

1 **Spliceosomal Prp8 intein at the crossroads of protein and RNA splicing**

2 **Running title:** Regulation of a spliceosomal intein

3

4 Cathleen M. Green¹, Zhong Li², Olga Novikova¹, Valjean R. Bacot-Davis¹, Fenghan
5 Gao², Nilesh K. Banavali^{3,4}, Hongmin Li^{*,2,4}, and Marlene Belfort^{*,1,2}

6

7 ¹Department of Biological Sciences and RNA Institute, University at Albany, 1400

8 Washington Avenue, Albany, NY 12222; ²Wadsworth Center, New York State

9 Department of Health, 120 New Scotland Ave, Albany, NY 12208; ³Wadsworth Center,

10 New York State Department of Health, Empire State Plaza, P. O. Box 509, Albany, NY

11 12201; ⁴Department of Biomedical Sciences, School of Public Health, University at

12 Albany, 1400 Washington Avenue, Albany, NY 12222

13

14 Key words: divalent metals, crystal structure, *Cryptococcus neoformans*, conditional

15 protein splicing

16

17 *Corresponding authors: Marlene Belfort (mbelfort@albany.edu) or Hongmin Li

18 (hongmin.li@health.ny.gov)

19 **ABSTRACT**

20 The spliceosome is a large ribonucleoprotein complex that removes introns from
21 pre-mRNAs. At its functional core lies the essential Prp8 protein. Across diverse
22 eukaryotes, this protein cofactor of RNA catalysis harbors a self-splicing element, called
23 an intein. Inteins in Prp8 are extremely pervasive and are found at seven different sites in
24 various species. Here, we focus on the Prp8 intein from *Cryptococcus neoformans*, a
25 human fungal pathogen. We solved the crystal structure of this intein, revealing structural
26 homology among self-splicing sequences in eukaryotes, including Hedgehog protein.
27 Working with the *C. neoformans* Prp8 intein in a reporter assay, we find that the
28 biologically relevant divalent metals copper and zinc inhibit intein splicing, albeit by two
29 different mechanisms. Copper likely stimulates reversible modifications on a catalytically
30 important cysteine, whereas zinc binds via the same critical cysteine with a K_d of ~1 nM.
31 An intein-containing Prp8 precursor model is presented, suggesting that metal-induced
32 protein splicing inhibition would disturb function of both Prp8 and the spliceosome.
33 These results indicate that Prp8 protein splicing can be modulated, and that this could
34 alter spliceosome function and RNA splicing under specific conditions.

35 INTRODUCTION

36 The spliceosome is a massive ribonucleoprotein complex that performs intron
37 splicing, an important process for maintaining genome diversity in eukaryotes. At the
38 heart of the spliceosome is **pre-mRNA processing factor 8** (Prp8), a large (~250 kDa)
39 and highly conserved protein (1). Prp8 helps generate mature mRNA by coordinating
40 critical rearrangements at the catalytic core of the spliceosome. This essential protein has
41 been implicated in human disease (2, 3), is evolutionarily linked to group II introns (4, 5),
42 and is structurally related to telomerase (6). Recent advances in structural biology have
43 shed new light onto both Prp8 and the spliceosomal machinery at atomic resolution,
44 unveiling an unprecedented level of detail into the molecular steps of intron splicing (5,
45 7-12).

46 A particular reason for our interest in Prp8 is that across several organisms, this
47 large protein contains a self-splicing intein at different positions, implying independent
48 acquisition. Inteins are *internal proteins* that invade at the DNA level and undergo
49 transcription and translation with the host gene (13-15). The intein-containing precursor
50 undergoes protein splicing, a process that excises the intein and ligates the flanking
51 sequences, called exteins, to form the functional protein. Inteins are often bipartite,
52 encoding a splicing domain for excision and ligation, and an endonuclease domain for
53 homing (16, 17). Since some inteins are mobile, they are generally considered selfish
54 genetic elements, but new research indicates that inteins can post-translationally regulate
55 proteins (18-25).

56 Inteins are found in all three domains of life, and are especially abundant in
57 bacteria and archaea (26). In eukaryotes, inteins are sparse, but have been found in

58 nuclear and chloroplast genomes with distinct patterns of insertion (27). Nuclear inteins
59 tend to be in proteins that are involved in energy metabolism and RNA processing,
60 whereas chloroplast inteins are found in proteins that carry out transcription and
61 replication. Out of all the intein-harboring proteins in eukaryotes, Prp8 is
62 overwhelmingly favored. There are over one hundred inteins identified across various
63 sites of Prp8 in different species.

64 Pathogenic fungi seem to be enriched for inteins (27, 28). Several notable human
65 pathogens contain Prp8 inteins, including *Aspergillus fumigatus*, *Histoplasma*
66 *capsulatum*, and *Cryptococcus neoformans*. Intriguingly, many organisms with Prp8
67 inteins also tend to be intron-rich (29). The presence of inteins in Prp8 and the correlation
68 with intron density beg the question of an intein benefit to the host, and especially to
69 pathogens. To begin to answer this question, we focus on the Prp8 intein from *C.*
70 *neoformans*. This is a mini-intein, naturally lacking the homing endonuclease domain, at
71 only 171 amino acid residues. The intein is also found at a highly conserved site at the
72 center of Prp8, and thus is at the core of the spliceosome (1, 5, 30).

73 Studying the Prp8 intein present in *C. neoformans* addresses questions of
74 conditional protein splicing in an important human pathogen in an entirely new domain
75 of life. Solving the Prp8 intein structure set the stage for beginning such studies and
76 provided evolutionary context, by revealing similarities to the metazoan Hedgehog
77 protein. Biochemical experiments then showed that the *C. neoformans* Prp8 intein is
78 differentially responsive to copper and zinc, metals encountered by pathogens in immune
79 cells during infection. Both metals inhibit protein splicing, largely through a critical,
80 catalytically active cysteine. These results suggest that inteins in eukaryotes might also be

81 post-translational sensors, echoing an emerging theme of cysteine-based regulation
82 observed in bacterial and archaeal systems. Further, creation of a Prp8 precursor model
83 illustrates how intein presence relates to the host protein and hints at how the intein could
84 influence spliceosome dynamics and RNA splicing.

85

86 **RESULTS**

87 **Prp8 is an intein hot-spot with diverse insertion sites**

88 Recent data mining revealed over 100 inteins in the Prp8 protein. These Prp8
89 inteins are present across assorted eukaryotic groups, some of which emerged as far back
90 as ~1,100 million years ago (Mya) (Fig. 1A, left) (27, 31-33). The vast majority of Prp8
91 inteins are found across different fungal species, particularly in Ascomycota, and the rest
92 are dispersed across other eukaryotic phyla (Fig. 1A, left). To characterize these diverse
93 Prp8 inteins, we performed comparative and phylogenetic analyses on a representative
94 subset based on the splicing motifs (Fig. 1; Figs. S1 and S2) (15, 34). In total, there are
95 seven distinct intein insertion points, denoted Prp8-**a** through Prp8-**g** (Fig. 1; Fig. S1) (33,
96 35, 36). With only a few exceptions, fungal Prp8 inteins occupy the same insertion site,
97 Prp8-**a** (Figs. 1A and 1B) (27, 31, 33). The newest insertion site, Prp8-**g**, was found at the
98 N-terminal end of Prp8 in the social amoeba *Acytostelium subglobosum* (37) and is
99 reported here for the first time (Fig. S1B).

100 The reconstructed phylogenetic tree reveals that Prp8 inteins group by insertion
101 site (Fig. 1A, right; Fig. S1A). Although insertion sites **b** through **g** have limited
102 representation, the observed clustering, as well as the level of sequence divergence
103 between inteins from different insertion sites, suggests multiple independent intein

104 invasion events throughout evolutionary history. Importantly, bifurcation of Prp8-a
105 inteins into two, well-supported clusters (**a1** and **a2**, with a bootstrap value of 92%)
106 indicates recurrent invasion of inteins into site **a** across diverse fungi. All seven insertion
107 sites were mapped to a simplified line diagram of Prp8 and are peppered across the
108 various domains (Fig. 1B).

109 A multiple sequence alignment of the intein splicing motifs, referred to as blocks
110 A, B, F, and G, demonstrates the sequence divergence among Prp8 inteins (Fig. 1C; Fig.
111 S2). Other than identical residues located in blocks A, B, and G (Fig. 1C, black shading;
112 Fig. S2), Prp8 inteins share limited sequence homology. Block A contains the first
113 residue of the intein, a highly conserved cysteine called C1, which performs the first
114 nucleophilic attack of the protein splicing pathway. This amino acid is identical across
115 the disparate Prp8 inteins (Fig. 1C, Block A; Fig. S2). The B block usually carries a
116 highly conserved motif known as TxxH (38). Across the Prp8 inteins, the B block
117 histidine of TxxH is present in all analyzed inteins, whereas the threonine is somewhat
118 conserved (Fig. 1C, Block B; Fig. S2). Also alike across all Prp8 inteins is a terminal
119 asparagine at the C-terminus of the intein in block G, which also participates in splicing
120 (Fig. 1C, Block G; Fig. S2). The first amino acid of the C-extein, known as the +1
121 residue, is usually a cysteine, serine, or threonine, and all Prp8 inteins use one of these as
122 the +1 nucleophile. The F block shows little conservation across Prp8 inteins. The
123 *Saccharomyces cerevisiae* (*Sce*) VMA1 intein, in the vacuolar ATPase, is as similar to
124 Prp8 inteins as other Prp8 inteins are to each other (Fig. 1C), indicating a close ancestral
125 relationship. The poor sequence alignment among Prp8 inteins reinforces that distinct
126 inteins recurrently invaded Prp8.

127 ***C. neoformans* Prp8 intein structure shows homology to eukaryotic self-splicing**
128 **elements**

129 We next solved the crystal structure of the *Cryptococcus neoformans* (*Cne*) Prp8
130 intein found at site **a** (Fig. 2A). This intein was chosen due to its small size, and because
131 it is found in a significant human pathogen. The *Cne* Prp8 intein was solved to 1.75 Å
132 resolution (Fig. 2A). This novel structure helped us to develop a sense of structural
133 relatedness of the *Cne* Prp8 intein to other inteins and intein-like elements, and to later
134 model the intein into both Prp8 and the spliceosome.

135 The *Cne* Prp8 intein structure represents only the second known fungal nuclear
136 intein structure. The first was of the *Sce* VMA1 intein, which was solved with its linker
137 domain, a connector between the splicing blocks and the internal endonuclease domain
138 (39, 40). As with all solved intein structures so far, the *Cne* Prp8 intein has the canonical
139 horseshoe shape, created by pseudo-two-fold symmetry that positions the catalytic N- and
140 C-termini in close proximity (Fig. 2A). Highlighting the splicing blocks (A, B, F, and G),
141 we see the active core that carries out autocatalytic protein splicing (Figs. 2A and 2B)
142 (15, 34). This catalytic center contains the residues essential for splicing: the nucleophilic
143 cysteine, called C1, in block A, and the terminal asparagine (N171) in block G (Fig. 2B).
144 The C1 and N171 are also positioned in the vicinity of the conserved B block TxxH
145 residues (T62 and H65), which are important for priming the intein for autocatalytic
146 excision at its amino terminus (Fig. 2B) (41). All of these residues contribute critically to
147 the protein splicing pathway, which involves a series of nucleophilic attacks, cyclization
148 of the terminal asparagine, and reformation of a peptide bond between the exteins to form
149 the functional protein (42). Overlaying the *Cne* Prp8 intein primary sequence with its

150 secondary structure shows the position of residues from each block within the context of
151 the 3D architecture (Fig. 2C). For example, blocks A and B are far apart in sequence, but
152 fold proximally in 3D space to execute protein splicing (Figs. 2B and 2C). This
153 representation also illustrates that the unresolved regions of the intein are between block
154 B and block F, and likely represent flexible linker sequences of a former endonuclease
155 domain (Fig. 2C).

156 We next performed a 3D BLAST to compare the *Cne* Prp8 intein structure to
157 other solved structures. Unsurprisingly, the *Cne* Prp8 intein structure as the query returns
158 the *Sce* VMA1 intein as the top hit (Fig. 3A, PDB 1GPP) (39). These are both fungal
159 inteins encoded in nuclear genomes. An overlay of the *Cne* Prp8 intein (red) and the *Sce*
160 VMA1 intein (splicing domain in cyan, linker/endonuclease domain in gray) displays
161 remarkable structural similarity in the splicing modules (Fig. 3A, RMSD of 1.06 Å). The
162 unstructured regions in the *Cne* Prp8 intein structure are where the *Sce* VMA1 intein
163 encodes a linker domain (Fig. 3A, dashed red lines). A closer look at the active centers of
164 the *Cne* Prp8 intein and the *Sce* VMA1 intein demonstrates unmistakable overlap of the
165 catalytic residues (Fig. S3A), further confirming the similarities.

166 Remarkably, another top hit from the 3D BLAST is the 17 kDa fragment of the
167 *Drosophila melanogaster* (*Dme*) Hedgehog C-terminal domain (HHc) (43). Hedgehog is
168 an essential signaling molecule in higher eukaryotes with an analogous self-cleaving
169 reaction performed by a highly conserved cysteine (44). The cleavage allows the N-
170 terminal domain of Hedgehog to ligate to a cholesterol molecule, which plays a critical
171 role in metazoan development. There has been considerable speculation about the
172 relatedness of Hedgehog and inteins (43, 45). It was recently reported, based on sequence

173 similarity, that the N-terminal portion of Hedgehog was acquired through horizontal gene
174 transfer from a prokaryote (46). However, a sequence alignment between the *Cne* Prp8
175 intein, the *Sce* VMA1 intein, and *Dme* HHc shows the three eukaryotic self-splicing
176 sequences have only an average of 22.6% sequence identity (Fig. 3B, 26.4% identity
177 VMA1 to Prp8, 19.2% HHc to Prp8, 22.2% VMA1 to HHc). One highly conserved
178 residue across the three proteins is the initiating cysteine, shared by all the sequences, as
179 well as a C-terminal valine (Fig. 3B, black shading). Despite the considerable sequence
180 divergence, a secondary structure overlay demonstrates that these sequences all code for
181 the same structural elements (Fig. 3B). Strikingly, the *Cne* Prp8 intein and *Dme* HHc
182 have an RMSD of 1.88 Å (Fig. 3C, PDB 1AT0), sharing a similar degree of structural
183 relatedness as a bacterial and a fungal intein (Fig. S3B, RMSD 2.22 Å, PDB 2IMZ) (47).
184 These results reinforce the evolutionary connection between inteins and Hedgehog
185 proteins.

186

187 ***C. neoformans* Prp8 intein is responsive to stress**

188 With the structure solved, we next sought to investigate Prp8 intein splicing and if
189 it is regulated in any way. For simplicity, the *Cne* Prp8 intein was studied in *Escherichia*
190 *coli*. Given that full-length Prp8 contains ~2500 amino acids, we cloned the *Cne* Prp8
191 intein into a reporter construct that uses maltose binding protein (MBP) and green
192 fluorescent protein (GFP) as foreign N- and C-exteins, respectively (21, 24) (Fig. 4A).
193 From this construct, termed MIG, which contains five native N- and C-extein residues,
194 expression is induced and splicing products, such as ligated exteins (LE), are visualized
195 using non-denaturing SDS-PAGE and scanning for GFP fluorescence (Fig. 4A, left). Off-

196 pathway C-terminal cleavage (CTC) is also detectable in the gels, and is caused when the
197 terminal asparagine (N171) cyclizes prior to the first step of protein splicing (Fig. 4A,
198 right).

199 First, we observed that the *Cne* Prp8 intein splices well in the foreign context to
200 yield ligated exteins. However, splicing was so rapid that the amount of precursor
201 remaining after induction (0 h) did not provide a suitable dynamic range for performing
202 splicing assays (Fig. 4B, WT). To slow down splicing and accumulate precursor,
203 mutations were made to the last residue of the N-extein (referred to as the -1), a site
204 previously shown to affect splicing rates (48). After random mutagenesis, a slower
205 splicing mutant was isolated (Fig. 4B, A-1V). The MIG Prp8 A-1V mutant has 40%
206 precursor at 0 h and is splicing active over time (Fig. 4B, A-1V). It is worth noting that
207 splicing rates are intein-dependent, given that other Prp8-**a** inteins from fungal pathogens
208 exhibit diverse splicing phenotypes when cloned into MIG (Fig. S4).

209 Next, using MIG Prp8 A-1V, we asked if a condition exists in which intein
210 splicing might be regulated. Treatments chosen were to mimic environmental stress that
211 *C. neoformans* experiences during infection, such as reactive oxygen species (ROS),
212 reactive nitrogen species (RNS), and metals, all of which prevail during the intracellular
213 respiratory burst (Fig. 4C) (49, 50). From this initial panel, the RNS compounds DEA
214 NONOate and Angeli's salt showed significant precursor accumulation (Fig. 4C, DEA
215 and AS). It also appears that copper and zinc can cause splicing inhibition of MIG Prp8
216 A-1V (Fig. 4C, Cu and Zn). Under these conditions, splicing was inhibited by ~50% (Fig.
217 4C). This preliminary compound screen indicates that the *Cne* Prp8 intein may be subject
218 to inhibition by specific conditions that occur during infection.

219 **Splicing inhibition is mechanistically distinct under copper and zinc treatment**

220 Metal binding has been reported for other inteins, and often engages catalytic
221 residues, which would stall protein splicing (47, 51-56). Therefore, we chose to follow up
222 on the observed copper and zinc inhibition by running *in vitro* MIG Prp8 A-1V assays to
223 further assess effects on protein splicing over time (Fig. 5).

224 We found that 1 mM CuSO₄ caused strong splicing inhibition for up to 22 h
225 compared to untreated controls (Fig. 5A, Untreated and Cu). This inhibition persisted up
226 to 30 h (Fig. S5A). To test a copper-binding hypothesis, the same assay was carried out,
227 but after 2 h of incubation with copper, ethylenediaminetetraacetic acid (EDTA) was
228 added in excess. EDTA chelates copper and should strip bound copper from the *Cne* Prp8
229 intein so that splicing can occur. However, addition of EDTA did not rescue splicing,
230 ruling out the possibility of inhibition purely by copper binding (Fig. 5A, Cu + EDTA).

231 Copper is a redox active metal that can cause cysteine oxidation, either by
232 promoting disulfide bond formation or by catalyzing reversible or irreversible oxidative
233 modifications (57). We next tested whether the *Cne* Prp8 intein cysteines are being
234 reversibly modified by copper, which would prevent the C1 from performing the first
235 nucleophilic attack, and has precedent in intein biology (21, 25). We added the reducing
236 agent tris-(2-carboxyethyl)phosphine (TCEP) to the lysates after a 2 h incubation with
237 copper. Strikingly, TCEP completely reversed the splicing inhibition (Fig. 5A, Cu +
238 TCEP). After reduction, MIG Prp8 A-1V precursor conversion into ligated exteins
239 occurred at a rate similar to that of no copper treatment (Fig. 5A, bottom), indicating
240 reversible cysteine oxidation.

241 The *Cne* Prp8 intein only has two cysteines: C1, and C61 in the B block,
242 immediately preceding the TxxH motif (see Fig. 1C, blue arrowheads). The C1 to C61
243 distance is 8.9 Å, too far for a disulfide bond to form (Fig. S5B), although, because of
244 residue flexibility, this does not strictly preclude formation of a disulfide (21, 58, 59). We
245 also found that C61 is not highly conserved across Prp8 inteins (Fig. S5C), and the most
246 commonly used residue at this site is valine. Therefore, to ask if C1 modifications are
247 sufficient to inhibit protein splicing, several mutants of C61 in MIG Prp8 A-1V were
248 tested for splicing activity (Fig. S6A) and treated with copper (Fig. S6B). The C61
249 mutants also showed precursor accumulation (Fig. S6B), suggesting that C1-C61
250 disulfide bonding is not the underlying inhibitory mechanism, and that copper induces at
251 least C1 oxidation, which is enough to cause the non-splicing phenotype.

252 We further confirmed cysteine modification by performing mass spectrometry on
253 purified *Cne* Prp8 intein. This showed a peak shifted by 32 Da, consistent with an
254 addition of two oxygen atoms (Fig. S7A). Additional validation pinpointed reversible
255 sulfenic acid modifications (-SOH) to C1 and C61 with copper treatment (Fig. S7B), but
256 these were also present in the untreated *Cne* Prp8 intein (Fig. S7A). This indicates that
257 the *Cne* Prp8 intein has highly reactive cysteines that can be modified by atmospheric
258 oxygen alone. Such extreme sensitivity has been observed for other inteins that are
259 regulated by cysteine modification (21). At this time, it is unclear whether the
260 modifications in this assay are the result of copper, oxygen in air, or both. Based on our
261 MIG data, reversible, copper-induced cysteine modifications are the most likely
262 explanation for the inhibition we observe (Fig. 5A), and are likely mediated mainly
263 through C1 (Fig. S6).

264 Next, zinc, a metal without redox activity, was added to MIG Prp8 A-1V lysates
265 given that it too was inhibitory in preliminary treatments (Fig. 4C). The addition of 1 mM
266 ZnSO₄ also caused protein splicing inhibition, and for similar time periods (Fig. 5B,
267 Untreated and Zn). To probe the mechanism of zinc inhibition, we followed up with the
268 same EDTA chelation and TCEP reduction. In contrast to copper, EDTA reversed protein
269 splicing inhibition with zinc (Fig. 5B, Zn + EDTA), but reduction did not (Fig. 5B, Zn +
270 TCEP). Thus, zinc likely causes inhibition by binding to the *Cne* Prp8 intein, as it is
271 redox inactive and TCEP treatment yielded no change.

