Title
Regulating ER protein homeostasis by differentially processing mRNAs

Author
LI, WEIHAN

Publication Date
2018

Peer reviewed|Thesis/dissertation

Regulating ER protein homeostasis by differentially processing mRNAs

by

Weihan Li

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Committee in Charge

Acknowledgements

Part of the text of this thesis is a reprint of the material as it appears in Li, Weihan, et al. "Engineering ER-stress dependent non-conventional mRNA splicing." eLife 7 (2018): e35388. The co-author listed in this publication directed and supervised the research that forms the basis for the thesis. I would like to thank Peter Walter for his fantastic mentorship, the Walter lab for the constant scientific communication and my thesis committee, Jonathan Weissman and Wallace Marshall, for their guidance. I would like to dedicate this work to my father Peixin Li, my mother Kun Han, my grandfathers Zenglie Li, my grandmothers Fule Hu and my wife Yi Song. Without their support, I wouldn't have the chance to take this adventurous and serendipitous journey from Xi'an China, to Hong Kong and eventually to US for my education.

Contributions

This work has been published on *eLife* at 2018 (Li, Weihan, et al. "Engineering ER-stress dependent non-conventional mRNA splicing." eLife 7 (2018): e35388.) Weihan Li designed the experiments, performed the experiments, analyzed the data and wrote the manuscript. Voytek Okreglak, Jirka Peschek, Philipp Kimmig and Peter Walter designed the experiments and edited the manuscript. Meghan Zubradt and Jonathan S Weissman were our collaborators, who performed the DMS-seq experiments and the data analysis.

Regulating ER protein homeostasis by differentially processing mRNAs

Weihan Li

Abstract

The endoplasmic reticulum (ER) protein folding capacity is balanced with the protein folding burden to prevent accumulation of un- or misfolded proteins. The ER-membrane resident kinase/RNase Ire1 maintains ER protein homeostasis by initiating two distinct mRNA processing programs. First, in both metazoans and Saccharomyces cerevisiae, Ire1 catalyzes the non-conventional cytoplasmic mRNA splicing of XBP1 (metazoans) or HAC1 (*S. cerevisiae*)—thereby initiating a transcriptional response that increases the ER folding capacity. Second, in metazoans and *Schizosaccharomyces pombe*, Ire1 selectively degrades ER-localized mRNAs—thereby post-transcriptionally reducing the ER's protein folding burden. Thus, Ire1 orthologs in *S. cerevisiae* and *S. pombe* are specialized to only one of the two functional outputs, while Ire1 in metazoans can perform both. Here, we show that the respective Ire1 orthologs have become specialized for their functional outputs by divergence of their RNase specificities. In addition, RNA structural features separate the splicing substrates from the decay substrates. Using these insights, we *de novo* engineered non-conventional mRNA splicing in *S. pombe*, which confers *S. pombe* with both Ire1 functional outputs.

Table of Contents

List of Figures

CHAPTER I: Introduction: The Ire1 branch of the unfolded protein response

In eukaryotes, the vast majority of secretory and transmembrane proteins are folded in the endoplasmic reticulum (ER). The ER protein folding homeostasis is maintained by a collective of signaling pathways, termed the unfolded protein response (UPR) [1, 2]. The most evolutionarily conserved branch of the UPR is mediated by the ER-transmembrane kinase/endoribonuclease (RNase) Ire1. Direct binding of unfolded proteins to Ire1's ER lumenal domain triggers Ire1 to oligomerize and form foci [3-7]. In turn, Ire1 oligomerization activates Ire1's cytosolic kinase/RNase domain [8], which restores ER homeostasis through two functional outputs. First, Ire1 initiates a process of non-conventional cytosolic splicing of *XBP1* mRNA (in metazoans) or *HAC1* mRNA (in *S. cerevisiae*). Translation of the spliced mRNA produces a transcription factor Xbp1 (Hac1 in *S. cerevisiae*), which drives a large transcriptional program to adjust the ER's protein-folding capacity according to the protein folding load in the ER lumen [9-13]. Second, Ire1 can reduce the ER folding burden by cleaving a set of mRNAs encoding ER-target proteins. The initial Ire1-mediated cleavage leads to mRNA degradation, in a process termed regulated Ire1-dependent decay (RIDD) [14-16]. The mechanism that distinguishes the non-conventional mRNA splicing from RIDD has largely remained unknown.

