more complete understanding of such unexpected roles of conserved extein residues in protein splicing.

Keywords: Intein splicing; Splicing regulation; conserved histidines; *Mycobacterium tuberculosis* SufB; SufB intein.

1. Introduction

Protein splicing is a self-catalyzed event that generates a continuous extein protein by ligating two intein separated extein regions with a peptide bond. This post-translational auto-excision of the intervening intein protein is critical for the formation of an active protein (1-4). Splicing and cleavage products (Figure 1C) are obtained through a series of nucleophilic displacement reactions mediated in a coordinated fashion by the catalytic residues (5-9). Control over this

reversible interruption of the functional form of the host exteins can play a regulatory role in protein activation (10, 11). The extein sequences upstream and downstream of N-terminal and C-terminal intein ends are termed 'N-extein' and 'C-extein' respectively. The intein folds into a horseshoe-shaped structure with a catalytic cleft that brings the conserved catalytic residues and the extein splice junction close enough for initiation of splicing reaction (12).

Inteins themselves have N-terminal and C-terminal regions with conserved sequence segments (termed as Blocks or motifs) that facilitates the splicing and cleavage reaction(s). N-terminal intein region comprises of A, N2, B, and N4 structural motifs or Blocks while F- and G-Blocks are part of the C-terminal intein region. Typically, A-Block contains Cys/Ser or Thr; B-Block includes His and Thr residues; F-Block usually has Asp and His and the 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 (Figure 1C), a nucleophilic attack by C1 or S1/T1 leads to an 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 esterification, the branched intermediate is resolved by terminal Asn cyclization through cleavage of C-terminal splice junction, and finally an (S-N/O-N) acyl shift fully ligates the two extein segments by an amide bond formation.

These bond rearrangements at splice junctions during the cleavage and splicing reaction(s) are assisted by non-catalytic intein residues through stabilization of various intermediate structure(s)(13). The Block A residues Cys, Ser, or Thr participate in the first step of splicing with significant assistance from the Block B His and Thr residues. 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).The Block F Asp residue drives the thioesterification through 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). The Block F Asp residue also deprotonates the C+1 residue to stabilizes the net positive charge on Cys1 that drives the transesterification reaction. The F- and G-Block His residues are critical in the coordination of terminal Asn cyclization. The F-Block His increases nucleophilicity of Asn, and the G-Block His accelerates the Asn cyclization by increasing the electrophilicity of backbone peptide (20-23). The final acyl-shift is energetically favorable and does not require assistance from either intein or extein residues(24). Inteins do show polymorphisms in the catalytic residues leading to variation in the splicing mechanism as seen in Class 2 and Class 3 intein splicing (11, 25-29).

Interactions between extein residues and the catalytic intein core in the regulation of splicing reaction have been studied by modulating intein activity, changing the conformation of catalytic cleft, and restraining the activity of catalytic residues (6, 30-34). Previous studies on intein-extein partnership in intein splicing have suggested mediation by extein residues both near and remote to the N- and C-terminal splice junctions (7, 30, 31, 33, 35-37). The 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 distortion and induce N-cleavage reaction (17). The participation of the first C-extein Cys+1 in the second and third steps of the splicing reaction has been demonstrated by its mutation dramatically augmenting or inhibiting splicing and generating off-pathway N-cleavage products

122 (12, 35, 38-40). Earlier, extein effects were assumed to be limited to residues proximal to the
123 intein (12, 31, 41), but recent work has shown that distal exteins are implicated as environmental
124 sensors with the role in regulating splicing depending on solution environment and temperature
125 in *Pho* RadA precursors (6, 10, 34).

126 SufB is a critical component of [Fe-S] cluster assembly and repair machinery called SUF
127 (mobilization of sulfur) complex. This is a stress response system that gets upregulated during
128 periods of oxidative stress and Fe starvation (42, 43). Though there are multiple pathways for
129 [Fe-S] cluster biogenesis among the three kingdoms of life such as NIF (nitrogen fixation) and
130 ISC (Iron-sulfur cluster), the SUF complex is unique in mycobacteria. Fe-S cluster containing
131 proteins execute a broad spectrum of cellular functions in organisms such as respiration, gene
132 regulation, RNA modification, DNA repair, and replication (44-46). Although the SUF system
133 has been well characterized in the *E. coli* system, (39, 42, 43, 47-49), the importance of SufB
134 intein splicing in the formation of functional Suf complex has been shown in mycobacteria (49).

135 The present study reports on the splicing and cleavage reactions of a full-length *Mtu* SufB (FL-
136 SufB) precursor protein. We delineated the different structural domains of *Mtu* SufB, analyzed
137 whether it follows a canonical or non-canonical intein splicing pathway, identified intein and
138 extein residues that participate in catalytic cleft formation, assessed both their conservation in
139 different mycobacterial species and their role in regulating cleavage and/or splicing reactions, and
140 analyzed distinctions from other intein precursor proteins. We found intein residues highly
141 conserved in different mycobacterial species that favor a canonical splicing mechanism (Figure
142 1A and 1B). We detected two distal histidines in the N-extein region, His-5 and His-38, that are
143 conserved in mycobacteria, archaea, and other microbes where SUF is the exclusive system for
144 Fe-S cluster biogenesis (Figure 2A and 2B). Biochemical analyses of H5A and H38A SufB

145 mutants confirmed their influence on splicing and cleavage. Structural modeling of *Mtu* SufB and
146 explicit-solvent molecular dynamics (MD) simulations of the model were used to analyze the SufB
147 precursor splicing active site dynamics, and these simulations suggested that N-terminal cleavage
148 could be supported by an interaction between H- 38 and H-5. These observations suggest that the
149 two distal H-5 and H-38 N-extein residues participate in SufB precursor stabilization and aid *Mtu*
150 SufB intein splicing.

151 Kinetic analyses of H-5A SufB mutant demonstrated 3-fold and 1.4-fold reductions in splicing
152 and cleavage rates respectively, relative to wild-type *Mtu* SufB precursor. Likewise, a 3.4-fold
153 and 3-fold reduction in splicing and cleavage rates were observed in the H-38A SufB mutant. A
154 side-by-side *Mtu* SufB structure prediction was done using homology (chimera) modeling,
155 secondary structure prediction through consensus with other protein sequences. Subsequently,
156 molecular dynamics (MD) simulations in aqueous media were carried out to find the equilibrated
157 structure. Furthermore, MD simulations clarified the structural features of the SufB intein active
158 site and indicated that N-terminal cleavage reaction is catalyzed by H-38 with the assistance of
159 H-5. Taken together, our study substantiates a distinct mechanism for N-terminal cleavage
160 reaction shown by *Mtu* SufB. Although H-38 is relatively distal to the N-terminal splice junction
161 when supported by His-5, it can efficiently activate the first step of splicing reaction. Finally, we
162 have proposed a novel mechanism for the N-cleavage reaction mediated via the concerted actions
163 of these conserved histidines in the N-extein region of *Mtu* FL-SufB precursor. Interestingly,
164 attempts to express the SufB double mutant (H-5A/H-38A) protein resulted in a truncated protein.
165 These observations suggest that H-5 and H-38 might have important biological role(s) in the SufB
166 precursor stabilization and perhaps the functionality of *Mtu* SufB protein.

167

168 **2. MATERIALS AND METHODS**

169 **2.1. Genetic constructs**

170 The full-length Fe-S cluster assembly protein SufB from the *Mycobacterium tuberculosis* H37Rv
171 strain (*Mtu*-SufB-FL) and its isolated intein (*Mtu*-SufB-I) genes were PCR-amplified using Pfu
172 Ultra High-Fidelity DNA Polymerase (Agilent Technologies) from heat-killed *Mtu* genomic DNA.
173 DNA purification by gel electrophoresis was followed by EcoRI and HindIII restriction
174 digestion and cohesive end ligation (T4 DNA Ligase, NEB Cat. No. M0202S) for cloning. The
175 genes were inserted into the multiple cloning site 1 (MCS1) of the low copy expression vector
176 pACYCDuetTM-1(Novagen), which was driven by a T7 promoter/lac operator with a
177 chloramphenicol resistance gene for selection. The constructs were screened via colony PCR and
178 confirmed by sequencing (Sequencing Core Facility, SUNY, Albany, and Eaton Bioscience Inc.
179 sequencing service) using the ACYCDuetUP1 (Novagen Cat. No.71178-3) and DuetDOWN1
180 primers (Novagen Cat. No. 71179-3), as well as the original primers. *Mtu*-SufB-FL mutants H5A,
181 H38A, C1A, N359A were generated by substituting respective key catalytic residues to alanine
182 via phosphorylated inverse PCR primers. Splicing inactive (SI) *Mtu*-SufB-FL double mutant
183 (C1A/N359A) was created via inverse PCR to add the N359A mutation into the C1A cleavage
184 mutant. Similarly, H5A/H38A *Mtu*-SufB-FL double mutant (DM) was created by adding H38A
185 mutation into the H5A mutant by inverse PCR. The cloned H38A and H5A/H38A *Mtu* SufB
186 mutant genes were also confirmed separately by sequencing (Agrogenomics, Odisha) using
187 ACYCDuetUP1 (IDT Cat. No. 103948189) and DuetDOWN1 (IDT Cat. No. 103948190) along
188 with the SufB Primers. All the above primers are listed in Table S1.

189

190 **2.2. Sequence analysis**

191 Protein sequences for *Mtu* SufB-FL (Accession number YP_006514844.1, GI: 397673309) and
192 the 477 amino acid intein-less SufB protein from *Mycobacterium smegmatis* strain MC2 155
193 (*Msm*-SufB-FL, accession number YP_887437.1, GI: 118472504) were pairwise sequence aligned
194 in ClustalW to distinguish extein and intein regions in *Mtu* SufB-FL (50). The identified *Mtu* SufB
195 intein sequence (*Mtu* SufB-I) was confirmed by sequence comparison using Blast with the
196 sequences deposited in Inbase (The Intein Database, www.inteins.com) (51). Different structural
197 domains of the intein-like homing endonuclease, and the N- and C-terminal inteins were clearly
198 demarcated by sequence alignment and structural analysis of SufB intein with homing
199 endonuclease domain (I-CreI) and intein homing endonuclease Ii (PDB: 2CW7). The *Mtu* SufB-
200 FL and *Msm*-SufB-FL sequences, combined with one archaeal and other bacterial SufB proteins
201 collected from the NCBI and intein databases (51), were aligned using Dialign2 software (52).
202 Conservation of different intein and extein residues was edited and color-coded manually.
203 Phylogenetic tree analysis was performed using the Maximum likelihood method in the MEGA X
204 program (53, 54) for both SufB inteins [Figure 2C (ii)] and SufB precursor sequences [Figure 2C
205 (i)] from different organisms.