272 To corroborate zinc binding, purified *Cne* Prp8 intein was titrated with zinc in an
273 isothermal titration calorimetry (ITC) experiment. This revealed tight binding of zinc to
274 the intein, with a K_d in the 1 nM range, whereas no copper binding was detected in a
275 biologically relevant range (Fig. 5C, Table 1). To further understand the mechanism of
276 zinc binding, we turned to molecular modeling using the crystal structure to predict a
277 coordination site. The model was built by surveying published zinc-bound structures,
278 assessing the composition of the zinc pockets, and looking for similarities in the *Cne*
279 Prp8 intein. The putative binding site uses two cysteines and two histidines, a common
280 tetrahedral geometry for a non-catalytic zinc coordination (60). Specifically, we propose
281 that C1, C61, and the two B block histidines, H63 and H65, shift slightly to coordinate a
282 zinc ion (Fig. 5D; Figs. S8A and S8B). These residues are all in or nearby critical splice
283 sites, which would explain the splicing inhibition. To test if this is indeed the zinc site,
284 we performed ITC with individual mutants C1A, C61A, H63A, and H65A and a triple
285 C1A/C61A/H65A mutant. This demonstrated a significant decrease in binding, with K_d
286 values increased by 500-fold or more, supporting the putative zinc pocket in the *Cne* Prp8

287 intein (Fig. S8C, Table 1). We confirmed the mutants are structurally sound using 1D
288 proton NMR, so the observed decrease in binding is likely due to loss of a coordination
289 residue (Fig. S8D).

290

291 **A precursor model of Prp8 relates intein retention to spliceosome function**

292 Finally, we wished to determine how protein splicing inhibition might affect both
293 Prp8 and the spliceosome. Therefore, we docked the intein into a known Prp8 structure
294 and generated a precursor model, where the intein is still covalently connected to the
295 exteins. In this model, the bonds flanking the intein were broken at site **a** in Prp8 from a
296 cryo-EM C complex spliceosome solved from *Schizosaccharomyces pombe* (Fig. 6A,
297 PDB 3JB9, chain A) (12). The *Cne* Prp8 intein structure was computationally inserted
298 using an energy optimization protocol, allowing insight into how intein presence might
299 affect Prp8 and RNA splicing in general.

300 From our precursor modeling, it appears that the *Cne* Prp8 intein is fairly well
301 accommodated in the Prp8 protein (Fig. 6A). Prp8 looks like a curved bean, and the
302 intein fits comfortably into its open cleft. This insertion, site **a**, is in a linker domain that
303 is structurally flexible, as demonstrated by poor resolution of this region in solved
304 structures (5). Given this flexibility, the intein might be able to adopt several
305 configurations at site **a**. The intein positioning shown here is an open orientation, and
306 likely the least disruptive, which would allow for the Prp8 exteins to fold properly.
307 Nonetheless, the presence of the intein likely still interrupts Prp8 function, given the
308 importance of this region, the supporting contacts Prp8 makes within the spliceosome,
309 and the RNA splicing defects in Prp8 mutants (1, 2) (see Discussion). Mapping the other

310 insertion sites onto a Prp8 structure from a *S. cerevisiae* spliceosome also reveals that
311 their presence would presumably disrupt Prp8 function, as they too cluster around the
312 active site (Fig. S9, PDB 5GMK, chain A) (10).

313 We next overlaid the Prp8 intein-containing precursor in the spliceosome from *S.*
314 *pombe* (Fig. 6B, PDB 3JB9) (12). Although this cleft region also occupies a relatively
315 sparse area of the spliceosome (Fig. S10), there are a few crucial players, both protein
316 and RNA, in the vicinity of the intein. For example, one essential splicing protein,
317 Cwf19, occupies the same 3D space as the intein (Fig. 6B, Intein-protein). Cwf19 is
318 implicated in displacing the intron lariat/U2 branch helix (61). Furthermore, there are a
319 few important RNAs in the area of the intein (Fig. 6B, Intein-RNA). These include the
320 U2 snRNA and the intron RNA close by, with U6 snRNA appearing distally. Our
321 precursor model suggests that RNA splicing would be arrested or otherwise disturbed by
322 the presence of the intein, particularly in the presence of environmental stressors that
323 inhibit protein splicing (Fig. 4C).

324

325 **DISCUSSION**

326 Here, we have shown that Prp8 inteins are widely distributed across eukaryotes
327 and have invaded the Prp8 protein repeatedly and independently (Figs. 1A and 1B),
328 suggestive of potential adaptation that provides an advantage to the host. The crystal
329 structure of the *Cryptococcus neoformans* (*Cne*) Prp8 intein showed similarities to the
330 metazoan Hedgehog protein and has facilitated studies of function, as well as provided a
331 basis for molecular modeling (Figs. 2, 3, and 6). Initial *in vitro* studies demonstrated that
332 some environmental stressors that are prevalent in infected macrophages are capable of

333 modulating protein splicing of the *Cne* Prp8 intein (Fig. 4C). Specifically, copper and
334 zinc are potent inhibitors of protein splicing, with each metal interacting with the intein in
335 distinct ways, albeit both through the catalytic cysteine, C1 (Fig. 5). Copper likely
336 hinders protein splicing by cysteine oxidation, and zinc inhibits by tenacious binding to
337 the intein (Fig. 5; Figs. S7 and S8). This work supports a growing theme in intein
338 research that underscores the reactivity of catalytic cysteines (21-23, 58, 59). We propose
339 that the *Cne* Prp8 intein, at the nexus of protein and RNA splicing, can sense metals to
340 pause RNA removal during stressful conditions. This is reinforced by an intein-
341 containing Prp8 precursor model, which suggests that protein splicing inhibition would
342 interfere with RNA splicing (Fig. 6; Fig. S10).

343

344 **Prp8 is an intein sink, with functional implications**

345 We demonstrated a broad distribution of Prp8 inteins with multiple insertion sites
346 (Figs. 1A and 1B), a pattern noted by others as well (27, 31, 33). Our data support the
347 notion that Prp8 was invaded repeatedly, including at least twice at site **a** (Fig. 1A, **a1**
348 and **a2**), and the intein retained, with functional implications. We discovered a novel
349 insertion in a social amoeba, site **g**, bringing the total number of known insertion sites in
350 Prp8 to seven (Fig. 1C; Fig. S1B). Similar trends were previously reported with the
351 mycobacterial iron-sulfur cluster assembly protein SufB, which has three distinct
352 insertion sites, and the mycobacteriophage terminase TerL, which has at least 5 intein
353 insertion sites (20, 21, 26). Such bioinformatics observations have led to fruitful research
354 on intein function, which is now beginning to show that inteins can be tuned to respond
355 to environmental cues (19, 21, 23, 24). A striking example is a mycobacterial intein in

356 DnaB helicase, located in the P-loop of the ATPase domain (62), which is sensitive to
357 ROS both *in vitro* and *in vivo* (58).

358

359 **Structural insights into the Prp8 intein**

360 Here, we present the second structure of a eukaryotic intein, and the sole structure
361 of a eukaryotic intein in an essential protein (Fig. 2A). The *Cne* Prp8 intein structure
362 provides insight into the similarity of inteins in eukaryotes (Figs. 3A-C), suggesting that
363 they likely evolved from a common ancestor. Remarkably, the *Cne* Prp8 intein also has a
364 comparable structure to the C-terminus of a Hedgehog protein (Fig. 3C), which executes
365 a cleavage and ligation reaction to cholesterol also by utilizing a cysteine (44). These
366 results suggest that eukaryotic inteins and Hedgehog proteins might be ancestrally
367 related, but why inteins do not exist in metazoan genomes is a puzzle yet to be explained.

368 Around a dozen intein structures have been solved so far, comprising mainly
369 bacterial and archaeal inteins (63). These have proven useful for studying inteins as novel
370 drug targets (51). As inteins often invade essential proteins in pathogens, inhibiting them
371 from splicing out is an attractive option for developing new antimicrobials (51, 64, 65).
372 Progress towards this goal has been made in prokaryotes using the mycobacterial RecA
373 recombinase intein. A co-crystal of the RecA intein and the antineoplastic compound,
374 cisplatin, helped resolve the mechanism of protein splicing inhibition (51). This showed
375 that the platinum ions of cisplatin bind to the RecA intein at its two catalytic cysteines,
376 C1 and C+1. Concurrent work studying cisplatin and the Prp8 intein also demonstrated
377 effective splicing inhibition, both *in vitro* and *in vivo*, although the mechanism is
378 different than the RecA intein (Li et al., in prep). Solving the *Cne* Prp8 intein structure,

379 along with the observed metal inhibition, is a catalyst for advancing these studies in an
380 essential protein in a eukaryote, at an opportune time given that the antifungal pipeline is
381 drying up (66).

382

383 ***C. neoformans* Prp8 intein is responsive to metals, with biological ramifications**

384 Pathogenic microbes occupy niches that expose them to the opposing toxicities of
385 metal ion excess and deprivation (67). A major reason why *C. neoformans* can
386 successfully colonize distinct tissues is its ability to adapt to the ever-changing
387 environments of the host (68, 69). *C. neoformans* infects the lungs and can disseminate to
388 the brain, causing deadly cryptococcal meningitis in immunocompromised individuals
389 (70). As an intracellular pathogen, *C. neoformans* experiences the oxidative burst of the
390 phagolysosome, which exposes it to acute metal stress (71). Levels of copper can reach
391 up to several hundred micromolar, leading to metal ion toxicity, while zinc
392 concentrations are initially high, but drop with ongoing infection (67, 72, 73). These
393 means to thwart fungal proliferation by metal depletion and compartmentalization are
394 known as nutritional immunity (74). Pathogens prevail by devising sophisticated
395 strategies to either acquire metals or shuttle them out of the cell (67, 71, 75). This
396 constant tug of war between the host and pathogen forces continual evolution, and novel
397 means of overcoming metal ion excess and depletion.

398 Here, we speculate that the *Cne* Prp8 intein might provide cryptococci another
399 means to sense metals during infection. For example, a pause in protein splicing may be
400 useful for overcoming high levels of copper or zinc (Figs. 4C and 5). Copper generates
401 destructive ROS intermediates and can displace iron from iron-sulfur clusters (72),

402 whereas both copper and zinc can dislodge divalent metals from other metalloprotein
403 complexes. Like other stressors known to inhibit RNA splicing, copper and zinc would
404 act post-translationally to block Prp8 intein splicing and inhibit spliceosome function
405 until levels of the metals are reduced by scavenger proteins or metal transporters (71). An
406 advantage of these post-translational strategies is reversibility and instantaneous
407 resumption of spliceosomal function with a return to normal conditions.

408 Post-transcriptional programs that regulate expression of intron-containing
409 transcripts in response to environmental cues have been described in the budding yeast,
410 *Saccharomyces cerevisiae* (76) and in *C. neoformans* (77). Work done on alternative
411 splicing in *C. neoformans* supports pausing of spliceosome function (77). This fungus is
412 intron-dense, with over 40,000 introns in its genome, and abundant alternative splicing
413 has been observed (29). Intriguingly, the most common type of aberrant splicing is intron
414 retention (77). Intron retention has even been shown to play a role in virulence and is
415 regulated by environmental conditions (77). If intron retention is an adaptive mechanism
416 for *C. neoformans* to finely tune expression levels in adverse environments, then
417 inhibiting Prp8 intein splicing is a possible means of controlling that intron retention.
418 Here, we propose that intron retention of transcripts is achieved through environmental
419 sensing by the Prp8 intein, which would inhibit protein splicing and thus, cause
420 transcripts to retain introns. The sensing ‘machinery’ of the *Cne* Prp8 intein is C1.
421 Cysteines are powerful and reactive amino acids that endow proteins with catalytic
422 activity, redox chemistry, and metal binding capacity (57). C1 has the full spectrum of
423 cysteine utility, making it an excellent stress sensor.

424

425 **Nexus between protein splicing inhibition and RNA splicing**

426 Metal sensing by the Prp8 intein *in vivo* is yet to be explored. Ideally, an
427 inteinless strain of the native host would be needed. However, obtaining a knockout of an
428 intein in an essential protein has proven challenging. Given these difficulties, we turned
429 to molecular modeling of structures to help predict *in vivo* effects.

430 The intein-containing Prp8 precursor model generated from a solved
431 *Schizosaccharomyces pombe* spliceosome structure revealed a snug accommodation of
432 the intein in a cleft of Prp8 (Fig. 6A, PDB 3JB9, chain A). This insertion (site **a**) is in a
433 linker located between the thumb domain and the endonuclease domain of the reverse
434 transcriptase (Fig. 1B; Fig. S9). This highly conserved region of Prp8 (55%-87% identity
435 over 113 residues) is also known as the 3' splice site fidelity region 3.2, and is likely
436 involved in RNA-mediated catalysis leading to intron removal (1). This site is not only at
437 the core of the protein, but also at the catalytic center of the spliceosome. Although Prp8
438 likely cannot perform its function with an intein present, the structural tolerance may
439 allow for proper folding of the intein, as well as that of Prp8. In the longer term, this
440 flexibility gives the intein freedom to adapt to its surroundings, supporting some degree
441 of Prp8 function in a precursor state. Mini-inteins, such as the one present in *Cne* Prp8,
442 are not mobile and may therefore be under more selective pressure to adapt to their
443 exteins. This is in line with work that shows partial activity of the RadA precursor with
444 its mini-intein intact (19).

445 If the *Cne* Prp8 intein remains unspliced until spliceosome assembly, perhaps due
446 to metal sensing and inhibition, it would almost certainly be disruptive. In the C complex
447 modeled here, the intein would occupy the same location as the Cwf19 protein (Fig. 6B,

448 Intein-protein), which is an essential spliceosome component. Cwf19 plays a central role
449 in helping its binding partner, Prp8, fine-tune motions of the spliceosome involved in
450 intron removal (61). If Cwf19 is unable to bind Prp8 at its cognate site, movements of the
451 spliceosome would be altered and intron splicing disrupted. Furthermore, certain RNAs
452 thread close to the intein (Fig. 6B, Intein-RNA). Although from our modeling there are
453 no observable direct interactions with U2 snRNA, U2 is proximal to the intein. The
454 binding of U2 to the intron branch-point region creates the bulged adenosine, which
455 initiates nucleophilic attack on the 5' splice site (78). If the intein were still present during
456 intron removal, the critical interaction of the intron and U2 would be disturbed or
457 changed in a way that disrupts intron splicing. If the *Cne* Prp8 intein were to sense a
458 stressor, stay lodged in Prp8 and pause intron splicing, Prp8 precursor accumulation
459 would undoubtedly stop the spliceosome from functioning (Fig. S10), leading to pre-
460 mRNA accumulation, as even point mutations in Prp8 are known to do (2). Thereby, this
461 work proposes that the *Cne* Prp8 intein is subject to modifications that would influence
462 Prp8 function, and stall spliceosome activity. Since the observed stress-induced intein
463 modifications are reversible, removal of the stressor would result in an immediate
464 restoration of Prp8 and spliceosome function.

465

466 **MATERIALS AND METHODS**

467 **Bioinformatic and phylogenetic analyses**

468 The Prp8 intein sequences used to build the phylogenetic tree in Fig. 1A and Fig.
469 S1A were accessed from Green et al. 2018. For comparative and phylogenetic analyses,
470 amino acid sequences of inteins were manually trimmed to the splicing blocks (A, B, F,

471 and G). All multiple sequence alignments of the amino acid sequences were performed
472 using ClustalOmega with default parameters (79) and edited manually (Fig. 1; Figs. S1A
473 and S2). Where alignments are shown shaded, black represents an identical amino acid,
474 dark gray is a conserved amino acid, whereby the same amino acid is at the same position
475 in a majority of the sequences, and light gray is a similar amino acid, defined as a semi-
476 conserved amino acid substitution from the same class. Phylogenetic analysis was
477 performed using the Neighbor-Joining (NJ) method in the MEGA7 program (80).
478 Statistical support for the NJ tree was evaluated by Interior-branch test (number of
479 replications, 1000) (81). The sequence logo for the B block was generated based on the
480 multiple sequence alignment using WebLogo3 (82) (<http://weblogo.threeplusone.com>).
481 The seven Prp8 intein insertions were mapped onto a model of a *Saccharomyces*
482 *cerevisiae* Prp8 (Fig. S9, PDB 5GMK). All intein, Prp8, and spliceosome structures were
483 viewed, edited, or aligned using PyMol 1.3. The 3D-BLAST protein structure search was
484 performed by BioXGEM with default parameters (<http://3d-blast.life.nctu.edu.tw>).

485

486 **Bacterial strains and growth conditions**

487 All strains used in the present study can be found in Table S1. *Escherichia coli*
488 DH5 α , MG1655(DE3), and BL21(DE3) were grown in Luria Broth (LB), unless
489 otherwise indicated, with aeration at 250 rpm. Media contained kanamycin (50 μ g/mL) or
490 chloramphenicol (25 μ g/mL) where appropriate. Plasmids were transformed into cells by
491 electroporation using a Bio-Rad Gene Pulser and recovered for 1 h at 37°C in SOC
492 medium (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10

493 mM MgSO₄, and 20 mM glucose). Transformants were selected by plating on LB agar
494 with the appropriate antibiotic and incubated at 37°C overnight.

495

496 **Construction of plasmids**

497 All plasmids used in the present study can be found in Table S2 and all
498 oligonucleotides, synthesized by Integrated DNA Technologies (IDT), are in Table S3.
499 Plasmid DNA was prepared using E.Z.N.A. Plasmid Mini Kit (Omega). DNA was
500 visualized in 1% agarose gels using EZ-Vision DNA Dye (Amresco). PCR fragments
501 were amplified using CloneAmp HiFi PCR Premix (Clontech) from genomic DNA of *C.*
502 *neoformans* var. *grubii* H99 or *C. gattii* NIH444 (Dr. Sudha Chaturvedi, New York State
503 Department of Health), *A. fumigatus* AF293 (Dr. Robert J. Cramer, Dartmouth College),
504 *B. dendrobatitidis* JEL423 (Dr. Timothy James, University of Michigan), or *H.*
505 *capsulatum* G186A (Dr. Chad Rappleye, Ohio State University). For insertion into the
506 MIG construct, the inserts included 5 native N- and C-extein residues flanking the intein.
507 For insertion into the overexpression vector, pET47b, the intein alone with 3 native N-
508 exteins was PCR amplified. Digested backbone was gel purified using Zymoclean Gel
509 DNA Recovery Kit (Zymo Research). Restriction enzymes (NEB), T4 ligase (NEB), and
510 In-Fusion HD Cloning Plus Kit (Clontech) were all used per manufacturer protocol.
511 Mutagenesis was performed using the QuikChange Lightning Site-Directed Mutagenesis
512 Kit (Agilent) for single amino acid mutations or the QuikChange Lightning Multi Site-
513 Directed Mutagenesis Kit (Agilent) for multiple amino acid mutations. For the A-1V
514 mutation, primers were designed to randomly mutate the A-1 to all other possible codons

515 using a degenerate primer with NNS at the mutated location. All clones were verified by
516 sequencing (EtonBio).

517

518 **MIG splicing assays**

519 MIG Prp8 (WT, A-1V, and derived C61 mutants) was transformed by
520 electroporation into MG1655(DE3). The cells were subcultured 1:100 from an overnight
521 culture into fresh LB medium and grown at 37°C with 250 rpm shaking to an OD₆₀₀ of
522 0.5. Cells were induced with 0.5 mM IPTG for 1 h at 30°C and pelleted by spinning for
523 10 min at 4,000 rpm. The pellets were lysed immediately using tip sonication (20 sec
524 on/30 sec off at 30% amplitude for 1 min total) in 50 mM Tris, pH 8.0, and 10% glycerol
525 or stored at -80°C until lysis. For any ROS/RNS or metal treatment, the indicated
526 compound was added to cells at the desired concentration prior to incubation at 30°C for
527 the specified time. EDTA was added to a final concentration of 10 mM and TCEP to a
528 final concentration of 40 mM. Upon completion of the assay or time point, the lysate was
529 frozen at -80°C. To visualize MIG splicing assay results, samples were separated under
530 non-reducing conditions on Novex WedgeWell 12% Tris-Glycine gels (Invitrogen) using
531 loading dye lacking β-mercaptoethanol and visualized using a Typhoon 9400 scanner
532 (GE Healthcare) with excitation at 488 nm and emission at 526 nm. Quantitation and
533 analysis were done using ImageJ and GraphPad Prism (v7.02).

534

535 **Prp8 intein purification**

536 For isothermal titration calorimetry and mass spectrometry, the *Cne* Prp8 intein
537 from *Cryptococcus neoformans* var. *grubii* H99 was amplified with 3 native N-extein

538 residues (EKA) and cloned into pET47b in front of an N-terminal His₆-tag and an HRV
539 3C protease site. For crystallization, the *Cne* Prp8 intein with 2 native N-extein residues
540 (KA) and the native C-extein S+1 was amplified and cloned into pET28a with a C-
541 terminal His₆-tag using a megaprimer approach as described previously (83).