Interestingly, the two Ire1 modalities co-exist in metazoan cells [14, 15, 17], yet are evolutionarily separated in the two yeast species, *S. cerevisiae* and *S. pombe*. The UPR in *S. cerevisiae* engages Ire1 exclusively in mRNA splicing, whereas in *S. pombe* it engages Ire1 exclusively in RIDD. There is no detectable RIDD in *S. cerevisiae* and no *HAC1*/ *XBP1* ortholog in *S. pombe*, nor is there a corresponding transcriptional program [16, 18]. It is intriguing to note that the fundamental task of maintaining ER protein homeostasis can be

achieved by two radically different processes catalyzed by two distantly related Ire1 orthologs. The two yeast species, *S. cerevisiae* and *S. pombe,* therefore provide a unique opportunity to dissect the two Ire1 functional outputs, which has remained an unsolved challenge in metazoans. Here, we set out to exploit this opportunity [19].

References

1. Walter, P. and D. Ron, *The unfolded protein response: from stress pathway to homeostatic regulation.* Science, 2011. **334**(6059): p. 1081-6.

2. Ron, D. and P. Walter, *Signal integration in the endoplasmic reticulum unfolded protein response.* Nat Rev Mol Cell Biol, 2007. **8**(7): p. 519-29.

3. Gardner, B.M. and P. Walter, *Unfolded proteins are Ire1-activating ligands that directly induce the unfolded protein response.* Science, 2011. **333**(6051): p. 1891-4.

4. Karagoz, G.E., et al., *An unfolded protein-induced conformational switch activates mammalian IRE1.* Elife, 2017. **6**.

5. Credle, J.J., et al., *On the mechanism of sensing unfolded protein in the endoplasmic reticulum.* Proc Natl Acad Sci U S A, 2005. **102**(52): p. 18773-84.

6. Aragon, T., et al., *Messenger RNA targeting to endoplasmic reticulum stress signalling sites.* Nature, 2009. **457**(7230): p. 736-40.

7. Lee, K.P., et al., *Structure of the dual enzyme Ire1 reveals the basis for catalysis and regulation in nonconventional RNA splicing.* Cell, 2008. **132**(1): p. 89-100.

8. Korennykh, A.V., et al., *The unfolded protein response signals through high-order assembly of Ire1.* Nature, 2009. **457**(7230): p. 687-93.

9. Cox, J.S., C.E. Shamu, and P. Walter, *Transcriptional induction of genes encoding endoplasmic reticulum resident proteins requires a transmembrane protein kinase.* Cell, 1993. **73**(6): p. 1197-206.

10. Mori, K., et al., *A transmembrane protein with a cdc2+/CDC28-related kinase activity is required for signaling from the ER to the nucleus.* Cell, 1993. **74**(4): p. 743-56.

11. Yoshida, H., et al., *XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor.* Cell, 2001. **107**(7): p. 881-91.

12. Calfon, M., et al., *IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA.* Nature, 2002. **415**(6867): p. 92-6.

13. Sidrauski, C., J.S. Cox, and P. Walter, *tRNA ligase is required for regulated mRNA splicing in the unfolded protein response.* Cell, 1996. **87**(3): p. 405-13.

14. Hollien, J. and J.S. Weissman, *Decay of endoplasmic reticulum-localized mRNAs during the unfolded protein response.* Science, 2006. **313**(5783): p. 104-7.

15. Hollien, J., et al., *Regulated Ire1-dependent decay of messenger RNAs in mammalian cells.* J Cell Biol, 2009. **186**(3): p. 323-31.

16. Kimmig, P., et al., *The unfolded protein response in fission yeast modulates stability of select mRNAs to maintain protein homeostasis.* Elife, 2012. **1**: p. e00048.

17. Moore, K. and J. Hollien, *Ire1-mediated decay in mammalian cells relies on mRNA sequence, structure, and translational status.* Mol Biol Cell, 2015. **26**(16): p. 2873-84.

18. Niwa, M., et al., *Genome-scale approaches for discovering novel nonconventional splicing substrates of the Ire1 nuclease.* Genome Biol, 2005. **6**(1): p. R3.