206 **2.3. Protein overexpression and purification**

207 Full-length (FL) un-spliced precursor and mutant SufB proteins carrying an N-terminal 6X-His-
208 tag were over-expressed in BL21 (DE3) *E.coli* cells via IPTG (Sigma 367-93-1) induction. Cells
209 were resuspended in lysis buffer (20mM sodium phosphate, 0.5M NaCl, pH 7.4) and lysed via tip
210 sonicator (Sonics vibra cell VCX-130). Proteins were over-expressed and isolated from inclusion
211 bodies (IB) via centrifugation. The IB materials were solubilized by 8M urea (Merck, 1084870500)

212 buffer (lysis buffer, 8M urea, 20 mM of imidazole (MP–biochemicals-288-32-4) and centrifuged
213 at 16,500g for 20 min to collect the supernatant. Then 6X-His-tagged full-length (FL) precursor
214 and mutant proteins were purified by Ni-NTA affinity column (Ni-NTA His trap, HP GE
215 Healthcare Life Sciences-17524802). Prior to sample application, columns were equilibrated with
216 binding buffer (20mM sodium phosphate, 0.5M NaCl, 40mM imidazole). After sample loading,
217 columns were washed several times (15 CV) in binding buffer. Finally, proteins were eluted as
218 purified fractions in elution buffer (20mM sodium phosphate, 0.5M NaCl, 500mM imidazole)
219 followed by quantification via Bradford's assay.

220 **2.4. *In-vitro* splicing and cleavage assays**

221 2.5 µM of purified proteins were allowed to refold in 1 ml of renaturation buffer (20 mM sodium
222 phosphate, 0.5 M NaCl, 0.5 M Arginine,1 mM EDTA, pH 7.4) in presence of 2mM TCEP-HCl
223 (sigma-51805-45-9) at 20˚C for 24 h. The 0hr sample was retrieved before renaturation and
224 splicing was quenched by addition of loading dye (0.1% bromophenol blue, 50% glycerol, ß-
225 mercaptoethanol, 10% SDS, tris 6.8) followed by rapid freezing at -200 C. Our controls, splicing
226 inactive(SI) SufB double mutant and empty expression vector pACYC Duet-1, were treated
227 similarly for the *in-vitro* assays. For the N-cleavage assay, proteins were refolded in presence of
228 reducing agents and nucleophiles such as 2mM TCEP-HCl, 50 mM DTT (Roche-10708984001),
229 and 0.5 M 250 Hydroxylamine (SRL-66164) with 1mM TCEP in renaturation buffer (55). For
230 splicing and cleavage analysis, sample extraction at each time interval was followed by the
231 addition of loading dye to stop the reactions and then boiling at 95˚C for 5 min. Resultant products
232 from various refolding reactions were resolved through 4~10% gradient SDS PAGE. Protein bands
233 were stained with Coomassie blue R-250 and densitometric analysis was performed by using
234 GelQuant.Net biochemical solutions. Percentage(s) of splicing and cleavage products were

235 measured by taking the percentage(s) of the ratio of the total splicing product (LE and I) over total
236 proteins (LE+I+P) and total N-cleavage product (NE+NC) over total proteins (NE+NC+P). The
237 0hr splicing value(s) were subtracted at each time point for baseline correction.

238 **2.5. Kinetic analyses**

239 Since *Mtu* 6X (His)-tagged WT SufB and SufB mutants (H-5A, H-38A, H-5A/H-38A, and 263
240 C1A/N359A) were purified and renatured at different temperatures, after normalizing splicing and
241 cleavage values at different time intervals, the plot was generated by taking the percentage of
242 splicing or cleavage product with respect to time (in min). Next, the curve was fitted in pseudo-
243 first-order kinetics, with an equation Y=Y0 + (Plateau-Y0) *(1-exp(-K*x)) [Where X=time, Y0
244 =Y value when time (X)= time0, Plateau= max Y value at time t, K = rate constant, expressed in
245 reciprocal of the X-axis (time units)] in graph pad prism software. The fitted curve was generated
246 by automatic outlier elimination fitting in a nonlinear regression equation. The rate constant (K)
247 and Vmax were generated by the software. Half-life (t1/2) was calculated by graph pad prism using
248 the formula (Ln2/K).

249 **2.6. Western blot**

250 Western blot analysis was performed using an anti-His antibody (Invitrogen, LOT 1902132) to
251 confirm the identity of splicing and cleavage products. Following resolution through SDS PAGE,
252 test proteins were transferred to a nitrocellulose membrane; at 50v, 2h. After a successful transfer,
253 blocking was done with 5% skim milk for 2h at room temperature. Then the blot was incubated
254 with HRP conjugated anti-His antibody (Invitrogen, LOT 1902132) at 1:5000 dilutions for 16h at
255 40 C. Then blot was washed with 1X TBST and developed using 13 ECL as the substrate. N-extein
256 detection was done with 1:2500 antibody dilution.

257 **2.7 Mass spectrometry and chromatography**

258 After renaturation, proteins were resolved through 4~10% SDS PAGE. Protein identification by
259 mass spectrometry was performed at Central Proteomics Facility, Institute of Life Sciences,
260 Bhubaneswar using the following standardized protocol. Well resolved protein gel bands were
261 destained with 25mM ammonium bicarbonate (ABC). Cleared gel bands were reduced with 20
262 mM DTT and then alkylated by 100 mM iodoacetamide solution. The gel plugs were dehydrated
263 sequentially with 50%, 100% acetonitrile (ACN) and then digested with 0.3 µg of trypsin (Absciex
264 4352157) in 25 mM ABC solution at 37˚C for 16 hours. Peptides were extracted sequentially using
265 extraction buffer containing 0.1%, 0.5%, and 5% tri-fluoro acetic acid in 50% ACN from gel
266 bands. Extracted peptides were desalted by using C18 Zip-Tip (Millipore) and then dried in a
267 centrifugal vacuum concentrator. Peptide samples were reconstituted with 0.1% TFA in 50% ACN
268 solution mixed 1:1 ratio with α-cyano-4-hydroxycinnamic acid solution (C8982; 10 mg/ml) and
269 then spotted onto metal target plate in triplicates. MS and MS/MS spectra were acquired using
270 MALDI TOF/TOF (AbSciex TOF/TOF 5800). Acquired spectra were searched against
271 *Mycobacterium tuberculosis* from NCBInr database using mascot search engine in Protein Pilot
272 Software for protein identification. Searches were performed allowing trypsin mis-cleavage up to
273 Carbamido-methylation of cysteine, oxidation of methionine was also included as variable and
274 fixed modification respectively. The peptide mass tolerance was 300 set as 100 ppm for precursor
275 ion and 0.8 Da for fragment ion with +1 charge. Since all the splicing and cleavage products were
276 identified as *Mtu* FL-SufB, the acquired spectra were researched against a customized database
277 containing probable splicing products and cleavage products of SufB protein of *Mycobacterium*
278 *tuberculosis* using mascot search engine in Protein Pilot Software with the same parameter. β-gal
279 and BSA were run as internal calibration. The protein score, percent coverage, theoretical
280 molecular weight, and Iso-electric pH value were obtained. The mass spectrometry proteomics

281 data have been deposited to the ProteomeXchange Consortium via the PRIDE [1] partner
282 repository with the dataset identifier PXD015199.

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

295 Molecular weight of the SufB DM (H5A/H38A) was determined by FPLC. H215A/H248A SufB
296 mutant protein (DM) was isolated and purified under denaturing conditions. The relative molecular
297 weight of DM was confirmed by Gel filtration chromatography using Hi prep 16/60 Sephacryl
298 S200 GPC column (HiPrep 16/60 GE Healthcare, Cat. no. 10255928). 8M Urea buffer was
299 removed from the protein sample by dialysis in 20mM sodium phosphate and 0.15M sodium
300 chloride buffer with pH 7.4. Then the sample was concentrated using Protein concentrator (Thermo
301 Scientific 88515). For FPLC analysis, the column was equilibrated with 2 column volume (cv) of
302 sodium phosphate buffer (pH 7.4) followed by loading of molecular weight standard Bovine
303 Serum Albumin (BSA) at a concentration of 1mg/ml. Then 2mg of purified DM protein was loaded

304 onto the GPC column. The run was performed at a flow rate of 1ml/min while keeping the
305 temperature at 20° C. The fractions eluted at an absorbance of 280nm were collected and analyzed
306 by SDS-PAGE under reducing and non-reducing conditions.

307 **2.8 Molecular Dynamics (MD) Simulations**

308 ***2.8.1. Homology modeling of Mtu SufB precursor***

309 A homology model of the *Mtu* FL-SufB precursor was built using the SufB chain of the *E. coli*
310 SufB-SufC-SufD complex (PDB ID: 5AWF, chain A), the *S. cerevisiae* intein homing
311 endonuclease (PDB ID: 1VDE, chain A), and the *T. kodakarensis* homing endonuclease (PDB ID:
312 2CW7, chain A) using Bioluminate 2.7 from Schrödinger. The model was assessed by
313 Ramachandran plot analysis and PROSA(Z-score) (56).

314 ***2.8.2. MD simulation analysis***

315 Explicit-solvent MD simulations were performed using GROMACS (V5.1.4) (57) with the OPLS-
316 AA force field (58) and the SPC/E water model (59, 60). The *Mtu* SufB model (846 aa) was
317 solvated in a cubic box with at least 1 nm distance between the protein and the edge of the box,
318 and neutralized with NaCl. The system was minimized with target Fmax of no greater than 1000
319 kJ mol-1 349 nm-1 with steepest descent minimization with a spherical cut-off at 1 nm was
320 imposed on all intermolecular interactions with verlet cut-off scheme (61, 62). The leap-frog
321 algorithm with a timestep of 2fs (63) and a canonical NVT ensemble was used to run the simulation
322 for 100 ns, with temperature maintained at 300K through velocity-rescale coupling and no
323 temperature coupling. H-bonds were constrained using lincs with the order of 4 (64). The particle-
324 mesh Ewald (PME) algorithm was used for implementing long-range electrostatic interactions
325 with the grid dimension of 0.16nm and interpolation order of 4 (65). Histidine at positions 5 and

326 36 were mutated to alanine by homology modelling, and MD simulations were performed on these
327 mutant proteins in an identical fashion. The MD simulation trajectories were analyzed with
328 GROMACS and Pymol, and plotted using Origin 8.0.

329 **3. RESULTS**

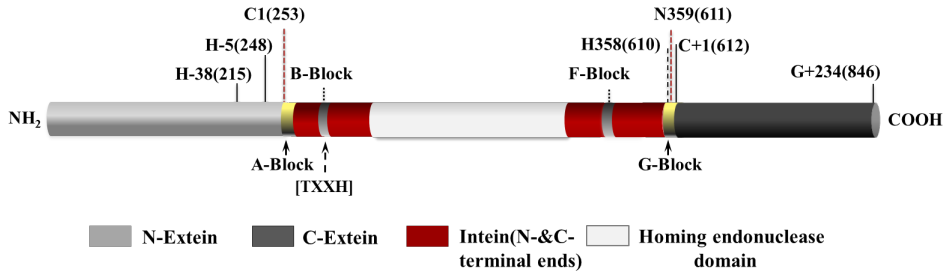
330 **3.1 Structural domains of *Mtu* full-length (FL) SufB precursor**

331 We obtained the protein sequences for Fe-S cluster assembly protein SufB [*Mycobacterium*
332 *tuberculosis* H37Rv], an 846 amino acid protein (Accession number WP_003407484.1, GI:
333 397673309) and the intein-less SufB protein [*Mycobacterium smegmatis* MC2 155], a 477 amino
334 acid protein (Accession number YP_887437.1, GI: 118472504) from NCBI protein database. A
335 pair-wise sequence alignment using the ClustalW server very clearly delineated the intein
336 sequence as well as identified ~95% sequence similarities between the extein sequences. A blast
337 search against the sequences deposited in Inbase also identified the intein sequence as 359 amino
338 acid *Mtu* SufB (*Mtu Pps1*) intein From *M. tuberculosis* strains H37Rv. Based on sequence
339 alignment and structural analyses of SufB intein and homing endonuclease domain I-CreI and
340 intein homing endonuclease Ii (PDB: 2CW7), the final demarcation of different structural
341 domains of *M. tuberculosis* and their respective sequences appeared to be as shown in Table S2.