542 The pET47b (or pET28a) *Cne* Prp8 intein construct was transformed by
543 electroporation into BL21(DE3) cells. The cells were subcultured 1:100 from an
544 overnight culture into fresh LB medium and grown to an OD₆₀₀ of 0.6. Cells were
545 induced with 0.5 mM IPTG and grown with shaking at 250 rpm overnight at 16°C. The
546 following morning, cells were harvested by centrifugation at 4,000 rpm for 10 minutes.
547 Pellets were frozen at -80°C until ready for lysis. Tip sonication was performed (30 sec
548 on/59 sec off at 30% amplitude for 4 min total) in buffer containing 20 mM Tris, pH 7.8,
549 500 mM NaCl, 25 mM imidazole, and 5% glycerol. Whole cell lysate was centrifuged at
550 20,000 x g for 20 min to separate the soluble fraction, which was loaded onto a nickel
551 affinity column equilibrated with the lysis buffer. Washes were carried out using buffer
552 containing 20 mM Tris, pH 7.8, 500 mM NaCl, 75 mM imidazole, and 5% glycerol and
553 elution buffer with 20 mM Tris, pH 7.8, 500 mM NaCl, 250 mM imidazole, and 5%
554 glycerol. Purified fractions of the *Cne* Prp8 intein were checked by separation on SDS-
555 PAGE and the cleanest elution samples were pooled. For the pET47b construct, the His₆-
556 tag on the *Cne* Prp8 intein was removed through digestion with HRV 3C protease
557 according to the manufacturer's protocol. The cleaved *Cne* Prp8 intein reaction was
558 passed back over a nickel affinity column and the flow-through was collected to ensure
559 no His₆-tagged *Cne* Prp8 intein or HRV 3C protease was in the sample. For analysis by
560 ITC, the flow-through intein was exchanged into 50 mM C₂H₃NaO₂ (pH 7.0), 100 mM

561 NaCl using a HiPrep 26/10 desalting column or a dialysis cassette. For mass
562 spectrometry, the flow-through intein was used directly for metal treatments and then
563 further purified by liquid chromatography (LC) prior to spraying on the instrument. For
564 the pET28a construct, the imidazole-eluted fractions were concentrated and subjected to
565 size exclusion chromatography by a gel filtrations 16/60 Superdex column (GE
566 Healthcare). For crystallization, the purified *Cne* Prp8 intein was concentrated to 9.5
567 mg/mL in a buffer composed of 25 mM HEPES, pH 7.5, and 150 mM NaCl.

568

569 **Mass spectrometry of Prp8 intein**

570 Purified *Cne* Prp8 intein was reduced with 40 mM TCEP and exchanged into
571 deoxygenated exchange buffer (20 mM Tris pH 7.5, 200 mM NaCl) using 7K MWCO
572 Zeba spin desalting columns (Thermo) to remove TCEP. The protein concentration was
573 measured and then treated with 10X of CuSO₄ and incubated at 30°C for 1 h. Following
574 treatment, the purified intein was denatured with 6 M urea at 37°C for 30 min. The urea
575 concentration was diluted down to less than 0.8 M with 50 mM Tris, pH 7.6, 1 mM
576 CaCl₂. Trypsin digest of the intein was performed by adding activated trypsin (Promega)
577 to a final ratio of 1:20 and incubating overnight at 37°C. The oxidation of *Cne* Prp8 intein
578 cysteines after treatment was analyzed by multiple reaction monitoring-initiated detection
579 and sequencing (MIDAS) as described (84). The trypsin-digested mixture was acidified
580 followed by LC-MS/MS analysis. LC-MS/MS analysis was performed on a microflow
581 LC-MS/MS system configured with a 3-pumping Micromass/Waters CapLC™ system
582 with an autosampler, a stream select module configured for precolumn plus analytical
583 capillary column, and a QTRAP 6500 (ABSCIEX) mass spectrometer fitted with Turbo

584 V microflow source, operated under Analyst 1.63 control. Injected samples were first
585 trapped and desalted isocratically on an LC-Packings PepMap™ C18 µ-Precolumn™
586 Cartridge (5 µm, 500 µm I.D. x 20 mm; Dionex, Sunnyvale, CA, USA) for 7 min with
587 0.1% formic acid delivered by the auxiliary pump at 40 µL/min after which the peptides
588 were eluted from the precolumn and separated on an analytical C18 capillary column (15
589 cm x 500 µm i.d., packed with 5 µm, Jupiter 300 C18 particles, Phenomenex, CA, USA)
590 connected inline to the mass spectrometer, at µL/min using a 50 min gradient of 5% to
591 80% acetonitrile in 0.1% formic acid. The oxidized peptide identification was conducted
592 through multiple reaction monitoring (MRM) triggered enhanced product ion (EPI) scan
593 using information dependent acquisition (IDA). The utilization of chromatographic
594 separation, MRM transitions, and EPI scan allows accurate peptide identification and
595 confirmation. The two MRM transitions including m/z 404.19 > 532.22 and m/z 786.04 >
596 895.41 for C[Oxi]LQNGTR.+2b5 and THEGLEDLVC[Oxi]THNHILSMYK.+3b8 were
597 used to trigger the EPI experiment respectively. The instrument was operated in a
598 positive ion mode with a turbo V ion drive electrospray source. The parameters for the
599 operation were as follows: curtain gas, 20 psi; heated nebulizer temperature 180° C, ion
600 spray voltage, 5500 V; gas1, 18 psi; gas 2, 15 psi, de-clustering potential, 65 V, EP, 10 V
601 and CAD gas, high.

602

603 **Isothermal titration calorimetry of Prp8 intein**

604 ITC measurements were carried out on a TA Instruments Nano ITC (TA
605 Instruments, Inc., New Castle, DE). Aqueous solutions of metal titrants (CuSO₄ or
606 ZnSO₄) were prepared to be 0.3-30-fold higher than the concentration of the *Cne* Prp8

607 intein, in the range of 0.05-5.0 mM. The titrant and *Cne* Prp8 intein samples (wild-type
608 and mutants) were degassed before each titration. The purified *Cne* Prp8 intein was
609 concentrated from 10 μ M to 16 μ M in 300 μ L and were placed in a 2.5 mL reaction cell,
610 and the reference cell was filled with 300 μ L deionized water. All titrations were carried
611 out at 37°C. After baseline equilibration, successive injections of an indicated titrant were
612 made into the reaction cell in 2.5 μ L increments at 400 s intervals with stirring at 250-350
613 rpm to ensure an equilibrium was achieved for a return to baseline. The resulting heats of
614 reaction were measured over 20 consecutive injections. Buffer control experiments [50
615 mM $C_2H_3NaO_2$ (pH 7.0), 100 mM NaCl, \pm 10 mM tris(2-carboxyethyl)phosphine] to
616 determine the heats of titrant dilution were carried out by making identical injections in
617 the absence of the *Cne* Prp8 intein. The net reaction heat was obtained by subtracting the
618 heats of dilution from the corresponding total heat of reaction. The titration data were
619 deconvoluted based on best-fit binding models containing either one or three sets of
620 interacting binding sites, using a nonlinear least-square algorithm through the
621 NanoAnalyze software. The binding enthalpy change (ΔH), dissociation constant (K_d),
622 and binding stoichiometry (n) were permitted to vary during the least-square
623 minimization process, and taken as best-fit values.

624

625 **Crystallization, structure determination, and refinement of Prp8 intein**

626 Initial crystallization conditions were obtained by screening the Hampton
627 crystallization screens (I, II, and Research Index HT), using the hanging-drop vapor
628 diffusion method. Upon optimization, large crystals were grown by mixing 1 μ L of *Cne*
629 Prp8 intein and 1 μ L of reservoir solution containing 22-28% PEG4,000, 0.1 M sodium

630 acetate, pH 4.2, 0.2 M ammonium acetate. The *Cne* Prp8 intein crystallizes in space
631 group *PI* with six intein molecules per asymmetric unit. Prior to data collection, all
632 crystals were transferred to a cryo-protectant solution containing crystallization buffer
633 supplemented with 20% glycerol. The crystals were flash-cooled directly in liquid
634 nitrogen. Diffraction data for the co-crystals were collected at 100 K using a Pilatus
635 detector at the BL9-2 beamline of the Stanford Synchrotron Radiation Laboratory. Data
636 were processed, scaled, and reduced using the programs HKL2000 (85) and Phenix suite
637 (86). The structure of the *Cne* Prp8 intein was determined by molecular replacement,
638 with the crystal structure of the *Cryptococcus gattii* Prp8 intein (Li et al., in prep) as a
639 search model using the PHENIX program suite. Structure refinement was carried out
640 using the Phenix program suite (Table S4).

641

642 **Modeling of zinc binding**

643 The models for zinc ion binding were constructed using the *Cne* Prp8 intein
644 crystal structure conformation (chain A) in the asymmetric unit and a structure deposited
645 in the PDB of human ubiquitin ligase (87), which has a zinc atom bound to two cysteines
646 and two histidine side chains (PDB 5TDA). This ubiquitin ligase structure was chosen
647 because of its high resolution (0.79 Å), and the occurrence of the four side chains present
648 in the *Cne* Prp8 intein coordinating a zinc ion. All the optimizations were performed
649 using the program CHARMM (version c35b3) (88, 89) and the CHARMM36 force field
650 for proteins (90). The best-fit orientations of a deprotonated C1, a deprotonated C61, a
651 neutral H63 side chain, and a neutral H65 side chain of each Prp8 intein crystal structure
652 conformation onto the corresponding C112, C115, H133, and H136 side chains of the

653 human ubiquitin ligase were first obtained using a Kabsch least-squared optimization
654 (91). The coordinates of the zinc ion were then transferred from the ubiquitin ligase
655 structure, and subsequently held fixed. All protein non-hydrogen atoms, except those in
656 C1, C61, H63, and H65, were harmonically restrained with a force constant of 1
657 kcal/mol/Å² to prevent unrelated structural changes in the intein. Root Mean Square
658 deviation (RMSD) restraints were then applied to the C1, C61, H63, and H65 residues to
659 gradually change their conformation to a zinc-coordinating one, as observed in the
660 ubiquitin ligase structure. The RMSD restraints had a force constant of 10000
661 kcal/mol/Å², were applied only to non-hydrogen atoms, and the restraint minimum was
662 gradually changed in 0.5 Å decrements from an initial RMSD of 4.5 Å to a final RMSD
663 of 0 Å. For each restraint minimum value, the following optimization protocol was
664 followed: a Steepest Descent (SD) (92) minimization for 5000 steps with an energy
665 change tolerance criterion of 0.001 kcal/mol, 1000 steps of Langevin dynamics at a
666 temperature of 150 Kelvin and a friction coefficient of 60.0 ps⁻¹, another 5000 SD
667 minimization steps with an energy change tolerance criterion of 0.001 kcal/mol. Each
668 model thus obtained was further minimized using 5000 SD steps with an energy change
669 tolerance criterion of 0.001 kcal/mol, the harmonic restraints on the other protein atoms
670 were removed, and a further 5000 SD steps of minimization with an energy change
671 tolerance criterion of 0.001 kcal/mol were carried out to get the final model.

672

673 **Precursor modeling**

674 The Prp8 protein structure (chain A) in the 3.6 Å cryo-EM structure of the
675 *Schizosaccharomyces pombe* (*Spo*) spliceosome (12) (PDB 3JB9) was used as a

676 structural template for constructing a homology model of the *Cne* Prp8 exteins. A pair-
677 wise Needleman-Wunsch (93) sequence alignment using EMBOSS Needle (94) shows
678 70.6% sequence identity and 83.1% sequence similarity between the full *Spo* and *Cne*
679 extein sequences. The *Cne* Prp8 extein homology model was constructed using
680 MODELLER (95) using the DOPE (96) and GA341 (97) energy functions to identify the
681 best model. The *Spo* Prp8 extein template has longer missing sections at the N- and C-
682 termini, as well as some smaller gaps in between (the missing residues are 1-46, 303-313,
683 1533-1538, 1781-1783, and 2031-2363). The 171 residue *Cne* Prp8 intein sequence is an
684 insert between residues 1530 and 1531 in the extein, which suggests only a relatively
685 short 5-residue gap (*Spo* Prp8 residues 1533-1538) as a factor limiting modeling accuracy
686 of the precursor state. The *Cne* Prp8 extein homology model was then combined with the
687 *Cne* Prp8 intein crystal structure to generate a model for the full *Cne* Prp8 precursor. The
688 orientation of the intein with respect to the extein was manually adjusted using VMD
689 software (98) to avoid any steric overlap and keep the intein and extein ends that are
690 joined by peptide bonds sufficiently close together. This combined intein-extein structure
691 was then used as a structural template to generate a continuous *Cne* Prp8 precursor
692 homology model using MODELLER. Further minimization on this precursor homology
693 model was performed using the program CHARMM, version c35b3 (88, 89) with the
694 CHARMM36 force field for proteins (90). All atoms not within the precursor amino acid
695 sequence range encompassing the intein and its neighboring extein regions (residues
696 1520-1720) were initially held fixed, and a low-temperature (150 K) optimization
697 protocol was used to improve the homology model. This protocol included 5 iterations of
698 the following steps: (a) 5000 steps of Steepest Descent (SD) minimization followed by

699 5000 steps of Adapted-Basis Newton-Raphson (ABNR) minimization, each with an
700 energy change tolerance of 0.001 kcal/mol, (b) 1000 steps of Langevin dynamics at a
701 temperature of 150 Kelvin and a friction coefficient of 5.0 ps⁻¹, (c) another 5000 steps of
702 SD and 5000 steps of ABNR minimization. All non-hydrogen atoms were then restrained
703 using harmonic restraints with a force constant of 1.0 kcal/mol/Å² and SHAKE
704 constraints (99) were applied on all hydrogen atoms, and 5000 steps of SD and 5000
705 steps of ABNR minimization was performed to obtain the final *Cne* Prp8 precursor
706 model.

707

708 **ACKNOWLEDGEMENTS**

709 The authors acknowledge Dr. Qishan Lin (UAlbany) for help with mass
710 spectrometry experiments. Fungal genomic DNA was donated by Dr. Sudha Chaturvedi
711 (Wadsworth Center), Dr. Robert Cramer (Dartmouth), Dr. Timothy James (University of
712 Michigan), and Dr. Chad Rappleye (Ohio State). We also wish to thank Dr. Tzanko I.
713 Doukov at the Stanford Synchrotron Radiation Lightsource (SSRL) for help in X-ray
714 diffraction data collection. Use of SSRL, SLAC National Accelerator Laboratory, is
715 supported by the U.S. Department of Energy (DOE) under Contract No. DE-AC02-
716 76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE
717 and by the National Institutes of Health (NIH), including P41GM103393. This work was
718 supported by the NIH grants GM39422 and GM44844 to M.B.

719

720 The atomic coordinates and structure factors (PDB 6MX6) have been deposited in
721 the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers
722 University, New Brunswick, NJ (<http://www.rcsb.org/>).

723

724 **REFERENCES**

- 725 1. Grainger RJ, Beggs JD. Prp8 protein: at the heart of the spliceosome. *Rna*. 2005
726 May;11(5):533-57. PubMed PMID: 15840809. Pubmed Central PMCID: 1370742.
- 727 2. Mayerle M, Guthrie C. Prp8 retinitis pigmentosa mutants cause defects in the
728 transition between the catalytic steps of splicing. *Rna*. 2016 May;22(5):793-809. PubMed
729 PMID: 26968627. Pubmed Central PMCID: 4836653.
- 730 3. Daiger SP, Sullivan LS, Bowne SJ. Genes and mutations causing retinitis
731 pigmentosa. *Clinical genetics*. 2013 Aug;84(2):132-41. PubMed PMID: 23701314.
732 Pubmed Central PMCID: 3856531.
- 733 4. Qu G, Kaushal PS, Wang J, Shigematsu H, Piazza CL, Agrawal RK, et al.
734 Structure of a group II intron in complex with its reverse transcriptase. *Nature structural*
735 *& molecular biology*. 2016 Jun;23(6):549-57. PubMed PMID: 27136327. Pubmed
736 Central PMCID: 4899178.
- 737 5. Galej WP, Oubridge C, Newman AJ, Nagai K. Crystal structure of Prp8 reveals
738 active site cavity of the spliceosome. *Nature*. 2013 Jan 31;493(7434):638-43. PubMed
739 PMID: 23354046. Pubmed Central PMCID: 3672837.
- 740 6. Dlagic M, Mushegian A. Prp8, the pivotal protein of the spliceosomal catalytic
741 center, evolved from a retroelement-encoded reverse transcriptase. *Rna*. 2011
742 May;17(5):799-808. PubMed PMID: 21441348. Pubmed Central PMCID: 3078730.

- 743 7. Zhan X, Yan C, Zhang X, Lei J, Shi Y. Structure of a human catalytic step I
744 spliceosome. *Science*. 2018 Feb 2;359(6375):537-45. PubMed PMID: 29301961.
- 745 8. Wilkinson ME, Fica SM, Galej WP, Norman CM, Newman AJ, Nagai K.
746 Postcatalytic spliceosome structure reveals mechanism of 3'-splice site selection. *Science*.
747 2017 Dec 8;358(6368):1283-8. PubMed PMID: 29146871. Pubmed Central PMCID:
748 5808836.
- 749 9. Bai R, Wan R, Yan C, Lei J, Shi Y. Structures of the fully assembled
750 *Saccharomyces cerevisiae* spliceosome before activation. *Science*. 2018 Jun
751 29;360(6396):1423-9. PubMed PMID: 29794219.
- 752 10. Wan R, Yan C, Bai R, Huang G, Shi Y. Structure of a yeast catalytic step I
753 spliceosome at 3.4 Å resolution. *Science*. 2016 Aug 26;353(6302):895-904. PubMed
754 PMID: 27445308.
- 755 11. Wan R, Yan C, Bai R, Wang L, Huang M, Wong CC, et al. The 3.8 Å structure of
756 the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science*.
757 2016 Jan 29;351(6272):466-75. PubMed PMID: 26743623.
- 758 12. Yan C, Hang J, Wan R, Huang M, Wong CC, Shi Y. Structure of a yeast
759 spliceosome at 3.6-angstrom resolution. *Science*. 2015 Sep 11;349(6253):1182-91.
760 PubMed PMID: 26292707.
- 761 13. Saleh L, Perler FB. Protein splicing in cis and in trans. *Chemical record*.
762 2006;6(4):183-93. PubMed PMID: 16900466.
- 763 14. Novikova O, Topilina N, Belfort M. Enigmatic distribution, evolution, and
764 function of inteins. *The Journal of biological chemistry*. 2014 May 23;289(21):14490-7.
765 PubMed PMID: 24695741. Pubmed Central PMCID: 4031506.

- 766 15. Perler FB, Olsen GJ, Adam E. Compilation and analysis of intein sequences.
767 Nucleic acids research. 1997 Mar 15;25(6):1087-93. PubMed PMID: 9092614. Pubmed
768 Central PMCID: 146560.
- 769 16. Gimble FS, Thorner J. Homing of a DNA endonuclease gene by meiotic gene
770 conversion in *Saccharomyces cerevisiae*. Nature. 1992 May 28;357(6376):301-6.
771 PubMed PMID: 1534148.
- 772 17. Liu XQ. Protein-splicing intein: Genetic mobility, origin, and evolution. Annual
773 review of genetics. 2000;34:61-76. PubMed PMID: 11092822.
- 774 18. Lennon CW, Stanger M, Banavali NK, Belfort M. Conditional Protein Splicing
775 Switch in Hyperthermophiles through an Intein-Extein Partnership. mBio. 2018 Jan
776 30;9(1). PubMed PMID: 29382734. Pubmed Central PMCID: 5790916.
- 777 19. Lennon CW, Stanger M, Belfort M. Protein splicing of a recombinase intein
778 induced by ssDNA and DNA damage. Genes Dev. 2016 Dec 15;30(24):2663-8. PubMed
779 PMID: 28031248. Pubmed Central PMCID: 5238726. Epub 2016/12/30. eng.
- 780 20. Kelley DS, Lennon CW, Sea P, Belfort M, Novikova O. Mycobacteriophages as
781 Incubators for Intein Dissemination and Evolution. mBio. 2016 Oct 04;7(5). PubMed
782 PMID: 27703073. Pubmed Central PMCID: 5050341.
- 783 21. Topilina NI, Green CM, Jayachandran P, Kelley DS, Stanger MJ, Piazza CL, et
784 al. SufB intein of *Mycobacterium tuberculosis* as a sensor for oxidative and nitrosative
785 stresses. Proc Natl Acad Sci U S A. 2015 Aug 18;112(33):10348-53. PubMed PMID:
786 26240361. Epub 2015/08/05. eng.

- 787 22. Nicastrì MC, Xega K, Li L, Xie J, Wang C, Linhardt RJ, et al. Internal disulfide
788 bond acts as a switch for intein activity. *Biochemistry*. 2013 Aug 27;52(34):5920-7.
789 PubMed PMID: 23906287. Pubmed Central PMCID: 3801215. Epub 2013/08/03. eng.
- 790 23. Reitter JN, Cousin CE, Nicastrì MC, Jaramillo MV, Mills KV. Salt-dependent
791 conditional protein splicing of an intein from *Halobacterium salinarum*. *Biochemistry*.
792 2016 Mar 8;55(9):1279-82. PubMed PMID: 26913597. Epub 2016/02/26. eng.
- 793 24. Topilina NI, Novikova O, Stanger M, Banavali NK, Belfort M. Post-translational
794 environmental switch of RadA activity by extein-intein interactions in protein splicing.
795 *Nucleic Acids Res*. 2015 Jun 22;43(13):6631-48. PubMed PMID: 26101259. Epub
796 2015/06/24. Eng.
- 797 25. Kelley DS, Lennon, C. W., Li, Z., Miller, M. R., Banavali, N. K., Li, H., and
798 Belfort, M. Mycobacterial DnaB helicase intein as oxidative stress sensor. *Nature*
799 *communications*. 2018.
- 800 26. Novikova O, Jayachandran P, Kelley DS, Morton Z, Merwin S, Topilina NI, et al.
801 Intein clustering suggests functional importance in different domains of life. *Molecular*
802 *biology and evolution*. 2016 Mar;33(3):783-99. PubMed PMID: 26609079. Pubmed
803 Central PMCID: 4760082. Epub 2015/11/27. eng.
- 804 27. Green CM, Novikova O, Belfort M. The dynamic intein landscape of eukaryotes.
805 *Mobile DNA*. 2018;9:4. PubMed PMID: 29416568. Pubmed Central PMCID: 5784728.
- 806 28. Theodoro RC, Bagagli E. Inteins in pathogenic fungi: a phylogenetic tool and
807 perspectives for therapeutic applications. *Memorias do Instituto Oswaldo Cruz*. 2009
808 May;104(3):497-504. PubMed PMID: 19547879.