19. Li, W., et al., *Engineering ER-stress dependent non-conventional mRNA splicing.* Elife, 2018. **7**.

CHAPTER II: Engineering ER stress dependent non-conventional mRNA splicing

Results

In previous studies, the KR domain of *S. cerevisiae* Ire1 homolog was purified using an *E. coli* expression system [7, 8]. Thus, we adopted this expression/purification strategy to isolate *Sp* Ire1-KR. We designed an *Sp* Ire1-KR construct containing a Glutathione S-transferase (GST)-tag and a Human Rhinovirus (HRV) 3C protease cleavage site at the N-terminus of the *Sp* Ire1-KR (Fig. 1A). We used GST-tag because it has been reported to facilitate the folding and stabilization of many recombinant proteins [19]. We isolated Ire1-KR using the purification procedure illustrated in Figure 1B.

First, we looked for the induction condition that yielded the highest recombinant protein level. *Sp* Ire1-KR, which was under a T7 promoter, was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG). We examined the protein expression levels by western blotting at different induction temperatures, IPTG concentrations and induction durations. We tested 25°C or 18°C for induction temperature because low induction temperature was reported to increase *S. cerevisiae* Ire1-KR yield [8]. The highest protein expression level was achieved using 0.5 mM IPTG for 4 hours at 25°C (Fig. 1C, lane 1).

For protein purification, we expressed *Sp* Ire1-KR at this optimized condition, harvested and lysed the cells. The cell lysate was subject to GST-affinity chromatography. In the column elution, we observed two dominant bands. Based on their molecular weight, they corresponded to GST-Ire1-KR (75 kDa) and free GST (26 kDa) (Fig. 1D, lane 2). Next, we added GST-tagged HRV 3C protease to cleave off the GST tag from GST-Ire1-KR and dialyzed the sample. After 16 h of protease incubation and dialysis, the sample contained

three dominant protein species: GST-Ire1-KR, Ire1-KR (49 kDa) and GST (Fig. 1D, lane 3). The cleavage products were subject to a second round of GST-affinity chromatography, which acted as negative chromatography to isolate the Ire1-KR without the GST-tag in the flow through (Fig. 1E). However, when we measured the UV absorbance of the sample, we noticed that the ratio of absorbance at 260 and 280 nm ($A_{260/280}$) was around 0.85, which was higher than $A_{260/280}$=0.57 for pure protein and indicated nucleic acid contamination [20]. To remove the possible nucleic acid contamination that might interfere with the subsequent *in vitro* assays, we applied the sample to anion exchange chromatography such that Ire1-KR would not bind to the anion exchange column while most of the nucleic acids do (see Materials and Methods). The $A_{260/280}$ was reduced to 0.62, indicating removal of the majority of the contaminating nucleic acids (Fig. 1F). Finally, to further purify the protein, we applied the sample to a gel filtration column (Superdex 200 16/60, GE Healthcare). We observed mainly three peaks in the elution profile (Fig. 1G). The high molecular weight fraction in the void volume of the gel filtration column (<50 ml) separated most of the impurities and aggregated proteins (Fig. 1H, lane 2, 3). The Ire1-KR oligomers eluted as the second fraction from the gel filtration column (Fig. 1H, lane 5-7). After concentrating the purified protein, we found that this fraction also contained a residue amount of uncleaved GST-Ire1-KR (Fig. 1I, lane 1). The *Sp* Ire1-KR dimers/monomers (Fig. 1H, lane 8-11) eluted as the third fraction (elution volume from 86 to 94 ml). This fraction contained Ire1-KR with high purity (>98%) (Fig. 1I, lane 2). We only used the dimer/monomer fraction for further characterization.

The *Sp* Ire1-KR protein yield was low, with 0.31 μg per liter culture. In order to increase the protein yield, we tried different expression hosts, including *S. cerevisiae* and *S. pombe* cells. However, both of these expression systems also gave low yields, similar to the *E. coli* expression system. Among the three expression hosts, *E. coli* has the shortest doubling time and is the easiest to handle, so we chose to optimize protein expression/purification for the *E. coli* expression system.