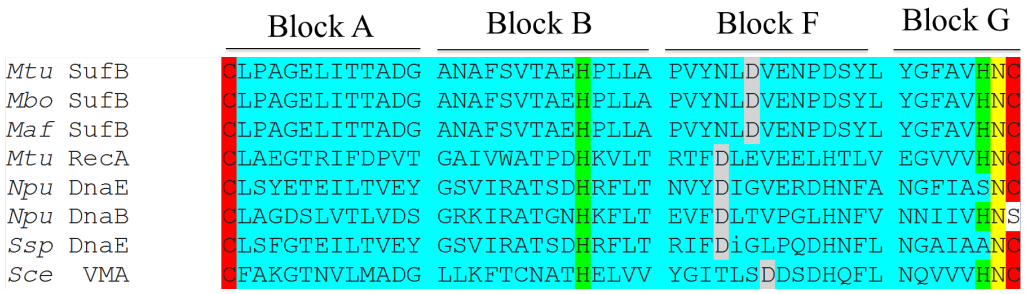
342 Residues encompassing N- & C-terminal extein and intein sequences from *Mtu* SufB precursor
343 are illustrated in Figure 1A. Multiple sequence alignment of *Mtu* SufB precursor and other intein
344 bearing proteins highlighted conservation of catalytic cysteines (C1 and C+1), Block B His67,
345 Block F Asp77, penultimate His358, and terminal Asn359 in the intein and extein regions (Figure
346 1B). Hence, SufB precursor protein (846 aa) is likely to follow a canonical intein splicing
347 pathway to generate the following products (Figure 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).

A



B



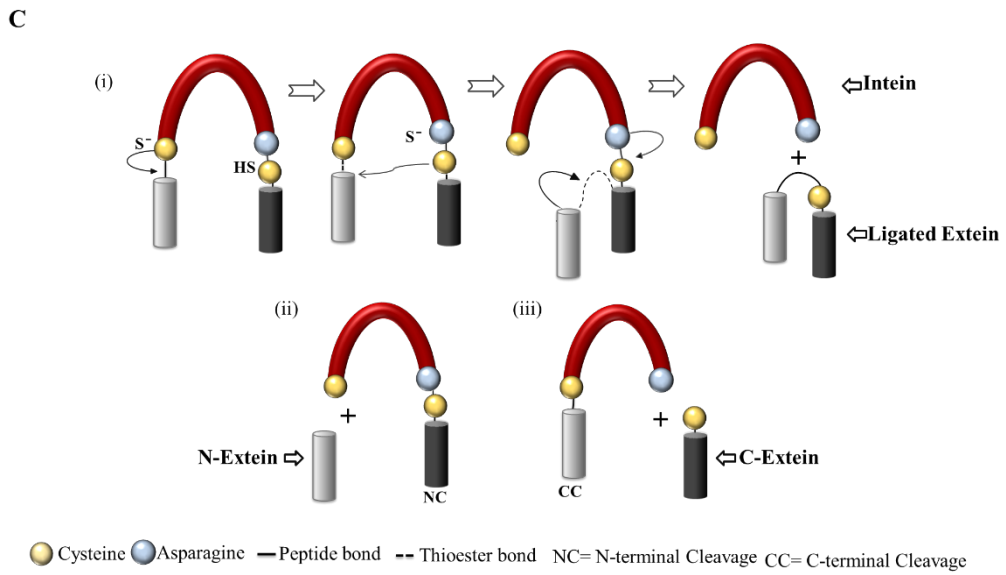


Figure 1. (A) Schematic representation of different structural domains of *Mtu* full-length SufB precursor. The conserved and the catalytic residues in the FL-SufB protein sequence are shown at the top. Further, numbering for different structural domains are also displayed; intein residues (1, 2……,n), 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.

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 (Figure 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 [Figure 2C (ii)] and precursor sequence [Figure 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.

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)   EKPTYDIQVV GLENFVANGI VAHNSFIYVP PGVHVDIPLQ AYFRINTENM (613)
M.lepromatosis (563)   EKPTYDIQVV GLENFVANGI VAHNSFIYVP PGVHVDIPLQ AYFRINTENM (612)
Ferroplasma    (196)   ---------- ---------- ----SFLYVP KNVTIDMPLQ TYFRMNGEQS (221)

M.tuberculosis (232)   GQFERTLIIA DEGSYVHYVE GCLPAGELIT TADGDLRPIE SIRVGDFVTG (281)
M.microti      (232)   GQFERTLIIA DEGSYVHYVE GCLPAGELIT TADGDLRPIE SIRVGDFVTG (281)
M.africanum    (232)   GQFERTLIIA DEGSYVHYVE GCLPAGELIT TADGDLRPIE SIRVGDFVTG (281)
M.bovis        (232)   GQFERTLIIA DEGSYVHYVE GCLPAGELIT TADGDLRPIE SIRVGDFVTG (281)
M.Bourgelatii  (225)   GQFERTLIIA DEGSYVHYVE GCLPAGELIT TADGDLRPIE SIQVGDYVTG (274)
M.Kansasii     (225)   GQFERTLIIA DEGSYVHYVE GCVPAGELIT TADGDLRPIE SIRVGDYVSS (274)
M.orygis       (232)   GQFERTLIIA DEGSYVHYVE GCLPAGELIT TADGDLRPIE SIRVGDFVTG (281)
M.gastri_W     (225)   GQFERTLIIA DEGSYVHYVE GCLPAGELIT TADGDLRPIE SIRVGDYVTG (274)
M.gastri_p     (479)   GQFERTLIIV GENAYVTYVE G--------- ---------- ---------- (499)
M.canettii     (237)   GQFERTLIIA DEGSYVHYVE GCLPAGELIT TADGDLRPIE SIRVGDFVTG (286)
M.caprae       (232)   GQFERTLIIA DEGSYVHYVE GCLPAGELIT TADGDLRPIE SIRVGDFVTG (281)
M.Pinnipeddi   (232)   GQFERTLIIA DEGSYVHYVE GCLPAGELIT 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)   GQFEHTIVVA DEGAKVHYIE GCLPEDELIS QGDkfv-PIN ELLTDDSVVS (270)
```

B

```
M.tuberculosis     (181) EYFGTVIPAGDNKFSALNTAVWSGGSFIYVPPGVHVDIPLQAYFRINTENMGQFERTLII(240)
M.marinum          (174) QYFGTVIPAGDNKFSALNTAVWSGGSFIYVPPGVHVDIPLQAYFRINTENMGQFERTLII(245)
M.fortuitum        (181) KYFGTVIPAGDNKFSALNTAVWSGGSFIYVPPGVHVDIPLQAYFRINTENMGQFERTLII(251)
M.smegmatis        (171) QYFGTVIPAGDNKFSALNTAVWSGGSFIYVPPGVHVDIPLQAYFRINTENMGQFERTLII(241)
M.shigaense        (175) QYFGSVIPAGDNKFSALNTAVWSGGSFIYVPPGVHVDIPLQAYFRINTENMGQFERTLII(245)
S.aureus           (159) KYFASVVPAADNKFAALNSAVWSGGSFIYVPKNIKLDTPLQAYFRINSENMGQFERTLII(229)
B.subtilis         (159) EHWAKVIPPTDNKFAALNSAVWSGGSFIYVPKGVKVETPLQAYFRINSENMGQFERTLII(229)

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

C

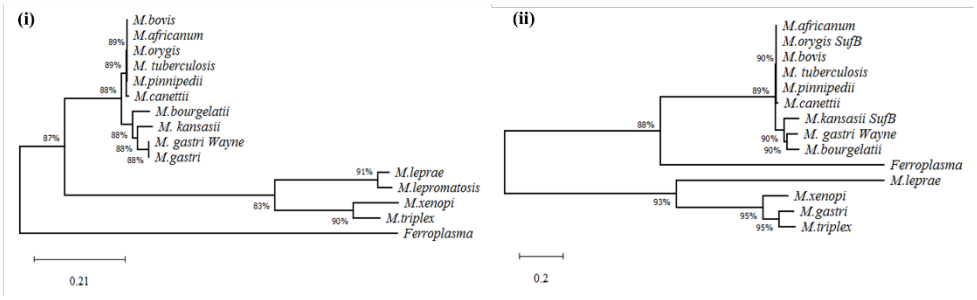


Figure 2. Histidine conservation in the N-extein region of *Mtu* SufB protein (**A**) Multiple sequence alignment of SufB, focusing on N-extein~intein junction in mycobacterial and archeal species. Conserved histidine residues at -5 and -38 positions in the N-extein region are highlighted in yellow and 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.

Further, two conserved histidines were identified in the N-extein region of different mycobacterial and archaeal SufB precursor proteins, His-5 and His-38 (Figure 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 (Figure 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 (Figure 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

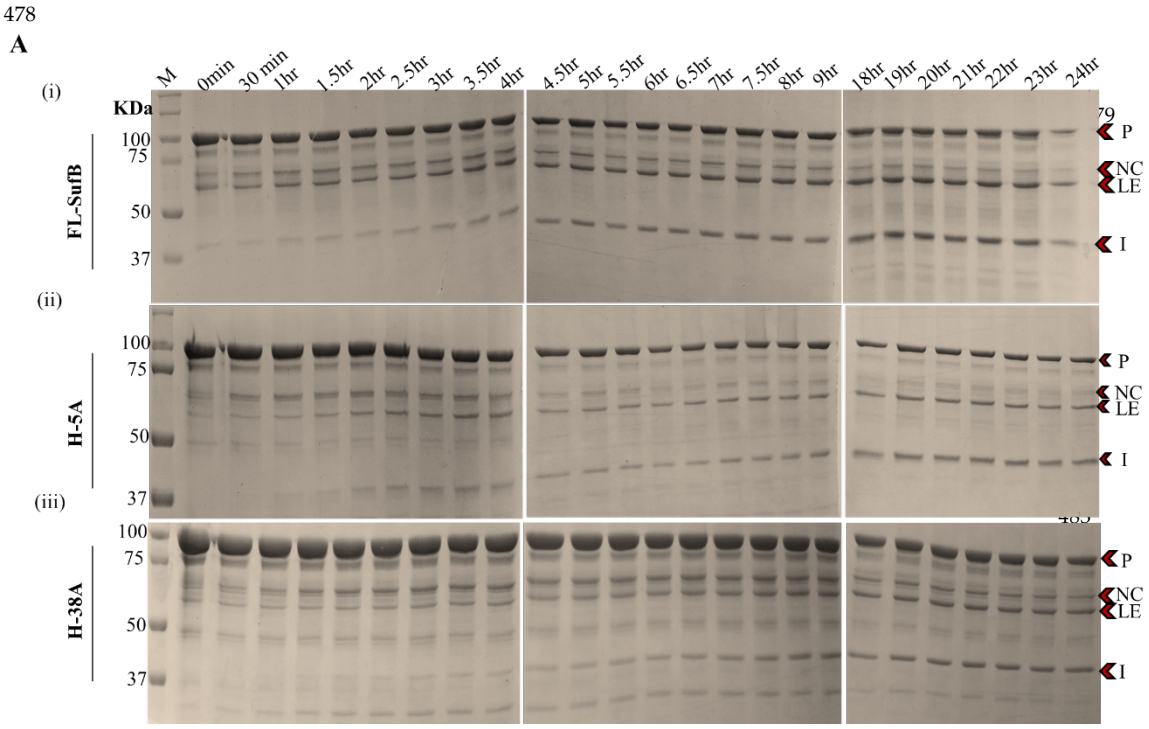
448 their regulatory roles on cleavage and/or splicing as well as in the functionality of [Fe-S] cluster
449 assembly protein SufB. Despite their remote positions these two histidines possibly come closer
450 to the active site and participate in protein splicing either via direct or indirect interaction with
451 the catalytic residues near cleavage site(s). A comprehensive kinetic study of *in-vitro* splicing,
452 and N-cleavage reactions are presented in the next section. Further mechanistic understanding of
453 the roles of highly conserved histidines was explored via molecular dynamics (MD) simulations
454 as shown later in this manuscript.