- 809 29. Janbon G. Introns in *Cryptococcus*. Memorias do Instituto Oswaldo Cruz.
810 2018;113(7):e170519. PubMed PMID: 29513783. Pubmed Central PMCID: 5851062.
- 811 30. Galej WP, Nguyen TH, Newman AJ, Nagai K. Structural studies of the
812 spliceosome: zooming into the heart of the machine. Current opinion in structural
813 biology. 2014 Apr;25:57-66. PubMed PMID: 24480332. Pubmed Central PMCID:
814 4045393.
- 815 31. Elleuche SaP, S. Fungal inteins: distribution, evolution, and applications. In:
816 Anke T, Schuffler, A., editor. The Mycota. 152018. p. 57-85.
- 817 32. Stajich JE, Berbee ML, Blackwell M, Hibbett DS, James TY, Spatafora JW, et al.
818 The fungi. Current biology : CB. 2009 Sep 29;19(18):R840-5. PubMed PMID:
819 19788875. Pubmed Central PMCID: 2913116.
- 820 33. Poulter RT, Goodwin TJ, Butler MI. The nuclear-encoded inteins of fungi. Fungal
821 genetics and biology : FG & B. 2007 Mar;44(3):153-79. PubMed PMID: 17046294.
- 822 34. Pietrokovski S. Modular organization of inteins and C-terminal autocatalytic
823 domains. Protein science : a publication of the Protein Society. 1998 Jan;7(1):64-71.
824 PubMed PMID: 9514260. Pubmed Central PMCID: 2143824.
- 825 35. Butler MI, Gray J, Goodwin TJ, Poulter RT. The distribution and evolutionary
826 history of the PRP8 intein. BMC evolutionary biology. 2006 May 31;6:42. PubMed
827 PMID: 16737526. Pubmed Central PMCID: 1508164.
- 828 36. Monier A, Sudek S, Fast NM, Worden AZ. Gene invasion in distant eukaryotic
829 lineages: discovery of mutually exclusive genetic elements reveals marine biodiversity.
830 The ISME journal. 2013 Sep;7(9):1764-74. PubMed PMID: 23635865. Pubmed Central
831 PMCID: 3749507.

- 832 37. Urushihara H, Kuwayama H, Fukuhara K, Itoh T, Kagoshima H, Shin IT, et al.
833 Comparative genome and transcriptome analyses of the social amoeba *Acytostelium*
834 *subglobosum* that accomplishes multicellular development without germ-soma
835 differentiation. BMC genomics. 2015 Feb 14;16:80. PubMed PMID: 25758444. Pubmed
836 Central PMCID: 4334915.
- 837 38. Volkmann G, Mootz HD. Recent progress in intein research: from mechanism to
838 directed evolution and applications. Cellular and molecular life sciences : CMLS. 2013
839 Apr;70(7):1185-206. PubMed PMID: 22926412.
- 840 39. Werner E, Wende W, Pingoud A, Heinemann U. High resolution crystal structure
841 of domain I of the *Saccharomyces cerevisiae* homing endonuclease PI-SceI. Nucleic
842 acids research. 2002 Sep 15;30(18):3962-71. PubMed PMID: 12235380. Pubmed Central
843 PMCID: 137108.
- 844 40. Moure CM, Gimble FS, Quijcho FA. Crystal structure of the intein homing
845 endonuclease PI-SceI bound to its recognition sequence. Nature structural biology. 2002
846 Oct;9(10):764-70. PubMed PMID: 12219083.
- 847 41. Dearden AK, Callahan B, Roey PV, Li Z, Kumar U, Belfort M, et al. A conserved
848 threonine spring-loads precursor for intein splicing. Protein science : a publication of the
849 Protein Society. 2013 May;22(5):557-63. PubMed PMID: 23423655. Pubmed Central
850 PMCID: 3649257.
- 851 42. Mills KV, Johnson MA, Perler FB. Protein splicing: how inteins escape from
852 precursor proteins. J Biol Chem. 2014 May 23;289(21):14498-505. PubMed PMID:
853 24695729. Pubmed Central PMCID: 4031507. Epub 2014/04/04. eng.

- 854 43. Hall TM, Porter JA, Young KE, Koonin EV, Beachy PA, Leahy DJ. Crystal
855 structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-
856 splicing proteins. *Cell*. 1997 Oct 3;91(1):85-97. PubMed PMID: 9335337.
- 857 44. Burglin TR. The Hedgehog protein family. *Genome biology*. 2008;9(11):241.
858 PubMed PMID: 19040769. Pubmed Central PMCID: 2614485.
- 859 45. Perler FB. Protein splicing of inteins and hedgehog autoproteolysis: structure,
860 function, and evolution. *Cell*. 1998 Jan 9;92(1):1-4. PubMed PMID: 9489693.
- 861 46. Roelink H. The Hedgehog signaling domain was acquired from a prokaryote.
862 bioRxiv2018.
- 863 47. Van Roey P, Pereira B, Li Z, Hiraga K, Belfort M, Derbyshire V.
864 Crystallographic and mutational studies of *Mycobacterium tuberculosis* recA mini-inteins
865 suggest a pivotal role for a highly conserved aspartate residue. *Journal of molecular*
866 *biology*. 2007 Mar 16;367(1):162-73. PubMed PMID: 17254599. Pubmed Central
867 PMCID: 1852430.
- 868 48. Amitai G, Callahan BP, Stanger MJ, Belfort G, Belfort M. Modulation of intein
869 activity by its neighboring extein substrates. *Proceedings of the National Academy of*
870 *Sciences of the United States of America*. 2009 Jul 7;106(27):11005-10. PubMed PMID:
871 19541659. Pubmed Central PMCID: 2708771.
- 872 49. Weiss G, Schaible UE. Macrophage defense mechanisms against intracellular
873 bacteria. *Immunological reviews*. 2015 Mar;264(1):182-203. PubMed PMID: 25703560.
874 Pubmed Central PMCID: 4368383.

- 875 50. Missall TA, Lodge JK, McEwen JE. Mechanisms of resistance to oxidative and
876 nitrosative stress: implications for fungal survival in mammalian hosts. *Eukaryotic cell*.
877 2004 Aug;3(4):835-46. PubMed PMID: 15302816. Pubmed Central PMCID: 500878.
- 878 51. Chan H, Pearson CS, Green CM, Li Z, Zhang J, Belfort G, et al. Exploring Intein
879 Inhibition by Platinum Compounds as an Antimicrobial Strategy. *J Biol Chem*. 2016 Oct
880 21;291(43):22661-70. PubMed PMID: 27609519. Pubmed Central PMCID: 5077202.
881 Epub 2016/09/10. eng.
- 882 52. Zhang L, Xiao N, Pan Y, Zheng Y, Pan Z, Luo Z, et al. Binding and inhibition of
883 copper ions to RecA inteins from *Mycobacterium tuberculosis*. *Chemistry*. 2010 Apr
884 12;16(14):4297-306. PubMed PMID: 20209535.
- 885 53. Zhang L, Zheng Y, Xi Z, Luo Z, Xu X, Wang C, et al. Metal ions binding to recA
886 inteins from *Mycobacterium tuberculosis*. *Molecular bioSystems*. 2009 Jun;5(6):644-50.
887 PubMed PMID: 19462022. Pubmed Central PMCID: 2790073.
- 888 54. Sun P, Ye S, Ferrandon S, Evans TC, Xu MQ, Rao Z. Crystal structures of an
889 intein from the split dnaE gene of *Synechocystis* sp. PCC6803 reveal the catalytic model
890 without the penultimate histidine and the mechanism of zinc ion inhibition of protein
891 splicing. *Journal of molecular biology*. 2005 Nov 11;353(5):1093-105. PubMed PMID:
892 16219320.
- 893 55. Mills KV, Paulus H. Reversible inhibition of protein splicing by zinc ion. *The*
894 *Journal of biological chemistry*. 2001 Apr 6;276(14):10832-8. PubMed PMID: 11152694.
- 895 56. Nichols NM, Benner JS, Martin DD, Evans TC, Jr. Zinc ion effects on individual
896 Ssp DnaE intein splicing steps: regulating pathway progression. *Biochemistry*. 2003 May
897 13;42(18):5301-11. PubMed PMID: 12731871.

- 898 57. Giles NM, Watts AB, Giles GI, Fry FH, Littlechild JA, Jacob C. Metal and redox
899 modulation of cysteine protein function. *Chemistry & biology*. 2003 Aug;10(8):677-93.
900 PubMed PMID: 12954327.
- 901 58. Kelley DS, Lennon CW, Li Z, Miller MR, Banavali NK, Li H, et al.
902 Mycobacterial DnaB helicase intein as oxidative stress sensor. *Nature communications*.
903 2018 Oct 19;9(1):4363. PubMed PMID: 30341292.
- 904 59. Callahan BP, Topilina NI, Stanger MJ, Van Roey P, Belfort M. Structure of
905 catalytically competent intein caught in a redox trap with functional and evolutionary
906 implications. *Nature structural & molecular biology*. 2011 May;18(5):630-3. PubMed
907 PMID: 21460844. Pubmed Central PMCID: 3087850.
- 908 60. McCall KA, Huang C, Fierke CA. Function and mechanism of zinc
909 metalloenzymes. *The Journal of nutrition*. 2000 May;130(5S Suppl):1437S-46S. PubMed
910 PMID: 10801957.
- 911 61. Casalino L, Palermo G, Spinello A, Rothlisberger U, Magistrato A. All-atom
912 simulations disentangle the functional dynamics underlying gene maturation in the intron
913 lariat spliceosome. *Proceedings of the National Academy of Sciences of the United States*
914 *of America*. 2018 Jun 26;115(26):6584-9. PubMed PMID: 29891649. Pubmed Central
915 PMCID: 6042132.
- 916 62. Walker JE, Saraste M, Runswick MJ, Gay NJ. Distantly related sequences in the
917 alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring
918 enzymes and a common nucleotide binding fold. *The EMBO journal*. 1982;1(8):945-51.
919 PubMed PMID: 6329717. Pubmed Central PMCID: 553140.

- 920 63. Aranko AS, Wlodawer A, Iwai H. Nature's recipe for splitting inteins. Protein
921 engineering, design & selection : PEDS. 2014 Aug;27(8):263-71. PubMed PMID:
922 25096198. Pubmed Central PMCID: 4133565.
- 923 64. Zhang L, Zheng Y, Callahan B, Belfort M, Liu Y. Cisplatin inhibits protein
924 splicing, suggesting inteins as therapeutic targets in mycobacteria. J Biol Chem. 2011 Jan
925 14;286(2):1277-82. PubMed PMID: 21059649. Pubmed Central PMCID: 3020735. Epub
926 2010/11/10. eng.
- 927 65. Paulus H. Inteins as targets for potential antimycobacterial drugs. Frontiers in
928 bioscience : a journal and virtual library. 2003 Sep 1;8:s1157-65. PubMed PMID:
929 12957838.
- 930 66. Fairlamb AH, Gow NA, Matthews KR, Waters AP. Drug resistance in eukaryotic
931 microorganisms. Nature microbiology. 2016 Jun 24;1(7):16092. PubMed PMID:
932 27572976. Pubmed Central PMCID: 5215055.
- 933 67. Ballou ER, Wilson D. The roles of zinc and copper sensing in fungal
934 pathogenesis. Current opinion in microbiology. 2016 Aug;32:128-34. PubMed PMID:
935 27327380. Pubmed Central PMCID: 4992176.
- 936 68. Alspaugh JA. Virulence mechanisms and *Cryptococcus neoformans* pathogenesis.
937 Fungal genetics and biology : FG & B. 2015 May;78:55-8. PubMed PMID: 25256589.
938 Pubmed Central PMCID: 4370805.
- 939 69. Watkins RA, King JS, Johnston SA. Nutritional Requirements and Their
940 Importance for Virulence of Pathogenic *Cryptococcus* Species. Microorganisms. 2017
941 Sep 30;5(4). PubMed PMID: 28974017. Pubmed Central PMCID: 5748574.

- 942 70. Mitchell TG, Perfect JR. Cryptococcosis in the era of AIDS--100 years after the
943 discovery of *Cryptococcus neoformans*. *Clinical microbiology reviews*. 1995
944 Oct;8(4):515-48. PubMed PMID: 8665468. Pubmed Central PMCID: 172874.
- 945 71. Smith AD, Logeman BL, Thiele DJ. Copper Acquisition and Utilization in Fungi.
946 *Annual review of microbiology*. 2017 Sep 8;71:597-623. PubMed PMID: 28886682.
- 947 72. Hodgkinson V, Petris MJ. Copper homeostasis at the host-pathogen interface. *The*
948 *Journal of biological chemistry*. 2012 Apr 20;287(17):13549-55. PubMed PMID:
949 22389498. Pubmed Central PMCID: 3340201.
- 950 73. Dos Santos FM, Piffer AC, Schneider RO, Ribeiro NS, Garcia AWA, Schrank A,
951 et al. Alterations of zinc homeostasis in response to *Cryptococcus neoformans* in a
952 murine macrophage cell line. *Future microbiology*. 2017 May;12:491-504. PubMed
953 PMID: 28492340.
- 954 74. Garcia-Santamarina S, Thiele DJ. Copper at the Fungal Pathogen-Host Axis. *The*
955 *Journal of biological chemistry*. 2015 Jul 31;290(31):18945-53. PubMed PMID:
956 26055724. Pubmed Central PMCID: 4521015.
- 957 75. Ding C, Festa RA, Sun TS, Wang ZY. Iron and copper as virulence modulators in
958 human fungal pathogens. *Molecular microbiology*. 2014 Jul;93(1):10-23. PubMed PMID:
959 24851950.
- 960 76. Bergkessel M, Whitworth GB, Guthrie C. Diverse environmental stresses elicit
961 distinct responses at the level of pre-mRNA processing in yeast. *Rna*. 2011
962 Aug;17(8):1461-78. PubMed PMID: 21697354. Pubmed Central PMCID: 3153971.
- 963 77. Gonzalez-Hilarion S, Paulet D, Lee KT, Hon CC, Lechat P, Mogensen E, et al.
964 Intron retention-dependent gene regulation in *Cryptococcus neoformans*. *Scientific*

- 965 reports. 2016 Aug 31;6:32252. PubMed PMID: 27577684. Pubmed Central PMCID:
966 5006051.
- 967 78. Papasaikas P, Valcarcel J. The Spliceosome: The Ultimate RNA Chaperone and
968 Sculptor. Trends in biochemical sciences. 2016 Jan;41(1):33-45. PubMed PMID:
969 26682498.
- 970 79. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable
971 generation of high-quality protein multiple sequence alignments using Clustal Omega.
972 Molecular systems biology. 2011 Oct 11;7:539. PubMed PMID: 21988835. Pubmed
973 Central PMCID: 3261699.
- 974 80. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics
975 Analysis Version 7.0 for Bigger Datasets. Molecular biology and evolution. 2016
976 Jul;33(7):1870-4. PubMed PMID: 27004904.
- 977 81. Sitnikova T, Rzhetsky A, Nei M. Interior-branch and bootstrap tests of
978 phylogenetic trees. Molecular biology and evolution. 1995 Mar;12(2):319-33. PubMed
979 PMID: 7700156.
- 980 82. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo
981 generator. Genome research. 2004 Jun;14(6):1188-90. PubMed PMID: 15173120.
982 Pubmed Central PMCID: 419797.
- 983 83. Li Z, Brecher M, Deng YQ, Zhang J, Sakamuru S, Liu B, et al. Existing drugs as
984 broad-spectrum and potent inhibitors for Zika virus by targeting NS2B-NS3 interaction.
985 Cell research. 2017 Aug;27(8):1046-64. PubMed PMID: 28685770. Pubmed Central
986 PMCID: 5539352.

- 987 84. Unwin RD, Griffiths JR, Whetton AD. A sensitive mass spectrometric method for
988 hypothesis-driven detection of peptide post-translational modifications: multiple reaction
989 monitoring-initiated detection and sequencing (MIDAS). *Nature protocols*.
990 2009;4(6):870-7. PubMed PMID: 19444244.
- 991 85. Otwinowski Z, Minor W. [20] Processing of X-ray diffraction data collected in
992 oscillation mode. *Methods in enzymology*. 1997;276:307-26. PubMed PMID: 27799103.
- 993 86. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al.
994 PHENIX: a comprehensive Python-based system for macromolecular structure solution.
995 *Acta crystallographica Section D, Biological crystallography*. 2010 Feb;66(Pt 2):213-21.
996 PubMed PMID: 20124702. Pubmed Central PMCID: 2815670.
- 997 87. Munoz-Escobar J, Matta-Camacho E, Cho C, Kozlov G, Gehring K. Bound
998 Waters Mediate Binding of Diverse Substrates to a Ubiquitin Ligase. *Structure*. 2017
999 May 2;25(5):719-29 e3. PubMed PMID: 28392261.
- 1000 88. Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B,
1001 et al. CHARMM: the biomolecular simulation program. *Journal of computational*
1002 *chemistry*. 2009 Jul 30;30(10):1545-614. PubMed PMID: 19444816. Pubmed Central
1003 PMCID: 2810661.
- 1004 89. Brooks B, Bruccoleri, R., Olafson, B., Swaminathan, S., Karplus, M. CHARMM:
1005 A program for macromolecular energy, minimization, and dynamics calculations. *Journal*
1006 *of computational chemistry*. 1983 (4):187-217.
- 1007 90. Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, et al. Optimization of the
1008 additive CHARMM all-atom protein force field targeting improved sampling of the
1009 backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *Journal of chemical*

- 1010 theory and computation. 2012 Sep 11;8(9):3257-73. PubMed PMID: 23341755. Pubmed
1011 Central PMCID: 3549273.
- 1012 91. Kabsch W. A discussion of the solution for the best rotation to relate two sets of
1013 vectors. Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and
1014 General Crystallography. 1978 (34):827-8.
- 1015 92. Curry H. The method of steepest descent for nonlinear minimization problems.
1016 Quarterly of Applied Mathematics. 1944 (2):250-61.
- 1017 93. Needleman SB, Wunsch, C. D. A general method applicable to the search for
1018 similarities in the amino acid sequence of two proteins. Journal of molecular biology.
1019 1970;48:443-53.
- 1020 94. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open
1021 Software Suite. Trends in genetics : TIG. 2000 Jun;16(6):276-7. PubMed PMID:
1022 10827456.
- 1023 95. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial
1024 restraints. Journal of molecular biology. 1993 Dec 5;234(3):779-815. PubMed PMID:
1025 8254673.
- 1026 96. Shen MY, Sali A. Statistical potential for assessment and prediction of protein
1027 structures. Protein science : a publication of the Protein Society. 2006 Nov;15(11):2507-
1028 24. PubMed PMID: 17075131. Pubmed Central PMCID: 2242414.
- 1029 97. Melo F, Sanchez R, Sali A. Statistical potentials for fold assessment. Protein
1030 science : a publication of the Protein Society. 2002 Feb;11(2):430-48. PubMed PMID:
1031 11790853. Pubmed Central PMCID: 2373452.

- 1032 98. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. Journal of
1033 molecular graphics. 1996 Feb;14(1):33-8, 27-8. PubMed PMID: 8744570.
- 1034 99. Ryckaert J-P, Ciccotti, G., Berendsen, H. J. Numerical integration of the cartesian
1035 equations of motion of a system with constraints: molecular dynamics of n-alkanes.
1036 Journal of Computational Physics. 1997;23:327-41.
- 1037

1038 **TABLES**

1039 **Table 1.** Metal binding K_d estimates using ITC

<i>Cne Prp8</i> intein	Metal	Approximate K_d
Wild-type	Cu^{2+}	>100 μM
Wild-type	Zn^{2+}	~1 nM
C1A	Zn^{2+}	>30 μM
C61A	Zn^{2+}	>570 nM
H63A	Zn^{2+}	>5 μM
H65A	Zn^{2+}	>30 μM
C1A/C61A/H65A	Zn^{2+}	>100 μM

1040

1041 **FIGURE LEGENDS**

1042 **Figure 1.** Prp8 is an intein hot-spot with multiple, independent insertion sites. **A.**
1043 Modified phylogenetic tree of eukaryotes (left) shows the phyla that contain Prp8 inteins,
1044 with representative genera listed. Evolutionary divergence times are denoted in millions
1045 of years ago (Mya). The number of Prp8 inteins in each phylum is shown in superscript,
1046 and insertion sites are shown on the branches (colored **a-g**). A phylogenetic intein tree
1047 (right) was reconstructed based on the amino acid sequences of intein splicing blocks for
1048 a subset of 50 Prp8 inteins. The radial tree shows numerous clusters, which correspond to
1049 grouping by insertion site. Abbreviated species names are shown (full names in Fig.
1050 S1A). Shading (dark gray, light gray, or white) indicates phylogenetic distribution. The
1051 divergence of the inteins despite extein conservation suggests independent invasions. **B.**
1052 A line diagram of Prp8 exteins shows the domains of the host protein from amino acid
1053 residues 127 to 2084. The arrows below indicate the site of intein insertion (**a-g**) with the
1054 corresponding residue number based on *Saccharomyces cerevisiae* Prp8 (PDB 5GMK,
1055 chain A). Shapes represent how many inteins are found at each site (square = 10 inteins,
1056 triangle = 1 intein) and are shaded to denote phylogenetic origin as in Fig. 1A. Prp8-**a** is
1057 the most common insertion site with ~100 inteins. **C.** Multiple sequence alignment of the
1058 splicing blocks of Prp8 inteins from each insertion site. Comparative analysis of residues
1059 found in blocks A, B, F, and G reveals that Prp8 inteins occupying other insertions sites
1060 are substantially different from one another, indicating independent acquisition. Identical
1061 residues are critical to self-splicing. Triangles indicate residues of general interest and
1062 those shaded blue are of specific interest. Numbers correspond to the *Cryptococcus*

1063 *neoformans* (*Cne*) Prp8 intein. Shading is as follows: black – identical amino acid, dark
1064 gray – conserved amino acid, light gray – similar amino acid substitution.