Low protein yield usually results from the recombinant protein accumulating in insoluble aggregates, or it being poorly expressed. To distinguish these two possibilities, we extracted total and soluble protein fractions, and examined them on a Coomassie-blue-stained SDS-PAGE gel (Fig. 2A). At the expected molecular weight (76 kDa), no visible bands appeared upon IPTG induction in neither total nor soluble protein fractions. This result indicated that the expression level of *Sp* Ire1-KR was very low in *E. coli*, which may lead to the low protein yield. The expression level of recombinant protein is influenced by multiple factors, including transcriptional/translational efficiency and recombinant protein's toxicity to host cells. Because we used a very strong promoter in an *E. coli* expression strain (BL21-CodonPlus-RIPL strain) that is optimized for translating recombinant proteins, we reasoned that transcription/translation efficiency was less likely to be the bottleneck. So we tested whether the recombinant protein was toxic to the host cells. To this end, we compared two *E. coli* strains, which were transformed either with an empty plasmid or a plasmid containing GST-Ire1-KR. It has been reported that T7 promoters are leaky in the absent of IPTG [21], leading to basal level expression of *Sp* Ire1-KR protein. To test whether the basal level of *Sp* Ire1-KR was toxic to the *E. coli* cells, we

inoculated the *E. coli* strains for 5 h or overnight (~16 h) in selective liquid medium with no IPTG, then plated the same number of cells (as determined by optical density OD_{600}) onto agar plates and measured the colony-forming units (CFUs). When the inoculation time was short (5 h), the CFUs were comparable between the Ire1-containing strain and the control strain, but the colonies of the Ire1-containing strain were smaller in size, about 1/3 in average (Fig. 2B, C, D), indicating that *Sp* Ire1-KR slowed down the growth of *E. coli* cells. When the inoculation time was long (overnight), the CFU for the Ire1-contaning strain was nearly 30 times fewer than that of the empty-plasmid-containing strain, showing a more prominent toxic effect from *Sp* Ire1-KR (Fig. 2B, C). Similarly, we found that the overexpressed *Sp* Ire1-KR was also toxic to *S. cerevisiae* and *S. pombe* cells, suggesting the protein's cytotoxicity lead to its low yields in multiple expression systems (Supplementary Figure 1A, B).

Because the toxic effect of *Sp* Ire1-KR in *E. coli* was less prominent when inoculation time was short (5 h), we optimized the protein induction procedure by shortening the liquid inoculation time before IPTG induction. In the previous procedure, a long incubation (~16 h) was performed in order to generate desired cell mass (12 l culture with OD_{600} around 1) from a single transformation colony. To shorten the incubation time while not compromising the final cell mass, we started the liquid culture with more cells by scraping off and collecting all the colonies on the transformation plate. In this way, only 5 h of incubation was needed to achieve the same cell mass (Fig. 2E). Using this optimized protein induction procedure, we increased the protein yield by about 100 folds (Fig. 2F), producing sufficient amounts of *Sp* Ire1-KR for *in vitro* biochemical assays.

To test *Sp* Ire1 RNase activity, we performed *in vitro* RNA cleavage assay using a 5'-radio labeled, 27-nucleotide RNA hairpin, which was derived from an endogenous *Sp* Ire1 cleavage site on *BIP1* mRNA. *Sp* Ire1-KR specifically cleaved this RNA substrate *in vitro*, with $k_{obs} = 21.1 \pm 3.6 \times 10^{-4}$ (Fig. 3A, B). Previous studies reported that the RNase activity of *S. cerevisiae* Ire1 can be further enhanced by cofactor ADP that binds to nucleotide binding pocket in the Ire1 kinase domain [22]. Similarly, when ADP was added, the RNase activity of *Sp* Ire1-KR increased by nearly 2 fold, with $k_{obs} = 38.4 \pm 10.6 \times 10^{-4}$ s^{-1} (Fig. 3A, B). Therefore, the recombinant *Sp* Ire1-KR specifically cleaved its endogenous RNA substrate *in vitro*. This RNase activity of *Sp* Ire1 can be further enhanced by cofactor ADP.

Discussion

In this study, we report an efficient protocol to purify the kinase/RNase domains of *S. pombe* Ire1, which is a unique Ire1 homolog specialized for RIDD. We found that *Sp* Ire1-KR recombinant protein is cytotoxic to multiple expression hosts, including *E. coli*, *S. cerevisiae* and *S. pombe*. Consequently, the protein yield was low, hindering its *in vitro* biochemical characterizations. To this end, we optimized the protein induction procedure in *E.* coli, and increased the protein yield by nearly 100 folds. The purified recombinant *Sp* Ire1-KR specifically cleaved its cognate RNA substrates *in vitro*. Such RNA cleavage efficiency was further enhanced by cofactor ADP.