455 **3.3 Kinetic study to evaluate the roles of conserved His-5 and His-38 on *Mtu* SufB intein**
456 **splicing**

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

467 The above proteins were purified under denaturing conditions and refolded at 20°C as described
468 in the materials and methods. FL-SufB precursor, H-5A and H-38A SufB mutants gave rise to
469 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 (Figure 3). As predicted none of these products were seen in the case of SI double mutant [Figure S2E (i)] and transformants expressing empty vector pACYC Duet-1 [Figure 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 (Figure 3B).



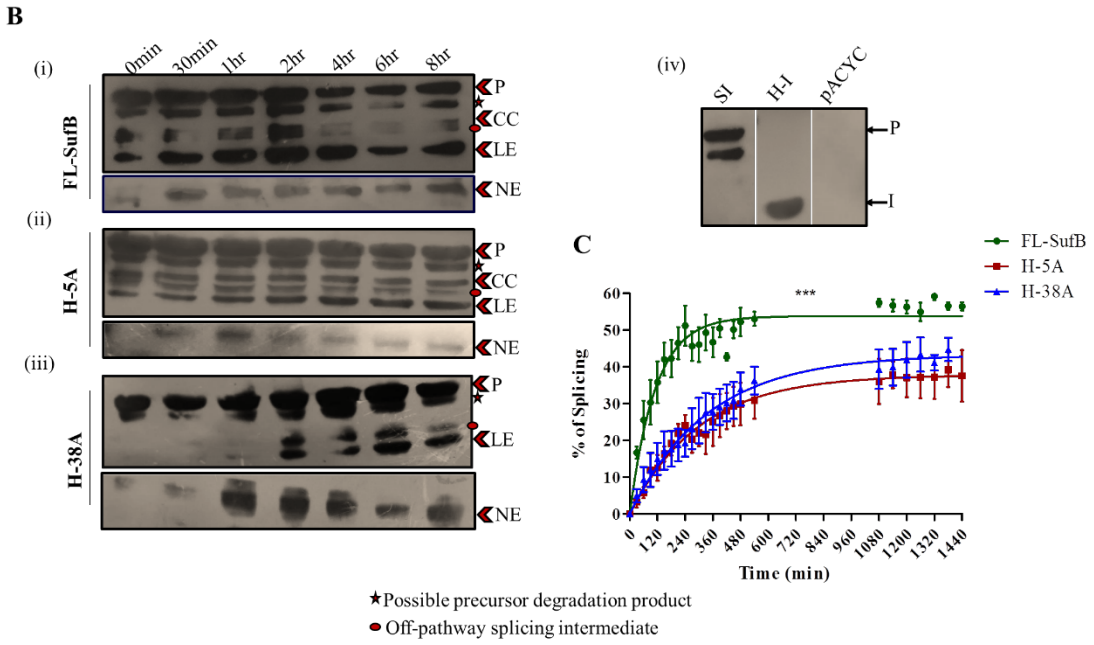


Figure 3. Effect of H-5 and H-38 mutation on *Mtu* SufB splicing. (**A**) Products from *in-vitro* refolding reactions were resolved through 4~10% gradient SDS PAGE as shown in, (i) FL-SufB precursor, (ii) H-5A and (iii) H-38A SufB mutant proteins exhibiting splicing over different time periods, (**B**) 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.

512 ***3.3.1 His-5Ala and His-38Ala SufB mutants exhibit attenuated splicing reaction***

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

520 Splicing efficiency was plotted over time as pseudo-first-order reaction kinetics in GraphPad
521 Prism version 9.0 or 8.0 (Figure 3C). At 20°C, WT FL-SufB shows a splicing efficiency (Y_{max})
522 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 ($t_{1/2}$)
523 of 1.1h (Table 1). Under the same experimental conditions, H-5A SufB mutant splices out with
524 maximum efficiency (Y_{max}) of (37.8±0.68%), a rate constant (K) value of (3.3±0.0001 x10^{-3} min^{-}
525 ^{1}) and a half-life ($t_{1/2}$) of 3.4 h. H-38A SufB mutant splices out with an efficiency (Y_{max}) of
526 (43.35±1.7%), the rate constant (K) (2.9±0.0003x10^{-3}), and half-life ($t_{1/2}$) 3.8h. This clearly
527 demonstrates a 3fold and 3.4-fold reduction ($p<0.0001$; one-way ANOVA) in splicing efficiency
528 for H-5A and H-38A mutants, respectively, relative to FL-SufB precursor. These observations
529 further strengthen our hypothesis that histidines at -5 and -38 positions may have regulatory role
530 on *Mtu* SufB intein splicing.

531

532

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

536

Kinetic parameters	FL-SufB	H-5A	H-38A 537
Y_{max}	53.81±1.03	37.82±1.68	43.35±1.7 538
Rate constant (K) (min^{-1})	$9.92 \pm 0.007 \times 10^{-3}$	$3.3 \pm 0.0003 \times 10^{-3}$	$2.9 \pm 0.0003 \times 10^{-3}$ 539
Half-life($t_{1/2}$)	1.1h	3.4h	3.8h 540

541

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

552

553

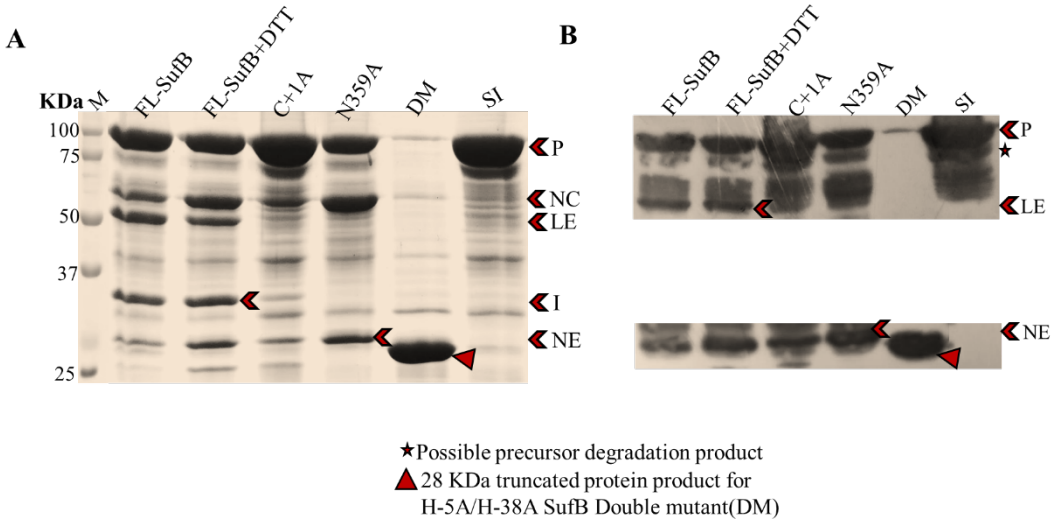


Figure 4. Identification of *Mtu* H-5A/H-38A SufB double mutant(DM): (A) SDS Gel showing splicing and N-cleavage reactions in *Mtu* FL-SufB precursor, C+1A, N359A, DM (H-5A/H-38A) and SI (C1A/N359A) SufB mutant (B) Confirmation of protein products for FL-SufB, C+1A, N359A, DM (H-5A/H-38A) and SI (C1A/N359A) 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.

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).

577 Since the optimum temperature for FL- SufB splicing is found to be 20°C, N-cleavage reactions
578 for the test proteins were also conducted at the same temperature. Purified proteins were renatured
579 in presence of different reducing agents, DTT (50mM), HA (0.5M), and TCEP(2mM) for
580 different durations of time (materials and methods). Products of N-cleavage reactions were
581 visualized by 4%~10% gradient SDS PAGE (Figure 5, 6, and Figure S4) and densitometric
582 analysis of the cleavage products was performed by Gelquant.NET biochemical solutions. The
583 percentage of N-cleavage reaction product was calculated as [(NE+NC/P+NE+NC) x 100].

584 ***3.4.1 H-5A and H-38A SufB mutants display reduced TCEP mediated N-cleavage reaction***

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

594

595

596

597

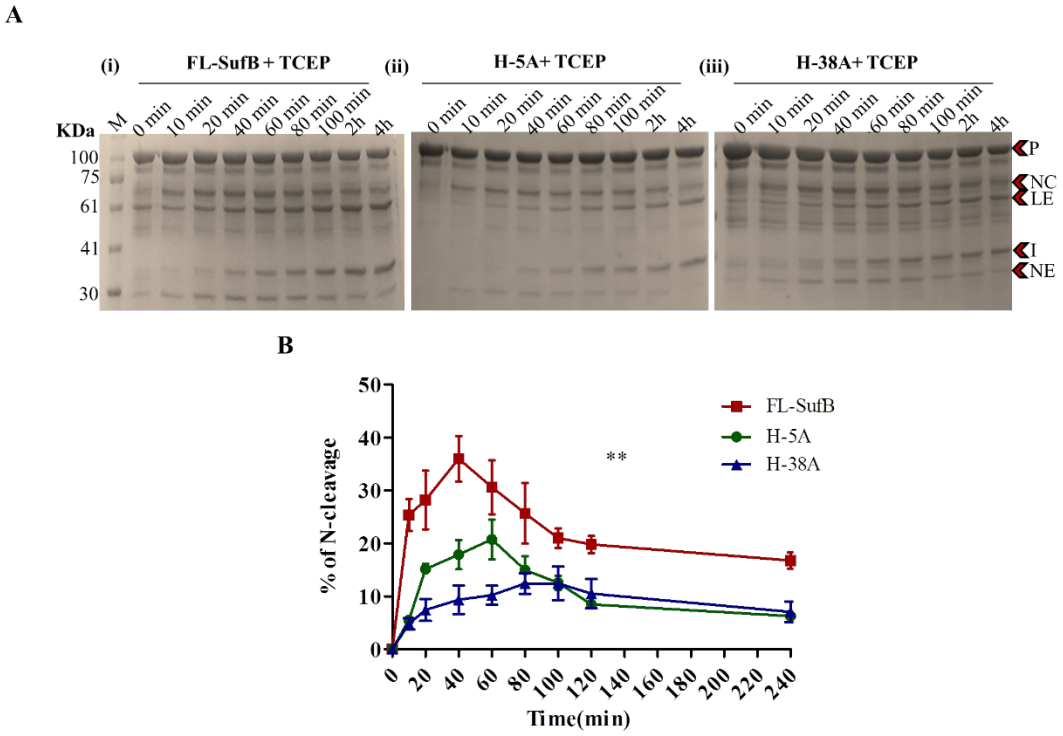


Figure 5. Effect of H-5A and H-38A mutation on TCEP mediated N-cleavage reaction in *Mtu* SufB: (A) SDS Gel image of (i) FL-SufB precursor (ii) H-5A and (iii) H-38A SufB mutant proteins displaying N-cleavage reactions over different time periods; B) (i) Comparative analysis of TCEP mediated reaction 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.