1065

1066 **Figure 2.** Structure and features of the *Cne* Prp8 intein. **A.** A crystal structure of the *Cne*
1067 Prp8 intein from site **a** was solved to 1.75 Å resolution. This structure has the canonical
1068 horseshoe shape and resolved the four splicing blocks, indicated in cyan (block A), green
1069 (block B), gray (block F), and purple (block G) (Fig. 1C). The amino (NT) and carboxy
1070 termini (CT) are annotated. Some regions of the structure are unresolved (dashed lines),
1071 and likely represent remnants of the original endonuclease domain or linker sequences. **B.**
1072 Features of the active site. The catalytic center is shown, highlighting the first residue
1073 (C1) and the penultimate residues (H170 and terminal N171). The B block TxxH motif is
1074 shown in green with the T62 and H65 shown as sticks. These residues are critical to
1075 carrying out autocatalytic protein splicing. **C.** Sequence of the *Cne* Prp8 intein (residues
1076 1-171) overlaid with secondary structure features. Blocks distant in sequence fold close
1077 in 3D space to allow protein splicing to occur. Unresolved regions (dashed lines) are
1078 between blocks B and F. Arrows represent β -strands, rectangles are α -helices. Residues
1079 noted in Fig. 2B are highlighted.

1080

1081 **Figure 3.** Homology of the *Cne* Prp8 intein to eukaryotic self-splicing elements. **A.**
1082 Overlay of the *Cne* Prp8 intein (red) with the *Sce* VMA1 intein (cyan and gray, PDB
1083 1GPP) with an RMSD of 1.06 Å. Structural similarities are most pronounced in the
1084 splicing domain (blocks A, B, F, and G) of the *Sce* VMA1 intein, whereas the
1085 linker/endonuclease domain (gray) is where the *Cne* Prp8 intein did not resolve (dashed

1086 lines). The amino (NT) and carboxy termini (CT) are annotated. **B.** Multiple sequence
1087 alignment of *Cne* Prp8 intein, *Sce* VMA1 intein, and *Drosophila melanogaster* (*Dme*)
1088 HHc. Sequence comparison (residues 1-193) reveals significant differences across the
1089 three proteins, which are only 22.6% identical. Overlaying secondary structure shows
1090 that, despite sequence divergence, the proteins have high structural similarity. Shading is
1091 as follows: black – identical amino acid, dark gray – conserved amino acid, light gray –
1092 similar amino acid substitution. Arrows represent β -strands, rectangles are α -helices.
1093 Unresolved regions shown as dashed lines. **C.** Overlay of the *Cne* Prp8 intein (red) with
1094 *Dme* HHc (yellow, PDB 1AT0). Structural 3D BLAST shows parallels between the
1095 eukaryotic intein and the eukaryotic self-splicing Hedgehog domain, with an RMSD of
1096 1.88 Å. The amino (NT) and carboxy termini (CT) are annotated.

1097

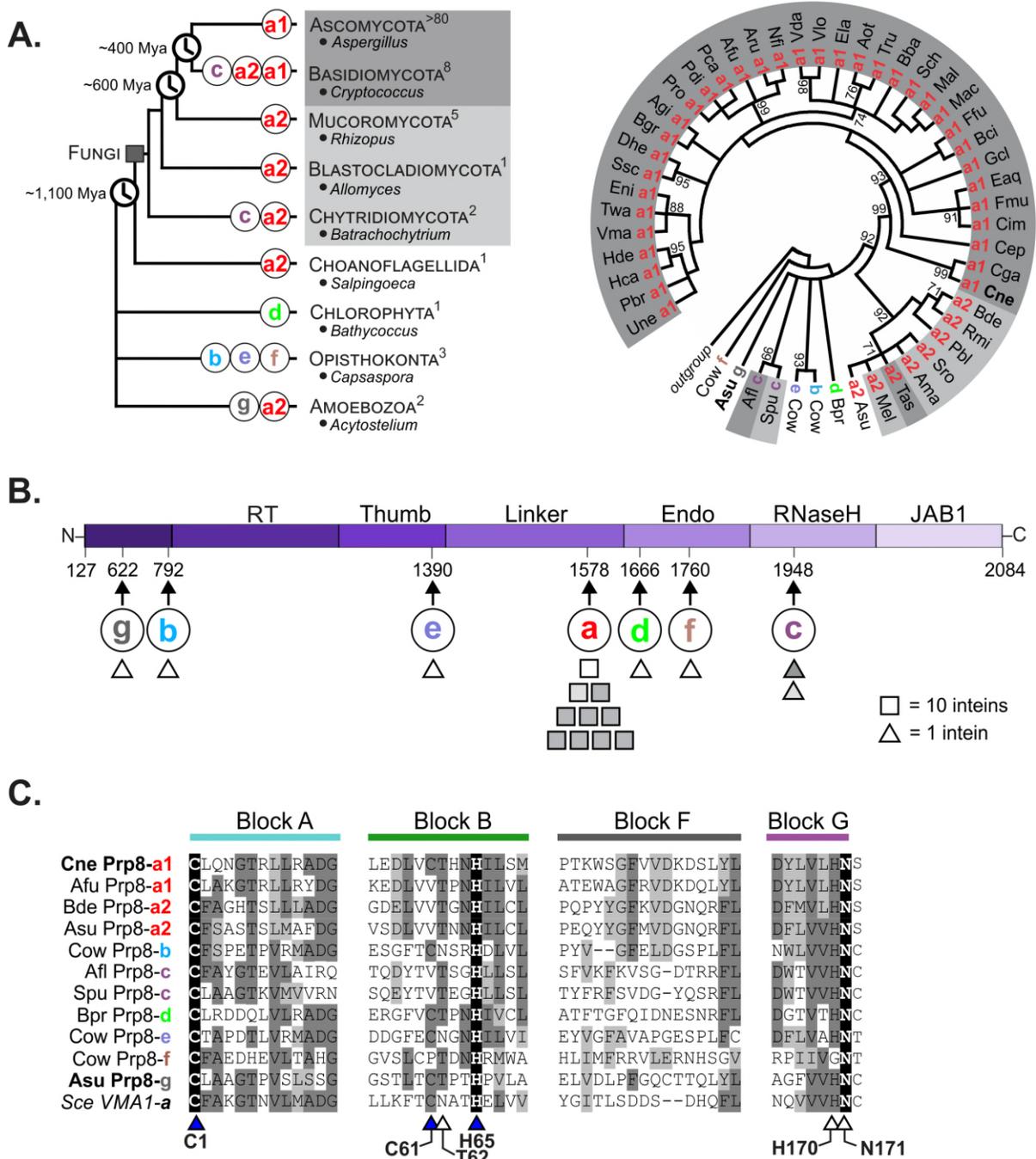
1098 **Figure 4.** MIG Prp8 A-1V is responsive to metal and RNS treatment. **A.** Schematic of the
1099 MIG reporter. The construct contains MBP-Intein-GFP (MIG) and is expressed in
1100 *Escherichia coli*. GFP allows monitoring of splicing using in-gel fluorescence. The
1101 precursor (P) can undergo protein splicing (left), yielding ligated exteins (LE) and free
1102 intein (INT, not seen on gels), or may undergo off-pathway cleavage reactions (right),
1103 such as C-terminal cleavage (CTC). The catalytic cysteine, C1, is labeled. **B.** MIG Prp8
1104 WT splices rapidly. A fluorescent gel of a splicing time-course shows that the wild-type
1105 *Cne* Prp8 intein in MIG is entirely spliced by the start of the assay (left, WT, 0 h). The A-
1106 1V mutant had precursor at the assay start and spliced over time (right, A-1V, 0 h).
1107 Quantitation is shown below in stacked plots. **C.** MIG Prp8 A-1V accumulates precursor
1108 under RNS and metal treatment. After 5 h *in vitro* treatment with a panel of

1109 environmental stressors, there was an increase in P compared to the untreated (UT, red
1110 line) with the RNS compounds DEA NONOate (DEA) and Angeli's salt (AS), and the
1111 metals copper (Cu), and zinc (Zn). H₂O₂ and magnesium (Mg) show no effect.
1112 Quantitation is shown below in a stacked plot. Data are representative of three biological
1113 replicates and mean standard deviations are shown. Representative gels are shown.
1114
1115 **Figure 5.** MIG Prp8 A-1V is differentially inhibited by copper and zinc. **A.** Copper
1116 inhibition is reversible by reducing agent only. MIG Prp8 A-1V splicing was completely
1117 inhibited by copper treatment (Cu) over 22 h, as minimal loss in P or increase in LE
1118 occurred compared to the untreated control. The inhibition was unaffected by treatment
1119 with metal chelator EDTA (Cu + EDTA). Upon adding copper and then reducing agent
1120 TCEP (Cu + TCEP), splicing was restored and P converted into LE over time. Red
1121 arrows indicate splicing rescue. The splice products were quantitated and the percent
1122 precursor is plotted as a proxy for splicing inhibition. Representative gels are shown. **B.**
1123 Zinc treatment is reversible by EDTA only. MIG Prp8 A-1V splicing was strongly
1124 inhibited by zinc treatment (+ Zn) over 22 h compared to untreated lysates. The zinc-
1125 based inhibition reversed when treated with EDTA (Zn + EDTA) and splicing was
1126 observed at a rate comparable to the untreated samples. Red arrows indicate splicing
1127 rescue. When adding zinc and then reducing agent TCEP (Zn + TCEP), splicing was
1128 unaffected. Plots are as in A above. Data are representative of three biological replicates
1129 and mean standard deviations are shown. Trend lines are fit to show the decay curve.
1130 Representative gels are shown. **C.** Zinc binds to the *Cne* Prp8 intein tightly. Using
1131 isothermal titration calorimetry (ITC), 16 μ M purified *Cne* Prp8 intein was titrated with

1132 0.05 mM ZnSO₄ over 20 injections at 37°C and pH 7.0 on a Nano ITC. The binding
1133 isotherm (bottom) shows integrated heat per mole of ZnSO₄ as a function of the molar
1134 ratio of ZnSO₄ to the *Cne* Prp8 intein and a K_d ~1 nM was calculated (Table 1). The
1135 NanoAnalyze ITC software automatically discarded outlier data points. Experiment was
1136 performed in triplicate. **D.** A putative zinc-binding pocket in the *Cne* Prp8 intein. Zinc
1137 binding was modeled using a structure of zinc bound human ubiquitin ligase (PDB
1138 5TDA) (87) and the *Cne* Prp8 intein structure. A coordination site is proposed using C1,
1139 C61, H63 and H65 (shown as sticks), with one zinc ion binding, shown as a gray sphere.
1140

1141 **Figure 6.** Modeling the *Cne* Prp8 intein into Prp8 and docking into the spliceosome
1142 reveals unfavorable interactions. **A.** *Cne* Prp8 intein in Prp8 exteins. The *Cne* Prp8 intein
1143 structure (red) was modeled into a structure of *Schizosaccharomyces pombe* (PDB 3JB9,
1144 chain A) Prp8 exteins (lavender). The S+1 is shown as blue spheres to indicate the site
1145 where a peptide bond was broken to insert the intein. The C1 is shown as yellow spheres
1146 and specifies start of the intein. The Prp8 intein fits into a cleft of the Prp8 structure, in a
1147 region that is highly conserved and functionally important. **B.** The *Cne* Prp8 intein in the
1148 spliceosome. The intein-containing Prp8 was overlaid into a spliceosome structure (PDB
1149 3JB9). This revealed a direct clash of the intein with spliceosomal protein Cwf19 (top
1150 panel, orange). There are also possible clashes with the intein and the intron lariat (split
1151 pea) and U2 snRNA (teal). U6 snRNA is also in the vicinity (pink).

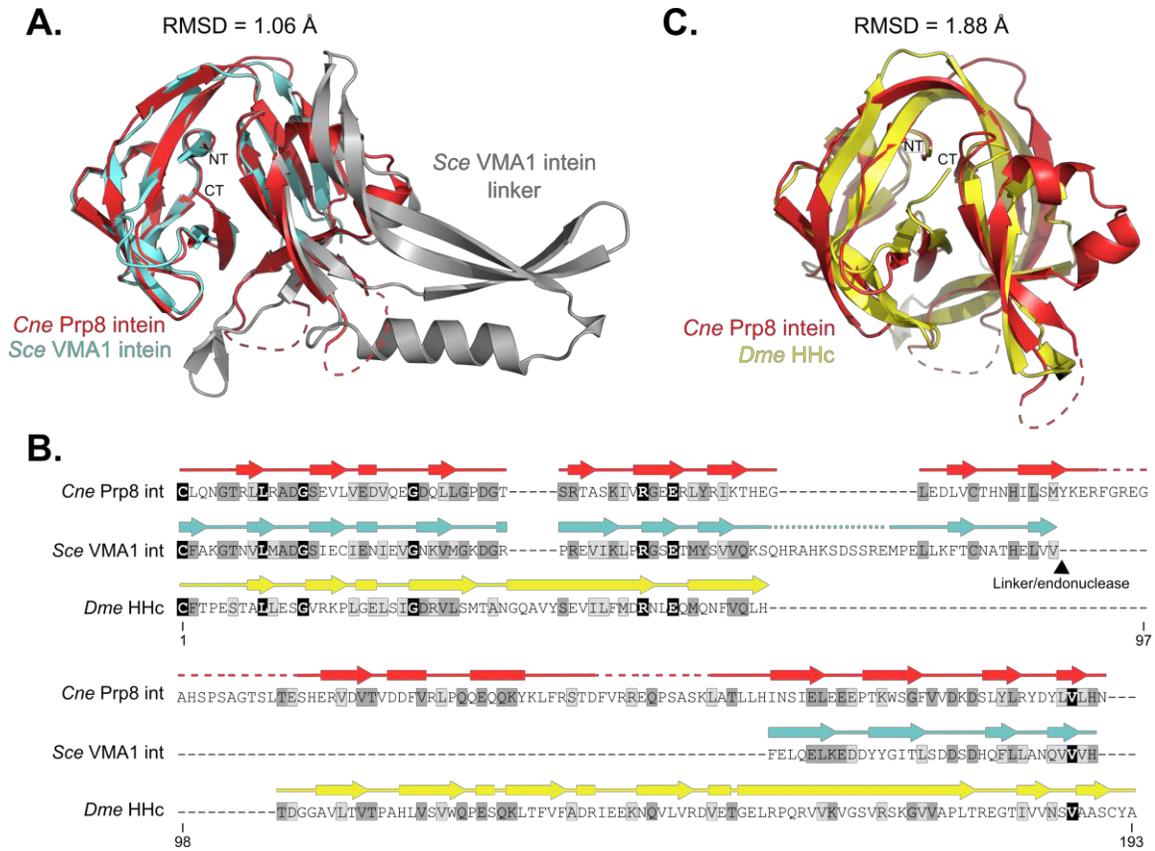
1152 FIGURES



1153

1154

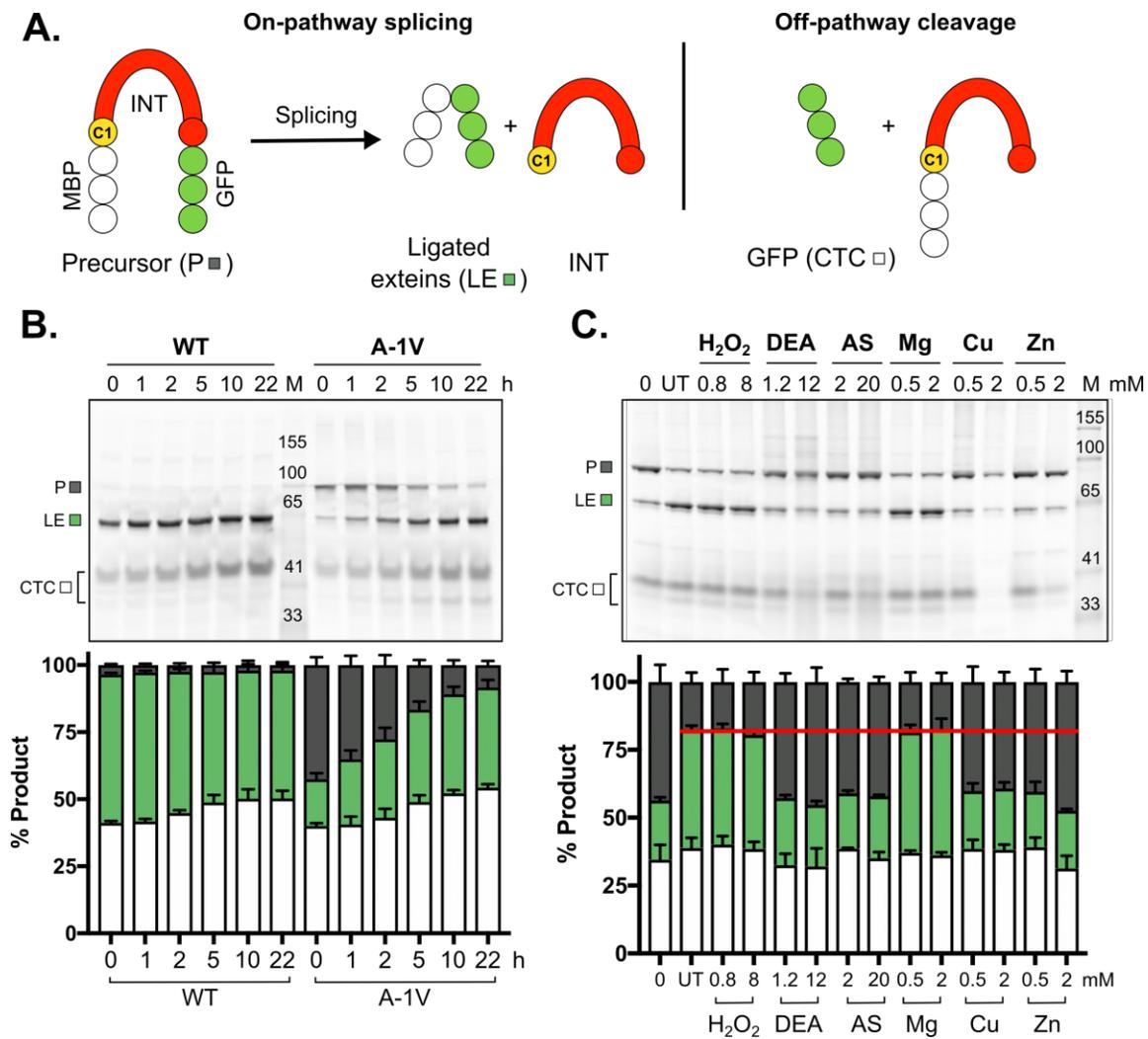
1155 **Figure 1.** Prp8 is an intein hot-spot with multiple, independent insertion sites.



1159

1160

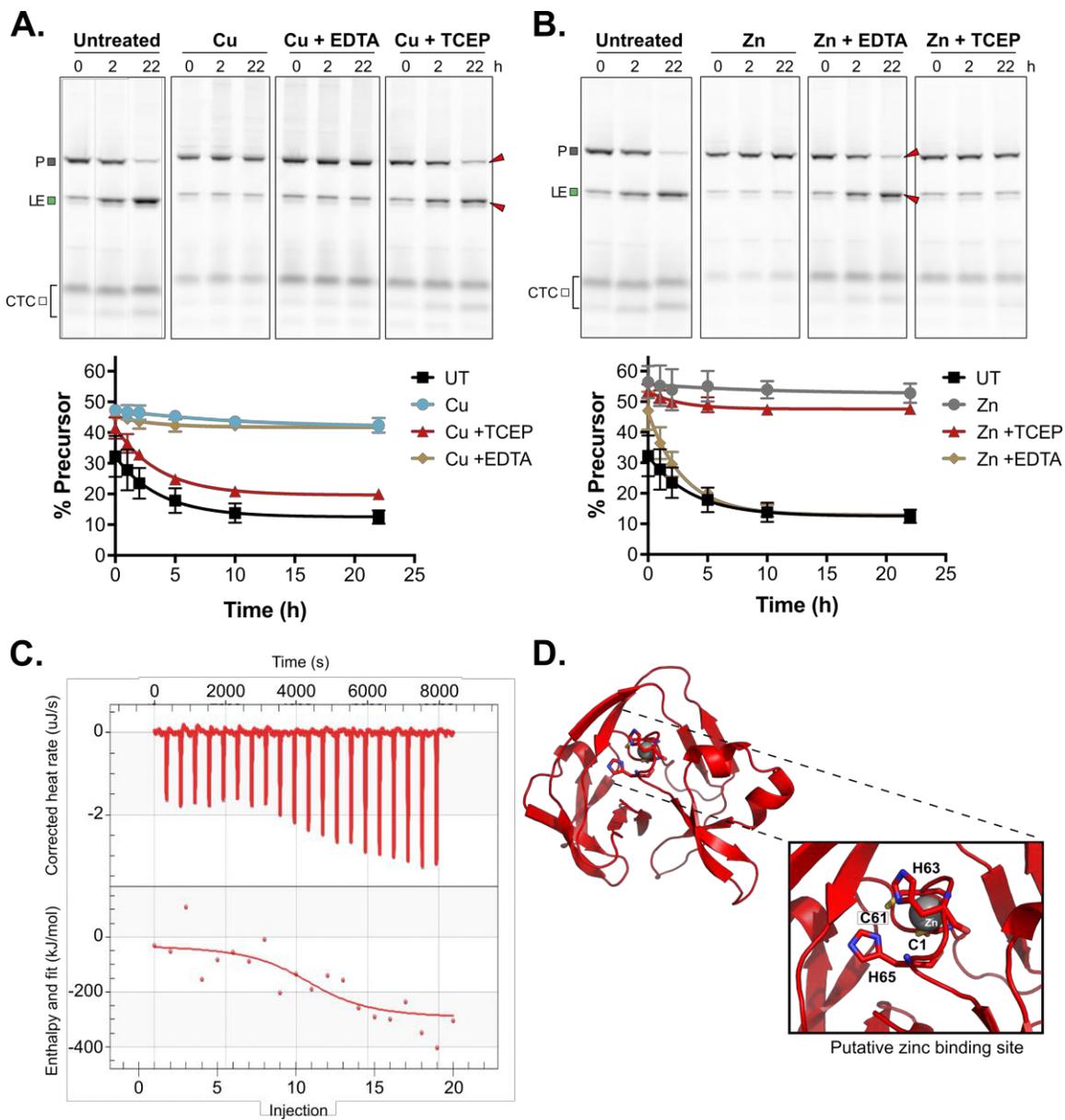
1161 **Figure 3.** Homology of the *Cne Prp8* intein to eukaryotic self-splicing elements.