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

623 and the curve was fitted to a pseudo-first-order reaction with the equation, Y=Y0 + (Plateau-Y0)
624 *(1-exp$^{(-K*x)}$), (Figure 6). H-5A and H-38A mutants exhibited about 1.4-fold and 1.5-fold
625 reduction (p=0.0025; one-way ANOVA) in % of N-cleavage, respectively, relative to FL- SufB
626 although the K and $t_{1/2}$ values were comparable [Figure 6C, Table 2 (ii)]. HA is an alpha
627 nucleophile that intercepts between amide-ester equilibrium and induces N-cleavage. It enhances
628 the nucleophilicity of residues that leads to N-cleavage (73, 74). The mutation effect on HA-
629 induced N-cleavage reaction was similar to what is observed in presence of TCEP. However, no
630 significant differences in N-cleavage efficiency were noticed between H-5A, H-38A mutants,
631 and FL- SufB in presence of DTT (Figure S4 and Table S4). Similar to HA, DTT is also a
632 nucleophile that enhances the nucleophilicity of residues and facilitates N- terminal cleavage
633 reaction (72, 75). Besides, DTT interacts directly with the residues that may increase the
634 flexibility of the active site structure (76). Possibly, this abolishes the effect of H-5A and H-38A
635 mutation on N-terminal cleavage. Supported by our current results we hypothesize that histidines
636 at -5 and -38 positions in the N-extein region may have a role in Cys1 activation directly or via
637 structural alignment of other active site residues. This likely explains why catalytic cysteines
638 Cys1 and Cys+1 are interacting differently in His mutants, hindering the N-cleavage reaction.
639 Affirmation of the above hypothesis was done by MD simulations (upcoming section).

640

641

642

643

644

645

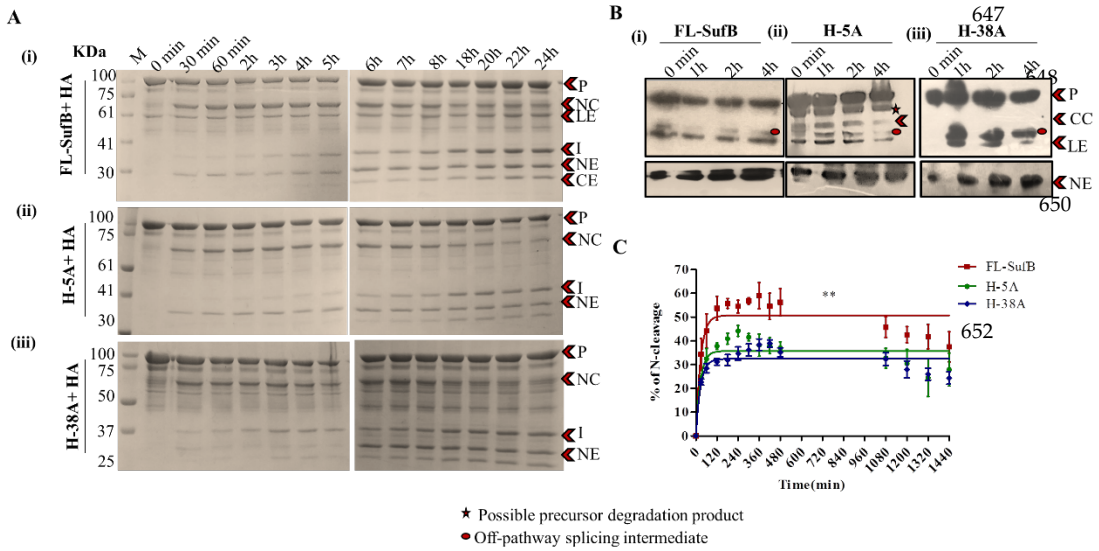


Figure 6. Effect of H-5A and H-38A mutation on HA mediated N-cleavage reaction in *Mtu* SufB : (A) SDS Gel showing HA induced N-cleavage reactions in (i) FL- SufB precursor (ii) H-5A and (iii) H-38A SufB mutant; **(B)** Confirmation of splicing and cleavage products for HA induced N-cleavage reaction 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.

675 **Table 2**. (i) Comparative analysis of TCEP mediated N-terminal cleavage reaction at 20^0C over
676 different time up to 240 mins. The data were extracted from Figure5B. (ii) Different kinetic
677 parameters (Y_{max}, rate constant, and half-life) for HA-induced N-terminal cleavage in FL-SufB
678 precursor, H-5A, and H-38A SufB mutants. These data were extracted from Figure 6C.

679 **(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

680

681 **(ii)**

682

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

683

684

685

686

687

688

689

690 **3.5 Confirmation of protein products**

691 The identity of splicing and cleavage products (P, CC, LE, and NE) for the N-terminal 6X (His)
692 tagged test proteins were confirmed via western blot using anti-His antibodies (Figure 3B, 4B,
693 6B and S4B). SI (C1A/N359A) double mutant SufB was used as a control for the study and gave
694 rise to unspliced precursor protein (P) [Figure 3B (iv)]. Interestingly, one additional protein band
695 was noticed above LE and below CC for FL- SufB, H-5A, and H-38A mutants possibly due to
696 the generation of a splicing intermediate as an off-pathway product. We also observed a product
697 just below the P protein for FL-SufB, SI-SufB, H-5A, and H-38A mutants (Figure3B, 6B, and
698 S4B). However, this protein was missing in cells expressing SufB intein or empty vector pACYC
699 Duet-1 [Figure 3B (iv)]. Hereby, we conclude that we are getting a degradation product of SufB
700 precursor in the cells over-expressing active and inactive *Mtu* full-length (FL) SufB precursor
701 protein.

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

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

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

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

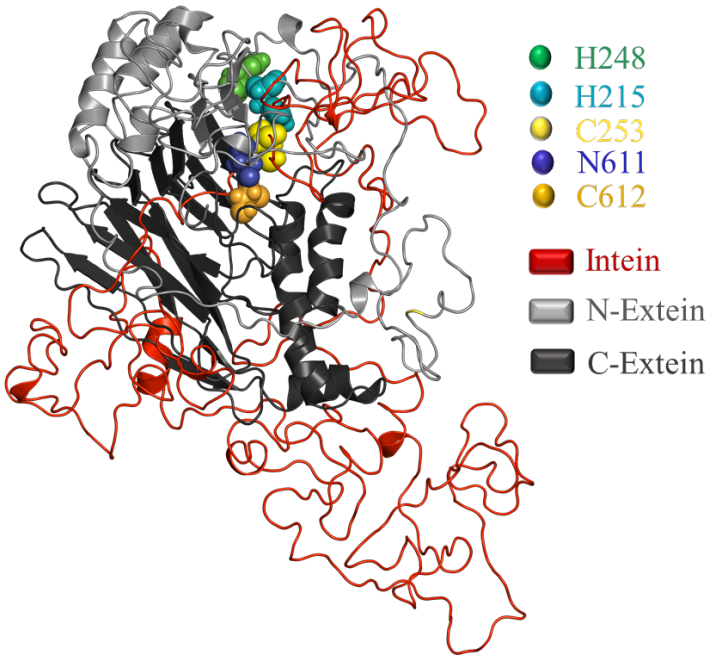
726 **3.6 Proposed mechanism for the cleavage of N- terminal intein~extein peptide bond**

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

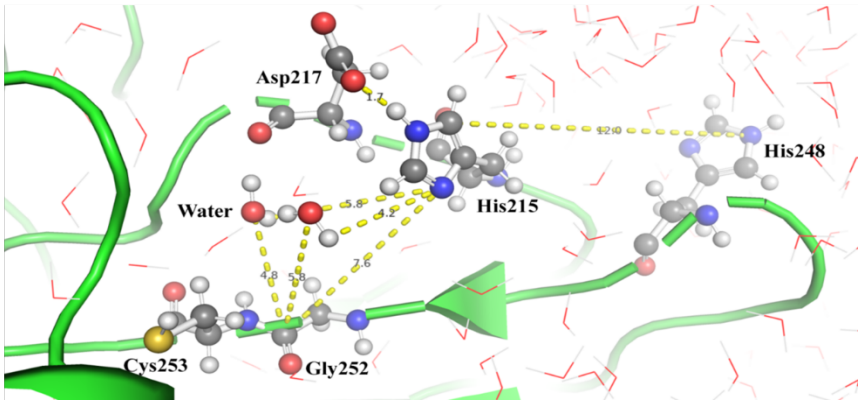
733 To ascertain this hypothesis, we performed MD simulations to predict a 3-dimensional (3D)
734 model for *Mtu* full-length (FL)-SufB precursor protein (Figure 7A). This also included

identification of possible critical residues at the SufB intein active site (N- and C-terminal splice junctions) (Figure 7B and Figure S8), given their proximity to catalytic residues Cys253 (C1), Cys612 (C+1), and Asn611 (N359).