1162

1163

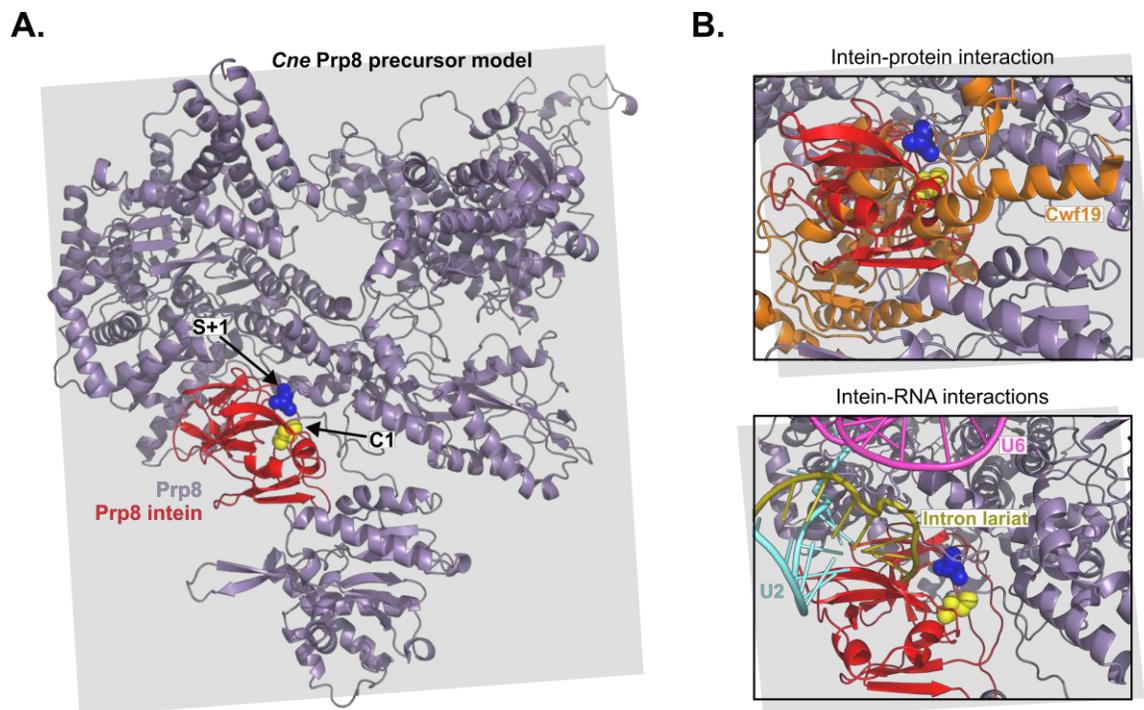
1164 **Figure 4. MIG Prp8 A-1V is responsive to metal and RNS treatment.**



1165

1166

1167 **Figure 5.** MIG Prp8 A-1V is differentially inhibited by copper and zinc.



1168

1169

1170 **Figure 6.** Modeling the *Cne* Prp8 intein into Prp8 and docking into the spliceosome

1171 reveals unfavorable interactions.

1172 **SUPPLEMENTAL INFORMATION**

1173

1174 **Supplemental Figures**

1175 **Figure S1.** Distribution of Prp8 inteins and novel insertion site **g**

1176 **Figure S2.** Amino acid multiple sequence alignment of Prp8 inteins utilized for
1177 phylogenetic analysis

1178 **Figure S3.** Overlays of the *Cne* Prp8 intein with other inteins

1179 **Figure S4.** Splicing of Prp8-a inteins from other fungal pathogens in MIG

1180 **Figure S5.** MIG Prp8 A-1V copper inhibition and cysteine analysis

1181 **Figure S6.** Copper inhibition of MIG Prp8 A-1V C61 mutants

1182 **Figure S7.** Mass spectrometry of cysteine modifications

1183 **Figure S8.** Putative zinc binding site

1184 **Figure S9.** Mapping of Prp8 intein insertion sites to Prp8 exteins and domains

1185 **Figure S10.** Model of the *Cne* Prp8 intein interrupting Prp8 and the spliceosome

1186

1187 **Supplemental Tables**

1188 **Table S1.** Bacterial strains

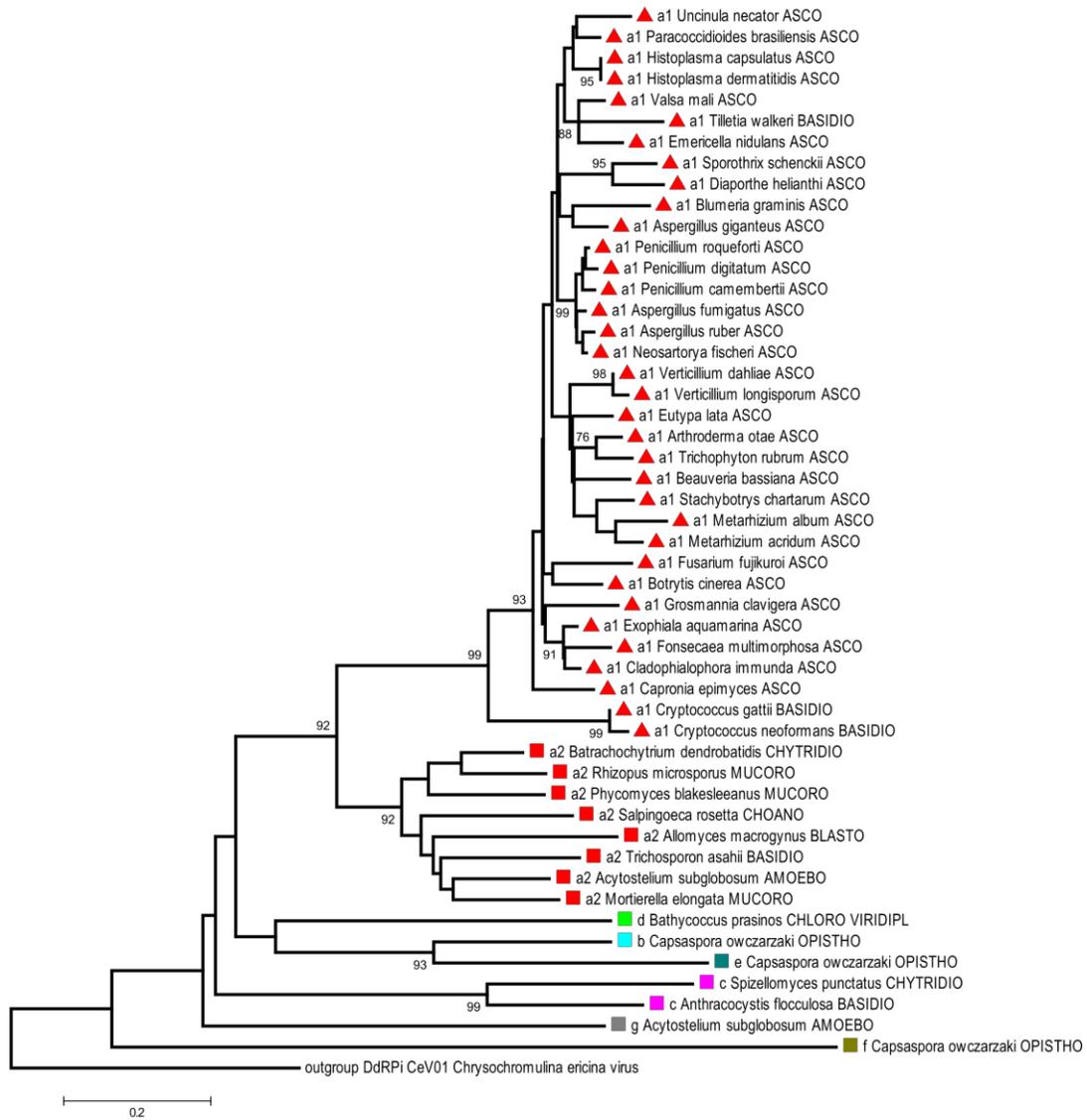
1189 **Table S2.** Plasmids and constructs

1190 **Table S3.** Oligonucleotide primers

1191 **Table S4.** Data collection, refinement statistics, and model details for *Cne* Prp8 intein
1192 crystal structure

SUPPLEMENTAL FIGURES

A.



B.

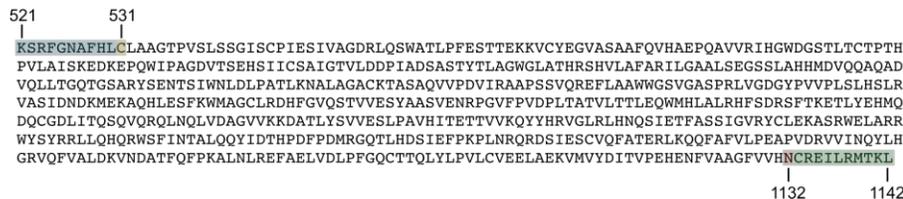


Figure S1. Distribution of Prp8 inteins and novel insertion site **g**. **A.** A phylogenetic tree of Prp8 inteins was reconstructed based on an amino acid multiple sequence alignment of

the splicing blocks (A, B, F, G) using the Neighbor-Joining algorithm and an Interior-branch test with 1000 replicates. Fifty representatives covering Prp8 intein diversity were selected and the full name of each intein-containing organism is listed. Colored symbols represent the insertion site and correspond to colors in Fig. 1A. Letters (a1, a2, b, c, d, e, f, g) represent each of the seven unique insertion sites. Phylum abbreviations are as follows: Amoebo – Amoebozoa; Asco – Ascomycota; Basidio – Basidiomycota; Blasto – Blastocladiomycota; Choano – Choanoflagellida; Chloro Viridipl – Chlorophyta Viridiplantae; Chytridio – Chytridiomycota; Mucoro – Mucoromycota; Opistho – Opisthokonta. **B.** Novel Prp8 insertion site **g**. In the amoeba *Acytostelium subglobosum* (*Asu*), an intein was identified at a new site in Prp8, here termed **g**. This is the seventh site in which a Prp8 intein has been found. The full site **g** intein sequence is shown, plus 10 flanking N-extein (blue) and C-extein (green) amino acids. The *Asu* C1 (yellow) and terminal asparagine (red) are highlighted. Residue numbering corresponds to the *Asu* Prp8 exteins. Accession number: XP_012753295.

```
1 50
a2_Asu CFSASTSLMAFDGVSDLVVTNNHILCLPEQYYGEMVDGNQRFLDFLVVHN
a2_Sro CFRRTTELLASGVEFLVVTGNHILCLPEPYVGFVVDGNKRFLDFLVTHN
a2_Bde CFAGHTSLLLADGDELVVTGNHILCLPQPYYGFKVDGNQRFLDFMVLHN
a2_Ama CFDRQTELKFYGGQPPLVTDNHILCLYEQYFGEMVDGNQRFLDRMVVHN
a2_Rmi CFDGETPLMLSDGGETLVVTGNHILCLPKKYYGFKVDGNQRFLDFMVLHN
a2_Mel CFSADTRLLRADGVKPLTVTNNHILCFEAYFGFLVDGNQRFLDFLVVHN
a2_Pbl CFSGDTFLLLADGGEALEVTGNHILCLPQTYYGFRIDGNQRFLDGLVVHN
a2_Tas CFSRDTRFIKYDRASDLVVTHNHILCLPEPYFGFKVDGNQRFLDFLVVHN
a1_Cga CLQNGTRLLRADGLEDLVCTHNHILSLPTKWSGFVVDKDSLYLDYLVLHN
a1_Cne CLQNGTRLLRADGLEDLVCTHNHILSMPTKWSGFVVDKDSLYLDYLVLHN
a1_Twa CLAKGTQLLAYDGKEELRVTANHIMVLETEWAGFRVDGDQLYLDYLVLHN
a1_Ssc CLAKGTRLCRYDGCEDLVVTPNHILTLETEWAGFRVDQDQLYLDHVVVHN
a1_Bgr CLALGTLLLQFDGQEDLVVTSNHILVLETEWAGFRVDQDHLYLDYLVLHN
a1_Une CLAKGTLLLRPDGLEDLVVTPNHILVLEREWAGFRVDHDQLYLDFVLHN
a1_Fmu CLANGTMLLRYDGREDLVVTPNHILVLSTKWSGFVVDKQDQLYLDYVVLHN
a1_Gcl CLAYGTMLLRYDGKEDLVVTSNHILALETKWAGFRTDCDQLYLDYLVLHN
a1_Sch CLANGTMLLRHDGVEDLVVTSNHILVFTTEWAGFRVDGDQLYLDYLVLHN
a1_Bba CLAKGTMLLRYDHEEDLVVTPNHILVLSTEWAGFRVDGDQLYLDNLVLHN
a1_Mal CLANGTMLLRHDRKEDLVVTPNHILVFARAWAGFRVDGDQLYLDHLVLHN
a1_Mac CLANGTMLLRYDHKEDLVVTSNHILVFATEWAGFRVDGDQLYLDHLVLHN
a1_Ffu CLAWNTRLLRYDHKEDLVVTPNHILVLSTKWAGFRVDRDQLYLDYIVLHN
a1_Aot CLANGTMLLRYDHKEDLVVTANHILVLATGWAGFRVDGDQLYLDYLVLHN
a1_Hca CLAKGTQLLRYDGKEDLVVTPNHILVLMTEWAGFRVDKQDQLYLDFVLHN
a1_Cep CLAYGTLLLRHDGKEDLVVTPNHILVLPTKWAGFRVDKQDQLYLDHMVLHN
a1_Cim CLANGTMLLRYDGVEDLVVTPNHILVLPTKWAGFRVDKQDQLYLDYLVLHN
a1_Eaq CLANGTMLLRYDGKEDLVVTPNHILVLVATGWAGFRVDKQDQLYLDYLVLHN
a1_Ela CLAKGTMLLRFDHKEDLVVTPNHILVLTTEWAGFRVDKQDQLYLDYLVLHN
a1_Tru CLAKGTMLLRYDHKDDLVVTANHILVLTTGWAGFRVDKQDQLYLDYLVLHN
a1_Pbr CLAKGTLLLRYDGKEDLVVTPNHILVLATEWAGFRVDQDQLYLDFVLHN
a1_Hde CLAKGTQLLRYDGKEDLVVTPNHILVLMTEWAGFRVDKQDQLYLDFVLHN
a1_Bci CLAWDTKLLRYDGKEDLVVTPNHILVLATNWAGFRVDKQDQLYLDYLVLHN
a1_Vda CLAKGTMLLRYDSKEDLVVTPNHILVLDTEWAGFRVDQDQLYLDYLVLHN
a1_Vlo CLAKGTMLLRYDSKEDLVVTPNHILVLDTEWAGFRVDQDQLYLDYFVLHN
a1_Vma CLAKGTQLLQYDGKEDLVVTPNHILVLATEWAGFRVDKQDQLYLDYLVLHN
a1_Pro CLAKGTRLLRYDGKEDLVVTPNHILVLKTEWAGFRVDKQDQLYLDYLVLHN
a1_Pdi CLAKGTRLLRFDGKEDLVVTPNHILVLKTEWAGFRVDKQDQLYLDYLVLHN
a1_Pca CLAKGTRLLRYDGKEDLVVTPNHILVLKTEWAGFRVDKQDQLYLDHLVLHN
a1_Eni CLANGTQLLRYDGKEDLVVTANHILVLATEWAGFRVDKQDQLYLDYVVLHN
a1_Aru CLAKGTRLLRYDGIEDLVVTPNHILVLTTEWAGFRVDKQDQLYLDYLVLHN
a1_Agi CLAKGTRLLQYDGQEDLVVTPNHILVLSTEWAGFRVDKQDQLYLDYLVLHN
a1_Afu CLAKGTRLLRYDGKEDLVVTPNHILVLATEWAGFRVDKQDQLYLDYLVLHN
a1_Nfi CLAKGTRLLRYDGKEDLVVTPNHILVLTTEWAGFRVDKQDQLYLDYLVLHN
a1_Dhe CLAKGTRLLALYDGGEDLVVTPNHILALETEWAGFRVDQDQLYLDHVVVHN
c_Spu CLAAGTKVMVVRNSQEYTVEGHLLSLTYFRFSVDGYQSR-FLDWTVVHN
c_Afl CFAYGTEVLAIRQTQYTVTSGHLLSLSFVKFKVSGDTRR-FLDWTVVHN
b_Cow CFSPETPVRMADGESGFTCNSRHDLVLPYV--GFELDGSPFLFNWLVVHN
d_Bpr CIRDDQLVLRADGERGFVCTPNHIVCLATFTGFQIDNESNRFLDGTVTHN
e_Cow CTAPDLVRMADGDDGFECNGNHILVIEYVGFVAVGESPLFCDFLVAHN
f_Cow CFAEDHEVLTAHGGVSLCPTDNHRMWAHLIMFRVLERNHSGVRPIIVGN
g_Asu CLAAGTPVSLSSGGSTLTCTPTHPVLAELVDLPEGQCTTQLYIAGFVVHN
outgroup CFDPETPIMMWSGFIKHRVTDNHILTLVGKYVGWQLEDKKGFRFGGLVVHN
Consensus CLA GT LLRYDG EDLVVT NHILVL T WAGFRVD DQLYLDYLVLHN
```

Figure S2. Amino acid multiple sequence alignment of Prp8 inteins utilized for phylogenetic analysis. Comparative analysis of amino acid residues found in blocks A, B, F, and G from the selected 50 representative Prp8 inteins, shown with abbreviated species

names (full names in Fig. S1). Letters (a1, a2, b, c, d, e, f, g) represent each of the seven unique insertion sites. Shading is as follows: black – identical amino acid, dark gray – conserved amino acid, light gray – similar amino acid substitution.

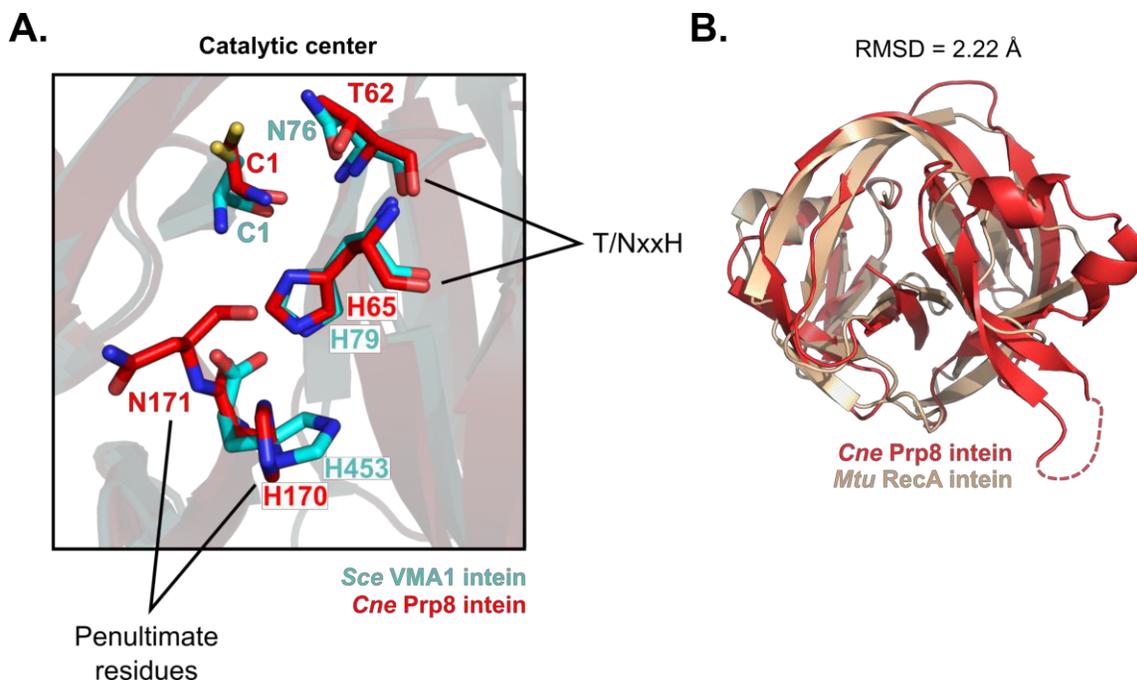


Figure S3. Overlays of the *Cne* Prp8 intein with other inteins. **A.** Overlay of the *Saccharomyces cerevisiae* (*Sce*) VMA1 intein and *Cne* Prp8 intein active sites. The *Sce* VMA1 intein (cyan, PDB 1GPP) was overlaid with the *Cne* Prp8 intein (red). The active site residues, crucial to protein splicing, are shown as sticks and labeled. A majority of these conserved residues overlap exactly, such as the catalytic C1, and the B block TxxH motif. The *Sce* VMA1 intein uses an asparagine (N76) rather than threonine in the TxxH motif, but the positioning is similar to the threonine (T62) of the *Cne* Prp8 intein. The penultimate histidines (H170 and H453) are in comparable positions except for the side chains, whose chi angles are different by 45°. The *Sce* VMA1 intein was not solved with the terminal asparagine. **B.** Structural comparison of bacterial *Mycobacterium tuberculosis* (*Mtu*) RecA intein and fungal *Cne* Prp8 intein. Overlay of the *Mtu* RecA intein (brown, PDB 2IMZ) and the *Cne* Prp8 intein (red) reveals structural similarities in major intein features, such as the anti-parallel β -sheet folding, that contribute to the

horseshoe shape. The Hint domain, comprised of splicing blocks A, B, F, and G, are generally aligned between the two inteins. The structures deviate at sequences between blocks B and F, where the *Cne* Prp8 intein encoded a linker or endonuclease domain. The two structures have an RMSD value of 2.22 Å.

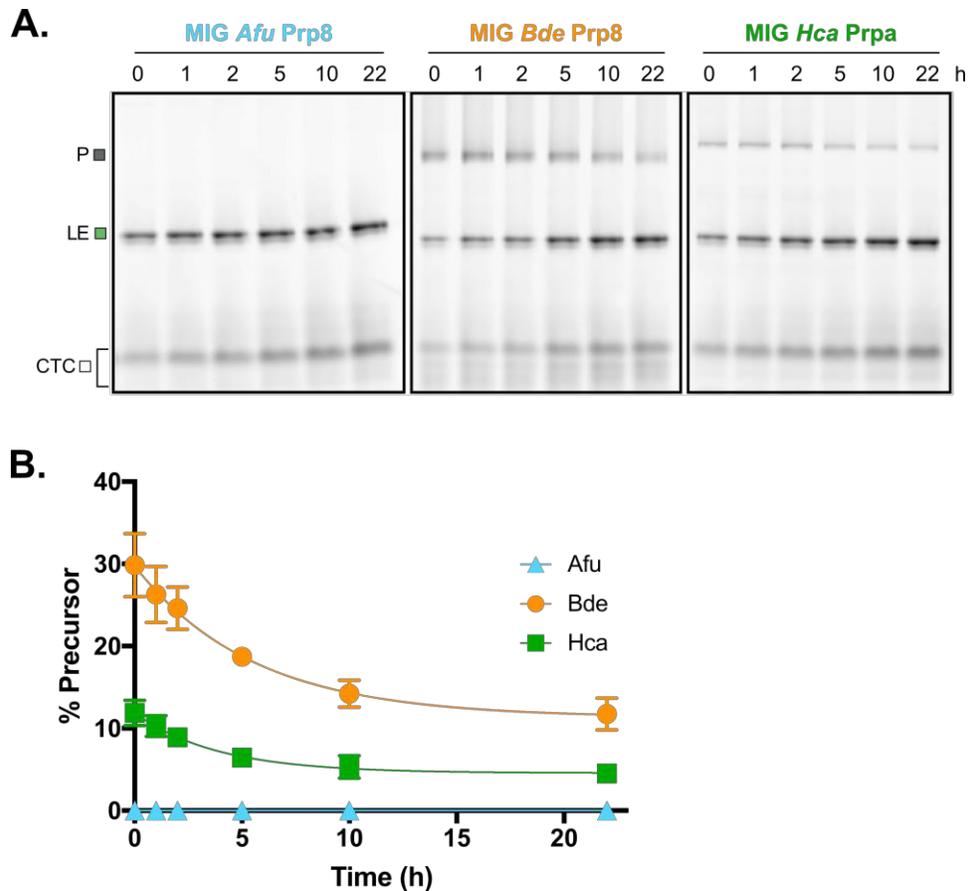


Figure S4. Splicing of Prp8-a inteins from other fungal pathogens in MIG. **A.** Diverse Prp8 intein splicing patterns. Several other Prp8 inteins from human fungal pathogens *Aspergillus fumigatus* (*Afu*), *Batrachochytrium dendrobatidis* (*Bde*), and *Histoplasma capsulatum* (*Hca*) were cloned into MIG. Splicing was observed over time by the loss of precursor (P) and increase in ligated exteins (LE), or simply by the presence of ligated exteins (for *Afu*). The gel shows that not all Prp8 inteins splice similarly, despite being placed in an identical extein context. **B.** Precursor amounts vary greatly. A quantitation of precursor (P) at each time-point shows that these Prp8 inteins are active, but splice at variable rates. The *Afu* Prp8 intein is almost entirely spliced at the start of the assay (0 h), whereas *Bde* has 31% precursor at 0 h and *Hca* has 14% precursor at 0 h. The variable

rates were determined by calculating the loss of precursor over time for MIG *Bde* Prp8 and MIG *Hca* Prp8, and are 3.6×10^{-2} per min and 1.9×10^{-2} per min, respectively. This suggests intein-mediated control of protein splicing. Data are representative of three biological replicates and mean standard deviations are shown. Trend lines are fit to show the decay curve.

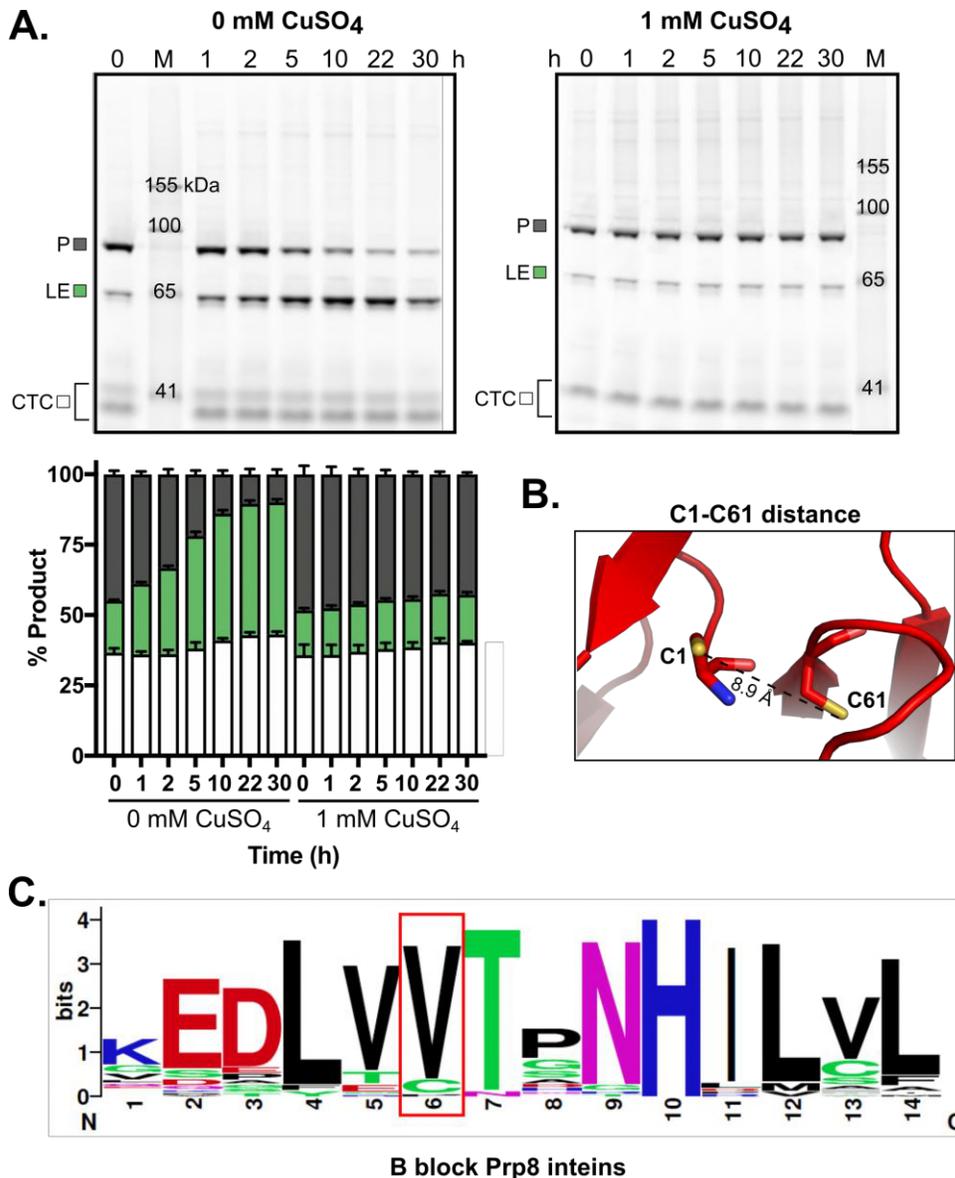


Figure S5. MIG Prp8 A-1V copper inhibition and cysteine analysis. **A.** Copper treatment causes inhibition. Induced MIG Prp8 A-1V cells were lysed and treated with 0 or 1 mM CuSO₄. The lysates were incubated for the indicated time at 30°C and then frozen. Samples were separated on SDS-PAGE and scanned for GFP fluorescence. In the absence of copper, MIG Prp8 A-1V spliced well over 30 h, converting precursor (P) into ligated extensins (LE). There was little to no conversion of P to LE over time with copper

was addition. The inhibition was still observed at 30 h after copper addition. Quantitation is shown below in a stacked plot. Data are representative of three biological replicates and mean standard deviations are shown. **B.** Relative position of two cysteines. There are only two cysteines present in the *Cne* Prp8 intein. Using the solved structure, a measurement of the distance between C1 and C61 (shown as sticks) was calculated to be 8.9 Å apart. **C.** Valine is the preferred residue at position 61. A sequence logo was constructed of the B block from 50 representative Prp8 inteins. This shows absolute conservation of the histidine (position 10) and a strong preference for threonine (position 7) in the TxxH motif. However, the B block cysteine (position 6, red box) is not highly conserved across Prp8 inteins and most encode valine at this site.

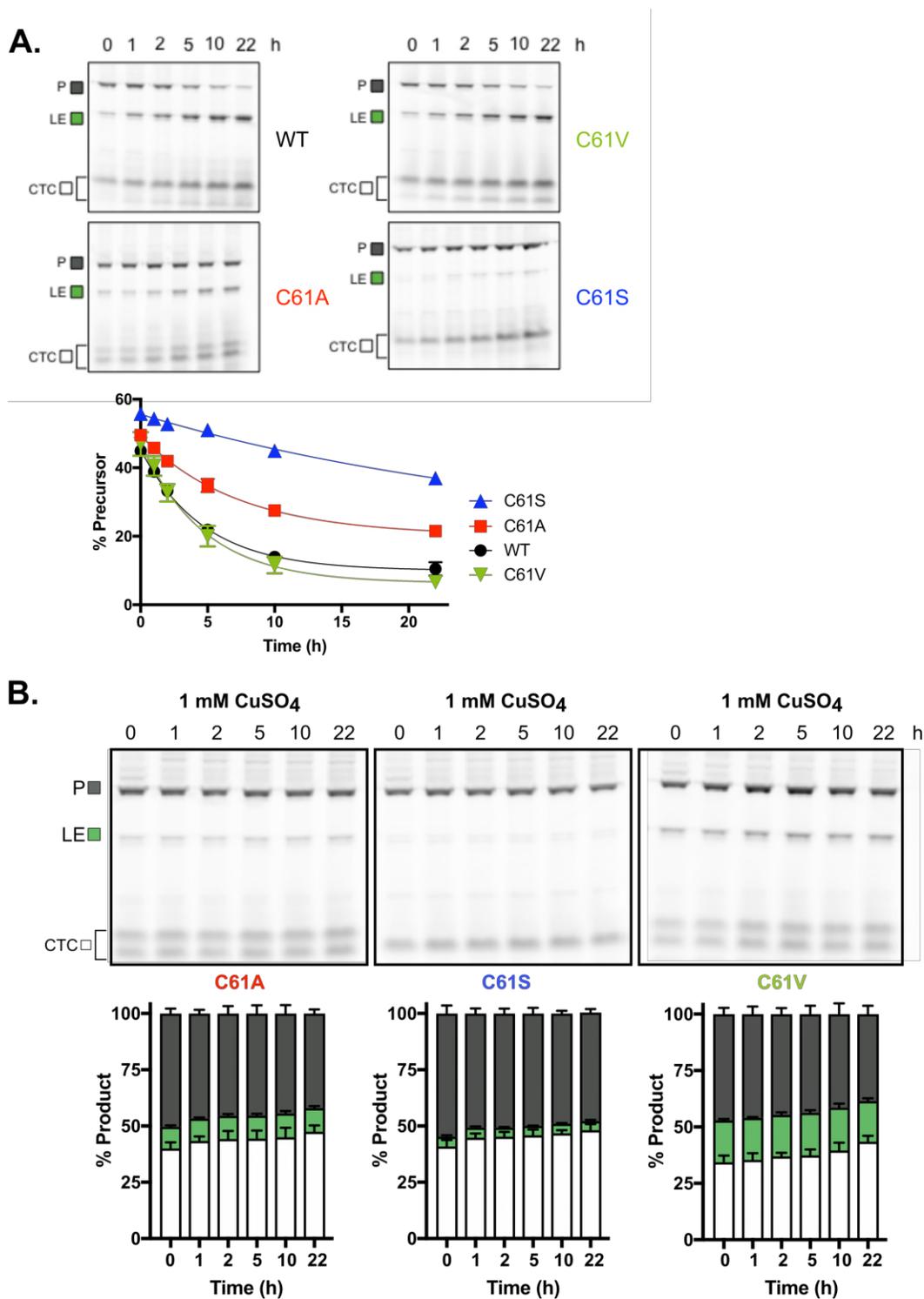
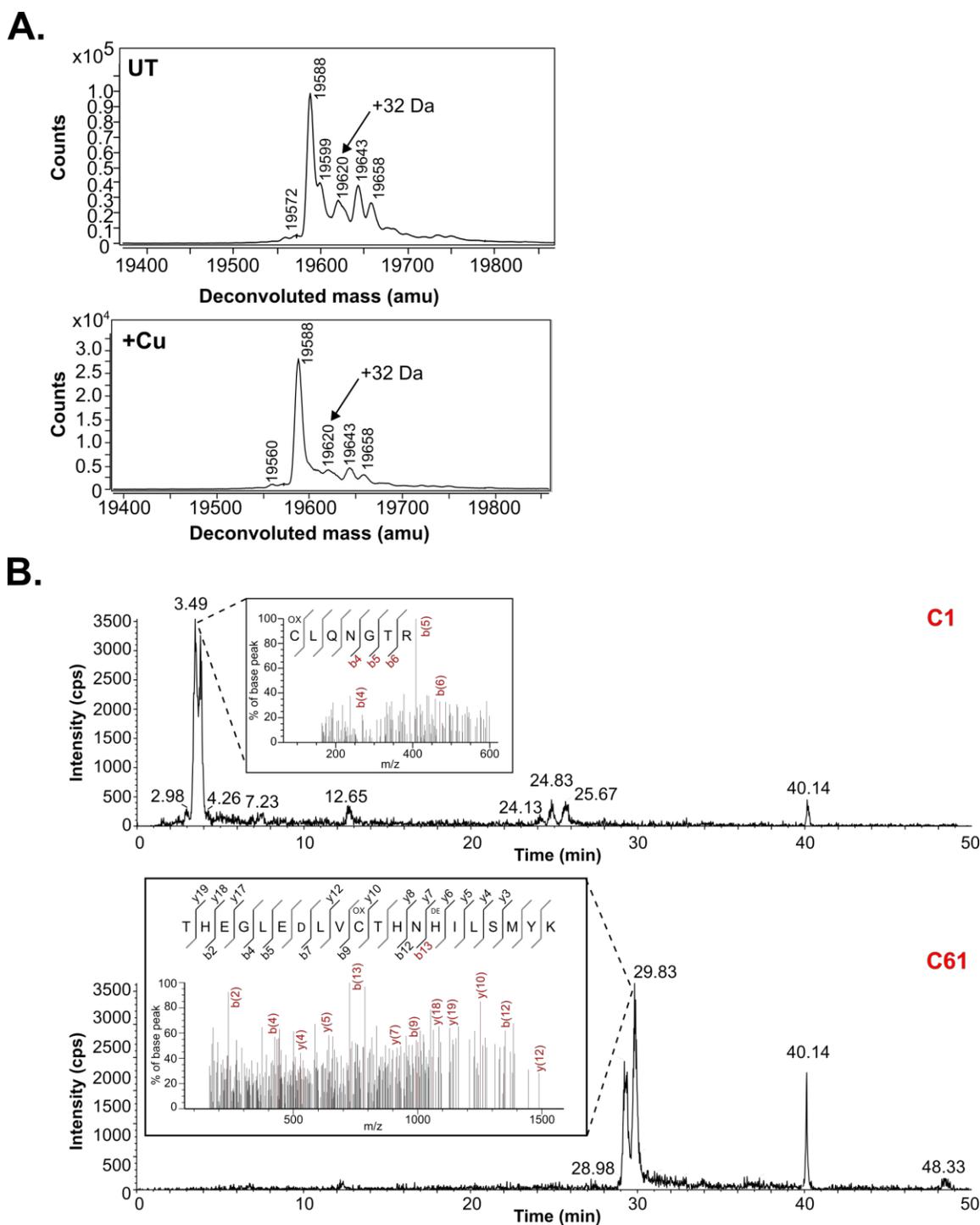


Figure S6. Copper inhibition of MIG Prp8 A-1V C61 mutants. **A.** Mutations to C61 in MIG Prp8 A-1V affect splicing rates. The B block C61 was mutated to valine (C61V),

alanine (C61A), and serine (C61S) and splicing observed over time in MIG. Splicing rates were determined by calculating the loss of precursor over time and are as follows: WT – 9.2×10^{-2} per min, C61V – 8.6×10^{-2} per min, C61A – 4.7×10^{-2} per min, and C61S – 1.6×10^{-2} per min. The C61V mutant splices similarly to WT, whereas C61A and C61S are slower. A quantitation is shown to the right with the amount of precursor (P) at each time-point. Data are representative of three biological replicates and mean standard deviations are shown. Trend lines are fit to show the decay curve. **B.** MIG Prp8 A-1V B block cysteine mutants are inhibited by copper. To test if copper inhibition was caused by C1 oxidation, C61 mutants were treated with CuSO_4 . After induction of MIG, the cells were lysed and 1 mM CuSO_4 was added. The lysates were incubated at 30°C and aliquots were collected at the indicated time. Samples were run on SDS-PAGE and scanned for GFP fluorescence. None of the C61 mutants show an increase in ligated exteins (LE) over time, with little loss of precursor (P). This indicates that at least C1 oxidation by copper is sufficient to cause the observed splicing inhibition, and that disulfide bonds are not involved. Quantitation is shown below in a stacked plot. Data are representative of three biological replicates and mean standard deviations are shown.



and separated and analyzed using LC-MS. The peaks were deconvoluted and the expected mass of the Prp8 intein, 19588 Da, is seen as the largest peak. A small, 32 Da shift (19620 Da) was visible with both no treatment and copper treatment only (arrow). This suggests highly reactive cysteines that are modifiable by atmospheric oxygen alone.

B. C1 and C61 are oxidized with copper treatment. Trypsin-digested fragments of copper-treated *Cne* Prp8 intein were separated and sprayed using LC-MS/MS (insets). Peptides (red peaks) containing C1 or C61 were detected and further analyzed using multiple reaction monitoring-initiated detection and sequencing (MIDAS) to confirm the identity and location of oxidation. The chromatogram shows elution time for both cysteines consistent with a single additional oxygen, or a sulfenic acid modification.