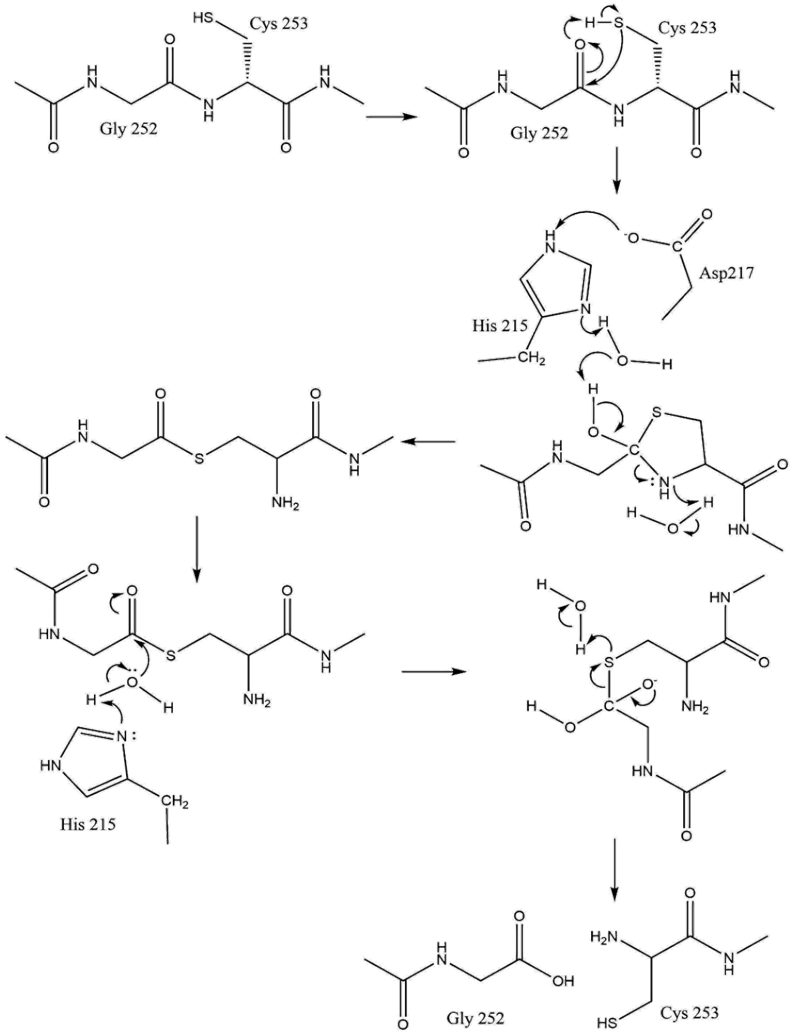
A



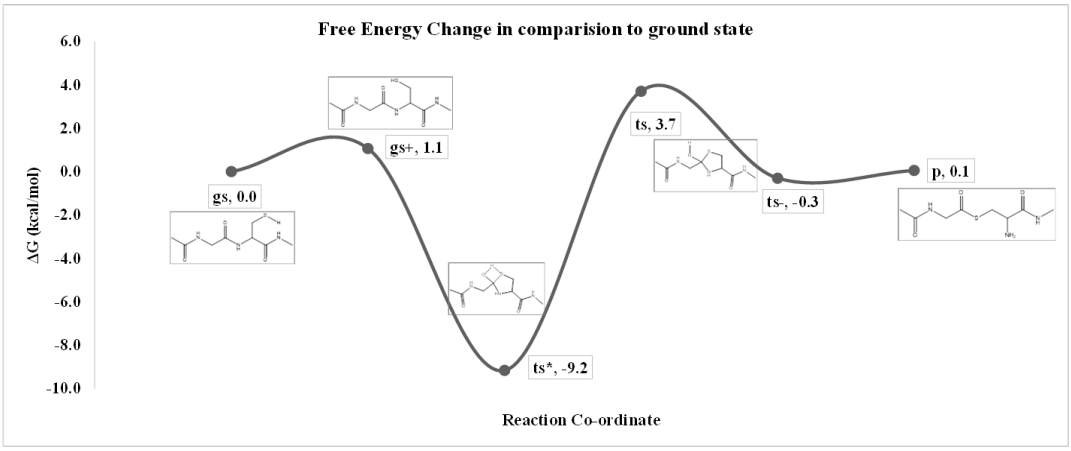
B



C


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

D



E

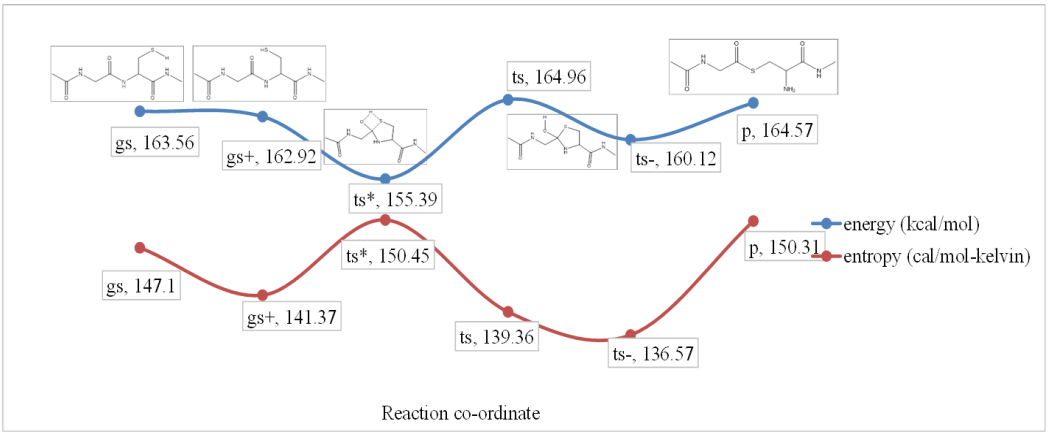


Figure 7. Structural model and proposed mechanism for N-cleavage reaction in *Mtu* FL-SufB precursor. (**A**) The 3D model of *Mtu* FL-SufB precursor obtained in presence of water and NaCl. Different domains and conserved catalytic residues are color coded as shown in the legend, (**B**) Spatial arrangement 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.

793 A chimera structure of *Mtu* FL-SufB protein was built in Schrodinger Bioluminate 2.7 using
794 different templates based on their sequence identity with *Mtu* FL-SufB (Figure 7A). The sequence
795 identity of *Mtu* FL-SufB to SufBCD complex from *Escherichia coli* (5AWF_A), PI-SceI; a
796 homing endonuclease with protein splicing activity (1VDE_A) and intein homing endonuclease II
797 (2CW7_A) was found to be 40%, 45%, and 22% respectively. The resultant structure of *Mtu* FL-
798 SufB was found to have many loops in the intein region because of low sequence identity to the
799 templates. In addition, nearly 13% of amino acids are found to be present in the disallowed region
800 in the Ramachandran plot (Figure S6A). Therefore, we carried out molecular dynamics (MD)
801 simulations of the aforementioned sub-optimized *Mtu* SufB structure in presence of water and
802 NaCl to mimic *in-vivo* conditions in GROMACS for proper folding of the protein. The structure
803 was solvated in water using the SPC/E water model (59, 60) and interactions were described
804 through the OPLS-AA force field (58). The system was minimized by steepest descent
805 minimization with a target maximum force not greater than 1000 kJ mol^{-1}nm. The simulations
806 were done using GROMACS with time steps of 2fs for a duration of 100ns.

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

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

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

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

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

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

859

860 **4. DISCUSSION**

861 ***4.1. Mtu SufB harbors a classic intein splicing system with two conserved histidine residues in***
862 ***the N-extein region***

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

874 Domain analysis of *Mtu* SufB precursor; an 846 aa residue protein clearly demarcated the intein
875 and extein structural regions (Figure 1, Table S2). We found that *Mtu* SufB intein spanning over
876 359 residues has an intact endonuclease domain. Catalytic residues critical for classic intein
877 splicing pathways like Cys1 (Block A), Cys+1 (Block G), penultimate His (Block G), and
878 terminal Asn (Block G) are conserved in different mycobacterial species including *Mtu*. Besides,
879 *Mtu* SufB intein also contains conserved Block B His67 that is known to catalyze the first N-S
880 acyl shift by destabilizing the scissile peptide bond due to loss of resonance and reduction in
881 energy barrier. *In vitro* refolding of *Mtu* SufB precursor gave rise to splicing and cleavage

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

889 ***4.2. A distinct extein~intein partnership mechanism guided by highly conserved H-5 and H-38***
890 ***regulates Mtu SufB N-cleavage reaction***

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

900 To ascertain the roles of H-5 and H-38, a detailed kinetic analysis was performed on *Mtu* SufB
901 full-length (FL) precursor and mutant proteins (H-5A and H-38A) [Figure 3, 4, and 5]. We found
902 that both H-5A and H-38A mutants exhibit a sluggish splicing reaction relative to FL- SufB
903 protein under optimum experimental conditions (Table 1). N-cleavage reaction kinetics of the

904 aforesaid histidine mutants were examined in presence of TCEP and nucleophiles such as
905 Hydroxylamine and DTT. H-5A and H-38A mutants demonstrated a significant reduction in N-
906 cleavage efficiency in comparison to FL- SufB protein in presence of TCEP and nucleophile HA
907 (Table 2 (i) and (ii), Figure5 and 6). But DTT-induced thiolysis reversed the effect of alanine
908 mutation in the histidine mutants. DTT is a thiol protectant reducing agent and exhibits a more
909 efficient thiolysis and di-sulfide reduction. A strong nucleophile such as DTT abolishes the
910 inhibitory effect of alanine mutation on SufB N-cleavage reaction. Herein, we may conclude that
911 H-5 and H-38 possibly have an assistive role in *Mtu* SufB N-cleavage reaction along with other
912 active site residues. Consequently, the effect of H-5A and H-38A mutation can be reversed in
913 presence of strong nucleophile(s) like DTT (Figure S4).

914 To further ascertain the roles of conserved histidines, a 3D model of the *Mtu* FL-SufB protein
915 was built by homology modeling. Molecular dynamics (MD) simulations were performed to
916 equilibrate the aforementioned sub-optimized *Mtu* SufB structure. We obtained an interesting
917 result during the trajectory analysis of the dynamical data. Although H-38 is relatively distal to
918 the N-terminal splice junction in the protein sequence, in the 3-dimensional structure, it comes
919 closer to the active site. MD results suggest that His248 (-5) and His215 (-38) act in a concerted
920 manner to facilitate *Mtu* SufB N-cleavage reaction. Thus, we propose that concomitant effects of
921 N-terminal active site residues such as Gly252 (-1), Cys253 (1), and His215 (-38) supported by
922 His248 (-5) contribute towards *Mtu* SufB N-terminal cleavage reaction. Next, QM/MM
923 calculations were done to determine the energetics of the proposed N-cleavage reaction. Taken
924 together, we hereby propose a detailed mechanism for *Mtu* SufB N-cleavage reaction. A
925 schematic for the same is depicted in (Figure 7C) along the line of canonical intein splicing
926 mechanism.

927 In the process to form thioester intermediate from peptide bond of Gly-Cys dipeptide capped with
928 a methyl group, the steps are as follows. The thiol group of Cys253 (1) comes closer to peptidyl
929 C=O by a conformational change to a rotational transition state. This state is enthalpy favored.
930 Then the thiol H is polarized towards the peptidyl O atom. The C=O and S-H experiences change
931 in charge density in the system. The C, S, H atoms in the system experience decrease in electron
932 density, and the O atom experiences an increase in electron density. In the next transition state,
933 C and S come closer as at this state the imaginary vibrational frequency displacement vector of
934 C=O is towards thiol. Thus, a C-S bond is formed. The proton migration from S to O is highly
935 favored energetically and entropically with a free energy change of -9.2 kcal/mol. The computed
936 value agrees well with earlier results (82, 83). After this step sp^2 hybridized peptide carbon
937 becomes sp^3 hybridized forming a tetrahedral transition state, a 2-hydroxy thiazolidine ring. This
938 state annihilates the peptide resonance, therefore free energy change of +3.7 kcal/mol in contrast
939 to the ground state of Gly252 (-1) and Cys253 (1) system. To again regain the sp^2 hybridization
940 and decrease the energy of the system from resonance stabilization, the 2-hydroxy group expels
941 proton to the environment. This process is energetically advantageous to the system elucidating
942 negative free energy. In a tetrahedral state, S and N atoms have a Mulliken charge of +0.05 and
943 -0.55. Thus, the N atom of the thiazolidine ring accepts a proton from the environment to the
944 system for charge neutrality. This process is also entropically preferred to form thioester
945 intermediate (82, 84). The conversion of a peptide bond to a thioester bond is energetically
946 equivalent at the start as well as the end of the reaction which suggests this process is entropy
947 driven. Further thioester is hydrolyzed by the imidazole side chain of His215 (-38) mediated by
948 water. Base catalyzed thioester hydrolysis in an aqueous environment is well known (85). It is
949 also important to point out that the least distance between His215 (-38) and His248 (-5) obtained

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

961 ***4.3. Biological significance of conserved histidines in the N-extein sequence of Mtu SufB***

962 [Fe-S] cluster-bearing proteins have important physiological roles in electron transfer, redox
963 regulation, metabolic pathways, cellular responses to external stimuli, and as regulators of gene
964 expression (42). The role of the [Fe-S] cluster is closely associated with the functionality of their
965 bound protein framework. In mycobacteria, the SUF system is the sole pathway for [Fe-S] cluster
966 assembly and repair, especially in response to oxidative stress and iron limiting conditions inside
967 macrophages (43, 49). During such an event, the intracellular Fe supply is either from siderophore
968 chelation or via the metabolism of [Fe-S] clusters associated with specific proteins (42). Thus,
969 iron homeostasis plays a major role in mycobacterial survival and virulence. SufB is a [Fe-S]
970 cluster scaffold protein and a vital component of the functional SUF system, and indirectly
971 promotes mycobacterial persistence under stress (49). Further, SufB has been implicated in
972 mycobacterial iron metabolism (42).

973 Earlier studies have specified the bonding of Fe in [Fe-S] clusters to mostly cysteines from the
974 protein backbone although there is increasing evidence for other ligands such as histidine,
975 aspartate, arginine, threonine, and tyrosine. The most common alternative ligand for [Fe-S]
976 cluster coordination is histidine that is highly conserved with a role in redox tuning and proton-
977 coupled electron transfer (86). His433 (SufB) and His360 (SufD) are identified as key protein-
978 ligands for the *de novo* [Fe-S] cluster assembly in *E. coli* (48). It has been shown that these non-
979 cysteine ligands can influence the stability and reactivity of [Fe-S] clusters.

980 We have identified two highly conserved His-5 and His-38 residues in the N-extein sequence of
981 *Mtu* SufB. These were shown to interact concordantly to initiate N-cleavage reaction via catalytic
982 Cys1 activation. Mutational analysis of alanine-substituted His-5 and His-38 residues revealed
983 significantly attenuated cleavage and splicing rates. Attempts to express and analyze the *Mtu*
984 SufB double mutant (His-5A/ His-38A) resulted in a truncated protein product. One of the
985 possible explanations could be the concomitant role of His-5 and His-38 in SufB precursor
986 stabilization. Lack of the histidine residues at the native position perhaps leads to loss of protein
987 stability forming off-pathway splicing intermediate and/or protein degradation product. In
988 summary, the two histidines (His-5 and His-38) work concurrently to activate/initiate N-cleavage
989 reaction leading to the generation of functional SufB protein via protein splicing. As proposed by
990 earlier studies, apart from Cys, His is also an important ligand for metal coordination (48, 86).
991 Thus, it is plausible that the highly conserved His-5 and His-38 possibly coordinate to Fe^{+2}/ Fe^{+3}
992 during [Fe-S] cluster biogenesis. Lastly, His-5 and His-38 together might also play a role in *Mtu*
993 SufB precursor stabilization prior to cleavage and splicing as we were unable to express a full-
994 length SufB double mutant (His-5A and His-38A), unlike the single mutants. However, whether

995 the SufB precursor stabilization is executed by coordination of conserved histidines to [Fe-S]
996 cluster or bonding to Fe^{+2}/ Fe^{+3}, it needs extensive experimental and computational confirmation.

997 Suf is an exclusive system for biogenesis of Fe-S clusters in many pathogenic organisms such as
998 *Staphylococcus aureus*, *Mycobacterium tuberculosis*, *Plasmodium* sp. and *Toxoplasma*, making
999 it an attractive drug target. For instance, D-Cycloserine is a clinical second-line drug currently
1000 used against *M. tuberculosis* and inhibits SufS. Investigation on *S. aureus* revealed that SUF is
1001 the target system for a polycyclic molecule 882 that has direct interaction with SufC (69).
1002 Likewise, the current work may have an implication towards the development of novel anti-TB
1003 drugs targeting SufB destabilization, to fight tuberculosis.

1004 **Data Availability:** All the data are available in the manuscript and supplemental data.

1005 The accession code for the *Mycobacterium tuberculosis* SufB protein used in this study is *P9WFP7*
1006 *(UniProtKB/Swiss-Prot) and* WP_003407484.1, GI: 397673309 (NCBI protein database).

1007 ***Confirmation of the identity of different splicing and cleavage products of Mtu FL-SufB protein***
1008 ***by mass spectrometric analysis:*** The data has been deposited to the ProteomeXchange Consortium
1009 via PRIDE partner repository with data set identifier PXD015199. The details of the submission
1010 are given below.

1011 **Project Name**: Identification of full length, splicing and cleavage products of SufB protein of
1012 SUF-complex of *Mycobacterium tuberculosis*.

1013 **Project accession**: PXD015199

1014 ***Reviewer account details:***

1015 **Username**- reviewer73867@ebi.ac.uk

1016 **Password**: bRxRcbvP

1017 ***Confirmation of the identity of H-38A (215)/H-5A (248) Mtu SufB double mutant product by***
1018 ***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; Writing- Reviewing and Editing; Visualization. **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, Belfort M. Lightning strikes twice: intron–intein coincidence. Proceedings of the National Academy of Sciences. 1998;95(4):1356-7.
2. Lennon CW, Belfort M. Inteins. Current Biology. 2017;27(6):R204-R6.
3. Starokadomskyy P. Protein splicing. Molecular Biology. 2007;41(2):278-93.
4. Novikova O, Topilina N, Belfort M. Enigmatic distribution, evolution, and function of inteins. Journal of Biological Chemistry. 2014;289(21):14490-7.
5. Elleuche S, Pöggeler S. Inteins, valuable genetic elements in molecular biology and biotechnology. Applied Microbiology and Biotechnology. 2010;87(2):479-89.
6. Lennon CW, Stanger M, Banavali NK, Belfort M. Conditional Protein Splicing Switch in Hyperthermophiles through an Intein-Extein Partnership. mBio. 2018;9(1):e02304-17.

7. Lennon CW, Stanger MJ, Belfort M. Mechanism of single-stranded DNA activation of recombinase intein splicing. Biochemistry. 2019;58(31):3335-9.
8. Nanda A, Nasker SS, Mehra A, Panda S, Nayak S. Inteins in Science: Evolution to Application. Microorganisms. 2020;8(12):2004.
9. Panda S, Nanda A, Nasker SS, Sen D, Mehra A, Nayak S. Metal effect on intein splicing: A review. Biochimie. 2021.
10. Lennon CW, Stanger M, Belfort M. Protein splicing of a recombinase intein induced by ssDNA and DNA damage. Genes & development. 2016;30(24):2663-8.
11. Kelley DS, Lennon CW, Li Z, Miller MR, Banavali NK, Li H, et al. Mycobacterial DnaB helicase intein as oxidative stress sensor. Nature communications. 2018;9(1):1-15.
12. Eryilmaz E, Shah NH, Muir TW, Cowburn D. Structural and dynamical features of inteins and implications on protein splicing. Journal of Biological Chemistry. 2014;289(21):14506-11.
13. Hiraga K, Soga I, Dansereau JT, Pereira B, Derbyshire V, Du Z, et al. Selection and structure of hyperactive inteins: peripheral changes relayed to the catalytic center. Journal of molecular biology. 2009;393(5):1106-17.
14. Binschik J, Mootz HD. Chemical Bypass of Intein-Catalyzed N–S Acyl Shift in Protein Splicing. Angewandte Chemie International Edition. 2013;52(15):4260-4.
15. Du Z, Shemella PT, Liu Y, McCallum SA, Pereira B, Nayak SK, et al. Highly Conserved Histidine Plays a Dual Catalytic Role in Protein Splicing: A p K a Shift Mechanism. Journal of the American Chemical Society. 2009;131(32):11581-9.
16. Du Z, Zheng Y, Patterson M, Liu Y, Wang C. 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. 2011;133(26):10275-82.
17. Friedel K, Popp MA, Matern JC, Gazdag EM, Thiel IV, Volkmann G, et al. A functional interplay between intein and extein sequences in protein splicing compensates for the essential block B histidine. Chemical science. 2019;10(1):239-51.
18. Mujika J, Lopez X. Unveiling the Catalytic Role of B-Block Histidine in the N–S Acyl Shift Step of Protein Splicing. The Journal of Physical Chemistry B. 2017;121(33):7786-96.
19. Pereira B, Shemella PT, Amitai G, Belfort G, Nayak SK, Belfort M. Spontaneous proton transfer to a conserved intein residue determines on-pathway protein splicing. Journal of molecular biology. 2011;406(3):430-42.
20. Ding Y, Xu M-Q, Ghosh I, Chen X, Ferrandon S, Lesage G, et al. 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. 2003;278(40):39133-42.
21. Klabunde T, Sharma S, Telenti A, Jacobs WR, Sacchettini JC. Crystal structure of GyrA intein from Mycobacterium xenopi reveals structural basis of protein splicing. Nature structural biology. 1998;5(1):31-6.
22. Frutos S, Goger M, Giovani B, Cowburn D, Muir TW. Branched intermediate formation stimulates peptide bond cleavage in protein splicing. Nature chemical biology. 2010;6(7):527-33.
23. Mujika JI, Lopez X, Mulholland AJ. Modeling protein splicing: reaction pathway for C-terminal splice and intein scission. The journal of physical chemistry B. 2009;113(16):5607-16.
24. Shao Y, PAULUS E. 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. 1997;50(3):193-8.