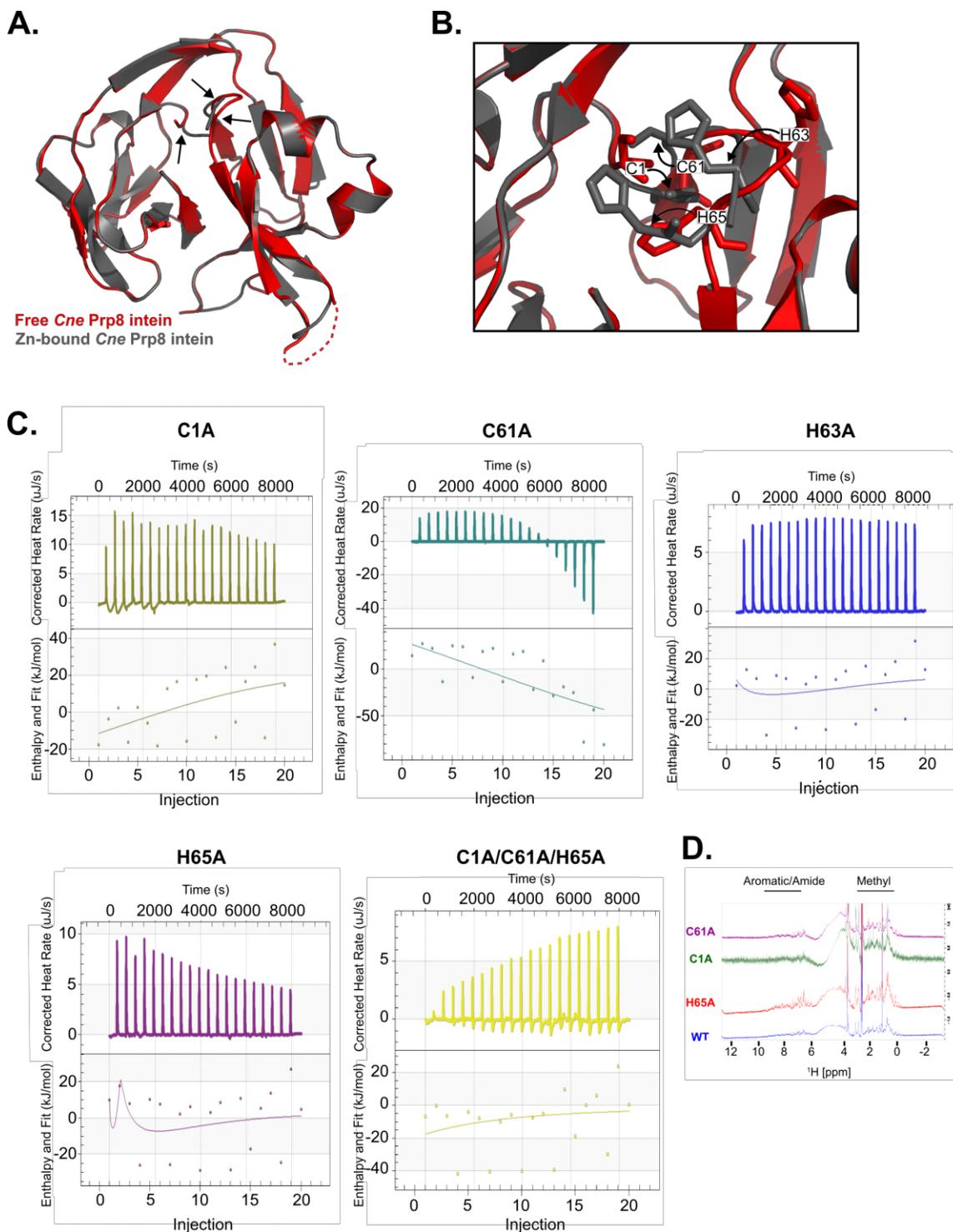


Figure S8. Putative zinc binding site. **A.** Overlay of the free *Cne* Prp8 intein with zinc-bound intein model. The free *Cne* Prp8 intein (red) is overlaid with the optimized zinc-

binding model (gray). Minor conformational changes of the intein are required for the pocket to form, indicated by arrows, with an RMSD of 0.104 Å. **B.** Zoom into zinc coordination residues. The residues suspected to form a binding pocket (C1, C61, H63, and H65) are shown as sticks in the free intein structure (red) or zinc-bound intein model (gray). Rotational freedom of these amino acid side chains allows movement of the residues to coordinate the zinc ion. Arrows indicate direction of movement from free intein to zinc-bound structure. **C.** Using isothermal titration calorimetry (ITC), purified 16 μM Prp8 intein with various binding site mutations was titrated with 0.5 mM ZnSO₄ over 20 injections at 37°C and pH 7.0 on a Nano ITC. Changes in heat after incremental ZnSO₄ injections are shown in the top panel and are corrected by subtracting the heat of ZnSO₄ titrated into buffer alone. The bottom plot shows integrated heat per mole of ZnSO₄ as a function of the molar ratio of ZnSO₄ to Prp8 intein. Binding was not detected in a biologically significant range for all mutants compared to the wild-type, suggesting that one, two, or all three residues play a role in zinc binding to the *Cne* Prp8 intein (see Table 1). Although the ITC data integration and binding model affirmation occasionally produced high error values, experiments were conducted in triplicate, and persistently completed highly reproducible enthalpic transitions. Repeated, corresponding transition levels along the binding isotherms at stoichiometric equivalent points consistently approached unity values despite these errors. Representative data is shown. **D.** 1D proton NMR spectrum of WT *Cne* Prp8 intein is shown (blue) and overlaid with the same spectra of three representative mutants (C1A, C61A, and H65A). The Aromatic/Amide amino acid region is similar across all proteins, suggesting that the mutants fold and are structurally sound.

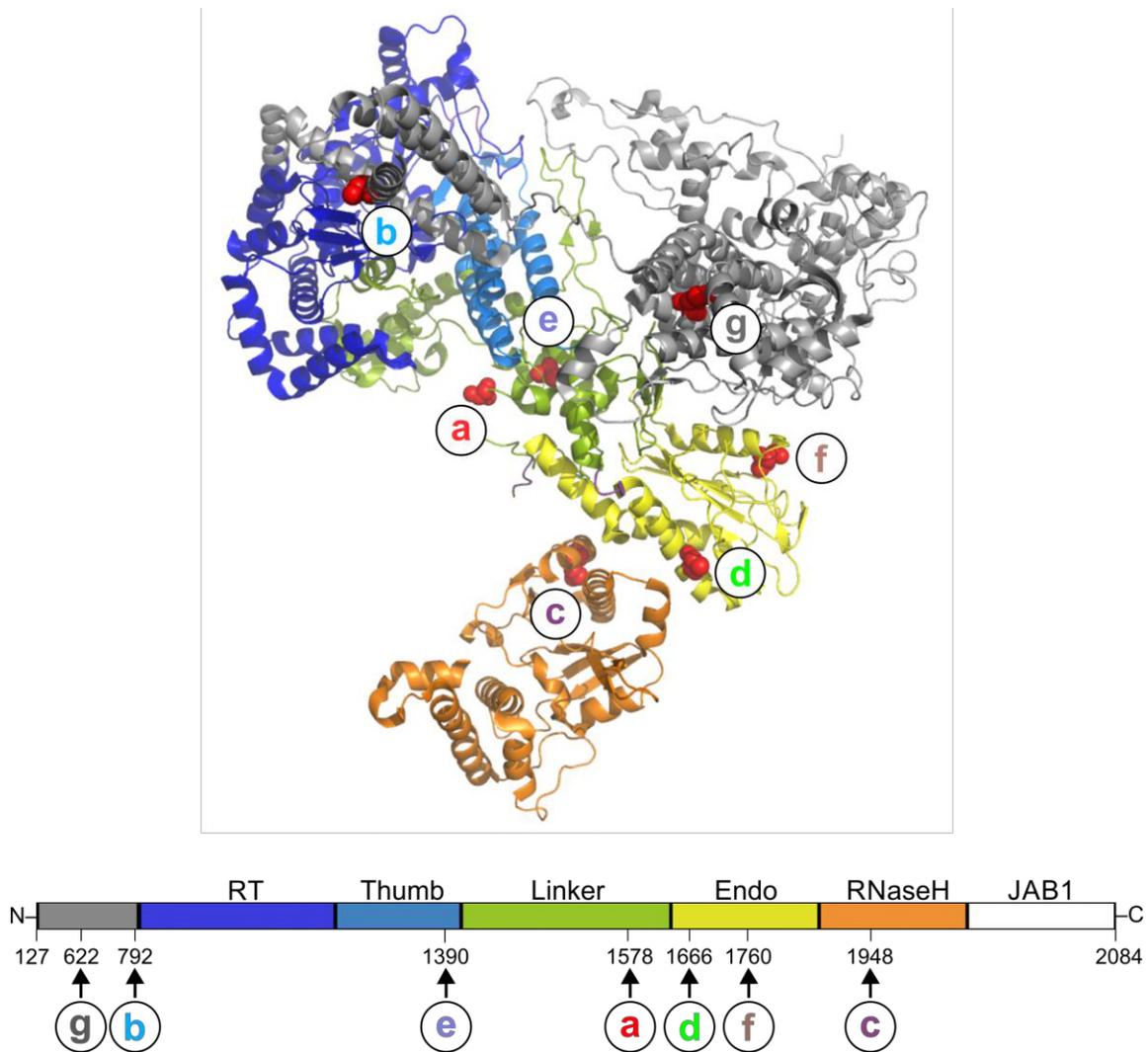


Figure S9. Mapping of Prp8 intein insertion sites to Prp8 extein domains. The seven unique insertion sites (**a-g**) were mapped to a solved structure of Prp8 from a *Saccharomyces cerevisiae* C complex spliceosome (PDB 5GMK, chain A from Wan et. al, 2016) by locating the +1 residue. This Prp8 structure was used because the insertion sites are all resolved. The +1 residues are shown as red spheres and labeled **a-g**. Most Prp8 inteins localize close to the active center of Prp8. Some insertions are in the N-terminal domain, which provides structural integrity to the spliceosome. A corresponding line diagram of Prp8 exteins shows the domains of the host protein from amino acid

residues 127 to 2084 with arrows indicating the site of intein insertion with the residue number and insertion site letter. The domains are as follows: N-terminal domain – gray, RT Palm/Finger – dark blue, Thumb/X – light blue, linker – green, endonuclease – yellow, and RNase H-like – orange.

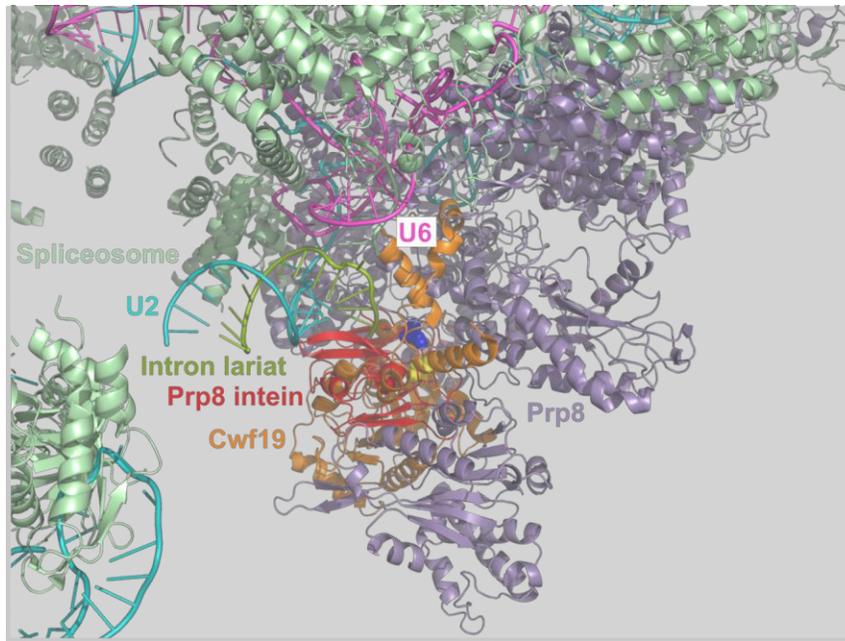


Figure S10. Model of the *Cne* Prp8 intein interrupting Prp8 and the spliceosome. The Prp8 intein-containing Prp8 precursor model was docked into a spliceosome structure from *Schizosaccharomyces pombe* (PDB 3JB9) to look for intein-spliceosome disruptions. Prp8 is shown as lavender and the Prp8 intein is shown as red, and the rest of the spliceosome is mint green. This reveals that the Prp8 intein would occupy a relatively sparse location in the spliceosome. The intein clashes are shown here (with labels) and noted in Fig. 6B.

SUPPORTING TABLES

Table S1. Bacterial strains

Strain	Features and comments	Source
DH5 α	F ⁻ <i>endAI recAI hsdR17</i> (rK ⁻ mK ⁻) <i>deoR supE44 thi-J gyrA96 relA</i>	Gibco-BRL
MG1655(DE3)	F ⁻ (λ DE3) <i>ilvG rfb50 rph1</i>	James Imlay
BL21(DE3)	F ⁻ <i>ompT hsdS_B</i> (r _B ⁻ m _B ⁻) <i>gal dcm</i> (λ DE3)	Novagen

Table S2. Plasmids and constructs

Plasmid/construct	Features and comments	Source
pACYCDuet-1	Expression vector, T7 promoter, Cam ^R .	Novagen
pACYC MIG SufB	Used as cloning backbone by removing <i>M. tuberculosis</i> SufB intein insert using ClaI/SphI.	Topilina et al. 2015
pACYC MIG Prp8	<i>C. neoformans</i> var. <i>grubii</i> H99 Prp8 intein flanked by short native exteins (N-extein: FWEKA; C-extein: SGFEE) cloned into ClaI/SphI sites between MBP and GFP coding sequences in pACYCDuet-1 backbone.	Present study
pACYC MIG Prp8 A-1V	Same as pACYC MIG Prp8 but with the indicated amino acid mutation at the last amino acid of the N-extein (position -1).	Present study
pACYC MIG Prp8 A-1V C61A/S/V	Same as pACYC MIG Prp8 A-1V but with the indicated amino acid mutation at the B block C61.	Present study
pACYC MIG <i>Afu</i> Prp8	<i>Aspergillus fumigatus</i> 293 Prp8 intein flanked by short native exteins (N-extein: FWERA; C-extein: SGFEE) cloned into ClaI/SphI sites between MBP and GFP coding sequences in pACYCDuet-1 backbone.	Present study
pACYC MIG <i>Bde</i> Prp8	<i>Batrachochytrium dendrobatidis</i> JEL423 Prp8 intein flanked by short native exteins (N-extein: FWEKA; C-extein: SGFEE) cloned into ClaI/SphI sites between MBP and GFP coding sequences in pACYCDuet-1 backbone.	Present study
pACYC MIG <i>Hca</i> Prp8	<i>Histoplasma capsulatum</i> G186A Prp8 intein flanked by short native exteins (N-extein: FWERA; C-extein: SGFEE) cloned into ClaI/SphI sites between MBP and GFP coding sequences in pACYCDuet-1 backbone.	Present study
pET28a	Expression vector, T7 promoter, C-terminal His ₆ tag, thrombin cleavage, Kan ^R .	Novagen
pET28a <i>Cne</i> Prp8 intein	<i>C. neoformans</i> var. <i>grubii</i> H99 Prp8 intein with 2	Present

	native N-extein residues (KA) in pET28a backbone at NcoI/XhoI sites.	study
pET47b	Expression vector, T7 promoter, N-terminal His ₆ tag, PreScission cleavage, Kan ^R .	Novagen
pET47b <i>Cne</i> Prp8 intein 3 N-exteins	<i>C. neoformans</i> var. <i>grubii</i> H99 Prp8 intein with 3 native N-extein residues (EKA) in pET47b backbone at BamHI/NotI sites.	Present study
pET47b <i>Cne</i> Prp8 intein 3 N-exteins C1A/C61A/H65A	<i>C. neoformans</i> var. <i>grubii</i> H99 Prp8 intein with 3 native N-extein residues (EKA) plus C1A/C61A/H65A proposed binding site knockout in pET47b backbone at BamHI/NotI sites.	Present study
pET47b <i>Cne</i> Prp8 intein 3 N-exteins C1A	<i>C. neoformans</i> var. <i>grubii</i> H99 Prp8 intein with 3 native N-extein residues (EKA) and a single C1A mutation in pET47b backbone at BamHI/NotI sites.	Present study
pET47b <i>Cne</i> Prp8 intein 3 N-exteins C61A	<i>C. neoformans</i> var. <i>grubii</i> H99 Prp8 intein with 3 native N-extein residues (EKA) and a single C61A mutation in pET47b backbone at BamHI/NotI sites.	Present study
pET47b <i>Cne</i> Prp8 intein 3 N-exteins H65A	<i>C. neoformans</i> var. <i>grubii</i> H99 Prp8 intein with 3 native N-extein residues (EKA) and a single H65A mutation in pET47b backbone at BamHI/NotI sites.	Present study

Table S3. Oligonucleotide primers

Oligo ID	Sequence (5' to 3')	Application
IDT3967	GGGGCATGCATTCTGGGAGAAAGCCTGTCTGCAG	<i>Cne</i> Prp8 intein forward primer for MIG with SphI site.
IDT3968	GGGATCGATTCCTCAAATCCTGAGTTGTGCAGTAC C	<i>Cne</i> Prp8 intein reverse primer for MIG with ClaI site.
IDT4027	ACCATTCTGCAGACANNSTTTCTCCCAGAATGCATG CATGTAAGGCC	MIG Prp8 N-1 random mutagenesis sense primer. N is any nucleotide, S is a G or C.
IDT4028	GGCCTTACATGCATGCATTCTGGGAGAAANNSTGTC TGCAGAATGGT	MIG Prp8 N-1 random mutagenesis antisense primer. N is any nucleotide, S is a G or C.

IDT5841	GAGGGACTCGAAGATCTGGTCGCTACCCATAACCA CATCCTTTC	MIG Prp8 A-1V mutagenesis C61A sense primer.
IDT5840	GAAAGGATGTGGTTATGGGTAGCGACCAGATCTTC GAGTCCCTC	MIG Prp8 A-1V mutagenesis C61A antisense primer.
IDT5987	GAGGGACTCGAAGATCTGGTCAGTACCCATAACCA CATCCTTTC	MIG Prp8 A-1V mutagenesis C61S sense primer.
IDT5986	GAAAGGATGTGGTTATGGGTACTGACCAGATCTTC GAGTCCCTC	MIG Prp8 A-1V mutagenesis C61S antisense primer.
IDT6026	TGAGGGACTCGAAGATCTGGTCGTTACCCATAACC ACATCCTTCT	MIG Prp8 A-1V mutagenesis C61V sense primer.
IDT6027	AGAAAGGATGTGGTTATGGGTAACGACCAGATCTT CGAGTCCCTCA	MIG Prp8 A-1V mutagenesis C61V antisense primer.
IDT5893	GAGGGAAGGCCTTACATGCATGCATTTTGGGAAAG AGCATGCCTTG	<i>Afu</i> Prp8 intein forward primer for MIG with SphI site with overhang InFusion cloning.
IDT5894	TCTCCTTTGCTCATATCGATTTCTTCAAATCCACTGT TATGCAAG	<i>Afu</i> Prp8 intein reverse primer for MIG with ClaI site with overhang for InFusion cloning.
IDT5891	GAGGGAAGGCCTTACATGCATGCATTTTGGGAAAA GGCATGTTTTGC	<i>Bde</i> Prp8 intein forward primer for MIG with SphI site with overhang InFusion cloning.
IDT5892	TCTCCTTTGCTCATATCGATCTCCTCAAACCCAGAG TTATGC	<i>Bde</i> Prp8 intein reverse primer for MIG with ClaI site.
IDT5889	GAGGGAAGGCCTTACATGCATGCATTTTGGGAACG AGCCTGTCT	<i>Hca</i> Prp8 intein forward primer for MIG with SphI site with overhang InFusion cloning.
IDT5890	TCTCCTTTGCTCATATCGATTTCTTCAAATCCGCTGT TATGT	<i>Hca</i> Prp8 intein reverse primer for MIG with ClaI site with overhang for InFusion cloning.

<i>Cne</i> Prp8 pET28a F	CTTTAAGAAGGAGATATACCATGGGCAAAGCCTGT CTGCAGAATGGTAC	<i>Cne</i> Prp8 intein with two native N-exteins (KA) forward primer for amplification and cloning into pET28a with NcoI end.
<i>Cne</i> Prp8 pET28a R	CAGTGGTGGTGGTGGTGGTGGTCTCGAGTGAGTTGTG CAGTACCAAATAGTC	<i>Cne</i> Prp8 intein with S+1 reverse primer for amplification and cloning into pET28a with XhoI end.
IDT6143	GGGGGATCCAGAGAAAGCCTGTCTGCAGAATG	<i>Cne</i> Prp8 intein with three native N-exteins (EKA) forward primer for amplification and cloning into pET47b with BamHI end.
IDT6142	GGGGCGGCCGCTTAGTTGTGCAGTACCAAATAGTC ATAAC	<i>Cne</i> Prp8 intein reverse primer for amplification and cloning into pET47b with NotI end.
IDT6149	TACCAGGATCCAGAGAAAGCCGCTCTGCAGAATGG TACTCGTCT	<i>Cne</i> Prp8 intein C1A sense mutagenesis primer.
IDT6158	AGACGAGTACCATTCTGCAGAGCGGCTTTCTCTGGA TCCTGGTA	<i>Cne</i> Prp8 intein C1A antisense mutagenesis primer.
IDT6299	ATGAGGGACTCGAAGATCTGGTCGCTACCCATAAC GCCATCCTTTCTATGTATAAAGAAA	<i>Cne</i> Prp8 intein C61A and H65A sense multisite mutagenesis primer.
IDT6298	TTTCTTTATACATAGAAAGGATGGCGTTATGGGTAG CGACCAGATCTTCGAGTCCCTCAT	<i>Cne</i> Prp8 intein C61A and H65A antisense multisite mutagenesis primer.
IDT6145	TCGAAGATCTGGTCTGTACCCATAACGCCATCCTTT CTATGTATAAAGAAAG	<i>Cne</i> Prp8 intein H65A sense mutagenesis primer.
IDT6144	CTTTCTTTATACATAGAAAGGATGGCGTTATGGGTA CAGACCAGATCTTCGA	<i>Cne</i> Prp8 intein H65A antisense mutagenesis primer.

Table S4. Data collection, refinement statistics, and model details for *Cne* Prp8 intein crystal structure

Data Collection	
Wavelength (Å)	0.97946
Resolution (Å)	32.7-1.75 (1.81-1.75)
Space group	P 1
Unit cell	
a,b,c (Å)	63.15 63.23 80.20
α,β,γ (°)	103.32 95.53 117.34
Total reflections	211,948 (19,137)
Unique reflections	90,641 (8,259)
Multiplicity	2.3 (2.3)
Completeness (%)	85.91 (77.71)
Mean $I/\sigma(I)$	12.48 (1.55)
R_{merge} (%)	0.069 (0.91)
Refinement	
R_{work}	0.197 (0.293)
R_{free}	0.235 (0.310)
No. of non-H atoms	7576
macromolecules	7026
solvent	550
RMS (bonds) (Å)	0.005
RMS (angles) (°)	1.02
Ramachandran	
favored (%)	98.42
allowed (%)	1.58
outliers (%)	0.00
Rotamer outliers (%)	0.00
Clashscore	3.45
Average B-factor (Å ²)	30.39
macromolecules	29.81
solvent	37.80

Statistics for the highest-resolution shell are shown in parentheses.