1115 25. Johnson MA, Southworth MW, Herrmann T, Brace L, Perler FB, Wüthrich K. NMR
1116 structure of a KlbA intein precursor from Methanococcus jannaschii. Protein Science.
1117 2007;16(7):1316-28.
1118 26. Southworth MW, Benner J, Perler FB. An alternative protein splicing mechanism for
1119 inteins lacking an N-terminal nucleophile. The EMBO Journal. 2000;19(18):5019-26.
1120 27. Brace LE, Southworth MW, Tori K, Cushing ML, Perler F. The Deinococcus radiodurans
1121 Snf2 intein caught in the act: detection of the Class 3 intein signature Block F branched
1122 intermediate. Protein Science. 2010;19(8):1525-33.
1123 28. Tori K, Dassa B, Johnson MA, Southworth MW, Brace LE, Ishino Y, et al. Splicing of the
1124 mycobacteriophage Bethlehem DnaB intein: identification of a new mechanistic class of inteins
1125 that contain an obligate block F nucleophile. Journal of Biological Chemistry. 2010;285(4):2515-
1126 26.
1127 29. Tori K, Perler FB. Expanding the definition of class 3 inteins and their proposed phage
1128 origin. Journal of bacteriology. 2011;193(8):2035-41.
1129 30. Amitai G, Callahan BP, Stanger MJ, Belfort G, Belfort M. Modulation of intein activity by
1130 its neighboring extein substrates. Proceedings of the National Academy of Sciences.
1131 2009;106(27):11005-10.
1132 31. Shah NH, Eryilmaz E, Cowburn D, Muir TW. Extein residues play an intimate role in the
1133 rate-limiting step of protein trans-splicing. Journal of the American Chemical Society.
1134 2013;135(15):5839-47.
1135 32. Shemella P, Pereira B, Zhang Y, Van Roey P, Belfort G, Garde S, et al. Mechanism for
1136 intein C-terminal cleavage: a proposal from quantum mechanical calculations. Biophysical
1137 journal. 2007;92(3):847-53.
1138 33. Sun P, Ye S, Ferrandon S, Evans TC, Xu M-Q, Rao Z. Crystal structures of an intein from
1139 the split dnaE gene of Synechocystis sp. PCC6803 reveal the catalytic model without the
1140 penultimate histidine and the mechanism of zinc ion inhibition of protein splicing. Journal of
1141 molecular biology. 2005;353(5):1093-105.
1142 34. Topilina NI, Novikova O, Stanger M, Banavali NK, Belfort M. Post-translational
1143 environmental switch of RadA activity by extein–intein interactions in protein splicing. Nucleic
1144 acids research. 2015;43(13):6631-48.
1145 35. Cheriyan M, Pedamallu CS, Tori K, Perler F. Faster protein splicing with the Nostoc
1146 punctiforme DnaE intein using non-native extein residues. Journal of Biological Chemistry.
1147 2013;288(9):6202-11.
1148 36. Chong S, Williams KS, Wotkowicz C, Xu M-Q. Modulation of protein splicing of the
1149 Saccharomyces cerevisiae vacuolar membrane ATPase intein. Journal of Biological Chemistry.
1150 1998;273(17):10567-77.
1151 37. Oeemig JS, Zhou D, Kajander T, Wlodawer A, Iwaï H. NMR and crystal structures of the
1152 Pyrococcus horikoshii RadA intein guide a strategy for engineering a highly efficient and
1153 promiscuous intein. Journal of molecular biology. 2012;421(1):85-99.
1154 38. Shemella PT, Topilina NI, Soga I, Pereira B, Belfort G, Belfort M, et al. Electronic
1155 structure of neighboring extein residue modulates intein C-terminal cleavage activity. Biophysical
1156 journal. 2011;100(9):2217-25.
1157 39. Topilina NI, Green CM, Jayachandran P, Kelley DS, Stanger MJ, Piazza CL, et al. SufB
1158 intein of Mycobacterium tuberculosis as a sensor for oxidative and nitrosative stresses.
1159 Proceedings of the National Academy of Sciences. 2015;112(33):10348-53.

1160 40. Mills KV, Johnson MA, Perler FB. Protein splicing: how inteins escape from precursor
1161 proteins. Journal of Biological Chemistry. 2014;289(21):14498-505.
1162 41. Cheriyan M, Chan S-H, Perler F. Traceless splicing enabled by substrate-induced
1163 activation of the Nostoc punctiforme Npu DnaE intein after mutation of a catalytic cysteine to
1164 serine. Journal of molecular biology. 2014;426(24):4018-29.
1165 42. Huet G, Daffé M, Saves I. Identification of the Mycobacterium tuberculosis SUF
1166 machinery as the exclusive mycobacterial system of [Fe-S] cluster assembly: evidence for its
1167 implication in the pathogen's survival. Journal of bacteriology. 2005;187(17):6137-46.
1168 43. Outten FW. Recent advances in the Suf Fe–S cluster biogenesis pathway: beyond the
1169 Proteobacteria. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research.
1170 2015;1853(6):1464-9.
1171 44. Lin C-W, McCabe JW, Russell DH, Barondeau DP. Molecular Mechanism of ISC Iron–
1172 Sulfur Cluster Biogenesis Revealed by High-Resolution Native Mass Spectrometry. Journal of the
1173 American Chemical Society. 2020;142(13):6018-29.
1174 45. Mettert EL, Kiley PJ. Fe–S proteins that regulate gene expression. Biochimica et
1175 Biophysica Acta (BBA)-Molecular Cell Research. 2015;1853(6):1284-93.
1176 46. Novikova O, Jayachandran P, Kelley DS, Morton Z, Merwin S, Topilina NI, et al. Intein
1177 clustering suggests functional importance in different domains of life. Molecular biology and
1178 evolution. 2016;33(3):783-99.
1179 47. Outten FW, Djaman O, Storz G. A suf operon requirement for Fe–S cluster assembly
1180 during iron starvation in Escherichia coli. Molecular microbiology. 2004;52(3):861-72.
1181 48. Yuda E, Tanaka N, Fujishiro T, Yokoyama N, Hirabayashi K, Fukuyama K, et al. Mapping
1182 the key residues of SufB and SufD essential for biosynthesis of iron-sulfur clusters. Scientific
1183 reports. 2017;7(1):9387.
1184 49. Huet G, Castaing J-P, Fournier D, Daffé M, Saves I. Protein splicing of SufB is crucial for
1185 the functionality of the Mycobacterium tuberculosis SUF machinery. Journal of bacteriology.
1186 2006;188(9):3412-4.
1187 50. Larking M, Blackshields G, Brown N, Chenna R, McGettigan G, McWilliam H, et al.
1188 ClustalW and ClustalX version 2. Bioinformatics. 2007;23(21):2947-8.
1189 51. Perler FB. InBase: the intein database. Nucleic acids research. 2002;30(1):383-4.
1190 52. Morgenstern B. DIALIGN 2: improvement of the segment-to-segment approach to
1191 multiple sequence alignment. Bioinformatics (Oxford, England). 1999;15(3):211-8.
1192 53. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary
1193 genetics analysis across computing platforms. Molecular biology and evolution. 2018;35(6):1547-
1194 9.
1195 54. Truszkowski J, Goldman N. Maximum likelihood phylogenetic inference is consistent on
1196 multiple sequence alignments, with or without gaps. Systematic biology. 2016;65(2):328-33.
1197 55. Mills KV, Paulus H. Reversible inhibition of protein splicing by zinc ion. Journal of
1198 Biological Chemistry. 2001;276(14):10832-8.
1199 56. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors
1200 in three-dimensional structures of proteins. Nucleic acids research. 2007;35(suppl_2):W407-W10.
1201 57. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High
1202 performance molecular simulations through multi-level parallelism from laptops to
1203 supercomputers. SoftwareX. 2015;1:19-25.

58. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. 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. 2001;105(28):6474-87.
59. Berendsen H, Grigera J, Straatsma T. The missing term in effective pair potentials. Journal of Physical Chemistry. 1987;91(24):6269-71.
60. Kusalik PG, Svishchev IM. The spatial structure in liquid water. Science. 1994;265(5176):1219-21.
61. Fletcher R, Powell MJ. A rapidly convergent descent method for minimization. The computer journal. 1963;6(2):163-8.
62. Grubmüller H, Heller H, Windemuth A, Schulten K. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. Molecular Simulation. 1991;6(1-3):121-42.
63. Barnes J, Hut P. A hierarchical O (N log N) force-calculation algorithm. nature. 1986;324(6096):446.
64. Hess B, Bekker H, Berendsen HJ, Fraaije JG. LINCS: a linear constraint solver for molecular simulations. Journal of computational chemistry. 1997;18(12):1463-72.
65. Darden T, York D, Pedersen L. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. The Journal of chemical physics. 1993;98(12):10089-92.
66. Perler FB. Protein splicing of inteins and hedgehog autoproteolysis: structure, function, and evolution. Cell. 1998;92(1):1-4.
67. Pearson WR. An introduction to sequence similarity ("homology") searching. Current protocols in bioinformatics. 2013;42(1):3.1. -3.1. 8.
68. Albrecht AG, Netz DJ, Miethke M, Pierik AJ, Burghaus O, Peuckert F, et al. SufU is an essential iron-sulfur cluster scaffold protein in Bacillus subtilis. Journal of bacteriology. 2010;192(6):1643-51.
69. Pérard J, de Choudens SO. Iron–sulfur clusters biogenesis by the SUF machinery: close to the molecular mechanism understanding. JBIC Journal of Biological Inorganic Chemistry. 2018;23(4):581-96.
70. Zhang B, Rapolu M, Liang Z, Han Z, Williams PG, Su WW. A dual-intein autoprocessing domain that directs synchronized protein co-expression in both prokaryotes and eukaryotes. Scientific reports. 2015;5:8541.
71. Kirby AJ, Souza BS, Medeiros M, Priebe JP, Manfredi AM, Nome F. Hydroxylamine as an oxygen nucleophile. Chemical evidence from its reaction with a phosphate triester. Chemical Communications. 2008(37):4428-9.
72. Bachmann A-L, Mootz HD. An unprecedented combination of serine and cysteine nucleophiles in a split intein with an atypical split site. Journal of Biological Chemistry. 2015;290(48):28792-804.
73. Simanenko YS, Popov A, Prokop'eva T, Savelova V, Belousova I. Hydroxylamine anion—an effective α-nucleophile in acyl transfer reactions. Theoretical and Experimental Chemistry. 1994;30(2):61-4.
74. Xu M-Q, Perler FB. The mechanism of protein splicing and its modulation by mutation. The EMBO journal. 1996;15(19):5146-53.
75. Volkmann G, Sun W, Liu XQ. Controllable protein cleavages through intein fragment complementation. Protein Science. 2009;18(11):2393-402.

76. Alliegro MC. 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. 2000;282(1):102-6.
77. Pruitt KD, Tatusova T, Maglott DR. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic acids research. 2005;33(suppl_1):D501-D4.
78. Chong S, Shao Y, Paulus H, Benner J, Perler FB, Xu M-Q. 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. 1996;271(36):22159-68.
79. Dearden AK, Callahan B, Van Roey P, Li Z, Kumar U, Belfort M, et al. A conserved threonine spring-loads precursor for intein splicing. Protein Science. 2013;22(5):557-63.
80. Frisch M, Trucks G, Schlegel H, Scuseria G, Robb M, Cheeseman J, et al. Gaussian 03, Rev. C. 02, Gaussian Inc. Wallingford, CT. 2004.
81. Becke AD. Density-functional thermochemistry. III. The role of exact exchange. The Journal of chemical physics. 1993;98(7):5648-52.
82. Burke HM, McSweeney L, Scanlan EM. Exploring chemoselective S-to-N acyl transfer reactions in synthesis and chemical biology. Nature communications. 2017;8(1):1-16.
83. Devaraj NK, Perrin CL. Approach control. Stereoelectronic origin of geometric constraints on N-to-S and N-to-O acyl shifts in peptides. Chemical science. 2018;9(7):1789-94.
84. Monbaliu J-CM, Dive G, Stevens CV, Katritzky AR. Governing parameters of long-range intramolecular S-to-N acyl transfers within (S)-acyl isopeptides. Journal of chemical theory and computation. 2013;9(2):927-34.
85. Kirsch JF, Jencks WP. Base catalysis of imidazole catalysis of ester hydrolysis. Journal of the American Chemical Society. 1964;86(5):833-7.
86. Bak DW, Elliott SJ. Alternative FeS cluster ligands: tuning redox potentials and chemistry. Current opinion in chemical biology. 2014;19:50-8.

1284

1285

1286

1287

1288

1289

1290