A number of strategies/tools for purifying toxic proteins have been previously described, including, but not limited to, targeting toxic proteins to periplasm, preventing plasmid loss using plate inoculation or more stringent antibiotic selections, turning off T7 promoter leaky expression by adding glucose to the medium or using an *E. coli* strain engineered for this purpose (BL21-CodonPlus-RIPL pLysE) [23, 24]. Our optimized protein expression procedure enriched this toolbox for expressing and purifying cytotoxic recombinant proteins. One limitation of our expression procedure is for recombinant proteins that are highly toxic that cells cease growing upon taking up the plasmid and no transformation colonies are observed. In this case, a combination of different tools, like combining our optimized expression procedure with BL21-CodonPlus-RIPL pLysE system, can be used to improve protein yield. This work laid the ground for further studies on the biochemical features that specialized *S. pombe* Ire1 to RIDD, and thus understanding the mechanistic distinctions between the two Ire1 functional outputs: non-conventional mRNA splicing and RIDD.

Materials and Method

Purification of *S. pombe* Ire1-KR

S. pombe Ire1-KR (653L to 1072Y) was cloned into the bacterial expression plasmid pGEX-6p-2 using BamHI and SalI restriction sites. In the final cloning product, the *S. pombe* Ire1-KR was fused to an N-terminal GST-tag and is being regulated by a T7 promoter. This expression plasmid was transformed into BL21-CodonPlus-RIPL competent bacteria cells (Agilent Technologies). The transformed cells were incubated on an ampicillin-containing lysogeny broth (LB) agar plate for 16 hours at 37° C. Then, all the colonies on the transformation plate were scraped off, mixed and inoculated in 50 ml ampicillin-containing LB medium. After 3-hour incubation at 37° C, the culture was diluted to 12 L with ampicillin-containing LB medium and was further incubated at 37° C (for about 2-4 hours) till OD_{600} reached 1. Next, 0.5 mM IPTG was added, and the culture was incubated at 25° C for 4 hours. Cell were pelleted by centrifugation and resuspended with GST binding buffer (50 mM Tris/HCl pH 7.5, 500 mM NaCl, 2 mM $Mg(OAc)_2$, 2 mM DTT, 10% Glycerol). The cells were homogenized using high-pressure homogenizer (EmulsiFlex, Avestin). The cell lysate was applied to a GST affinity chromatography column (GSTrap 5 ml FF, GE Healthcare) and was eluted with GST elution buffer (50 mM Tris/HCl pH 7.5, 200 mM NaCl, 2 mM $Mg(OAc)_2$, 2 mM DTT, 10% Glycerol, 10 mM glutathione). Then, GST-tagged HRV 3C protease (PreScission Protease, GE Health) was added to the column elution to remove the GST tag. Meanwhile, the sample was dialyzed in the GST binding buffer using Slide-A-Lyzer (Thermo Fisher Scientific). The dialysis and protease treatment was conducted at 4° C for over night (about 16 hours). Next, the sample (in the GST binding buffer) was applied to an

anion exchange chromatography column (HiTrap Q HP, GE Healthcare) and a GST affinity chromatography column (GSTrap 5 ml FF, GE Healthcare). *Sp* Ire1-KR was collected in the column flow through in both cases. To further purify the recombinant protein and replace the buffer with storage buffer (50 mM Tris/HCl pH 7.5, 200 mM NaCl, 2 mM $Mg(OAc)_2$, 2 mM TCEP, 10% Glycerol), *Sp* Ire1-KR was applied to a gel filtration column (superdex 200 hiload 16/60, GE Healthcare). The Ire1-KR dimer/monomer peak in the gel filtration column elution (elution volume from 86 to 94 ml) was harvested, further concentrated to 14 μM and flash frozen in liquid nitrogen.

Colony forming assay

E. coli cell cultures were inoculated from colonies on fresh transformation plates and incubated in 4 ml LB medium containing ampicillin at 37° C for 5 or 16 hours. Their OD_{600} were measured, and about 250 cells were plated onto each ampicillin-containing LB agar plate. The plates were incubated at 37° C for about 20 hours and were photographed. The sizes and numbers of colonies were measured using imageJ.

Ire1 *in vitro* RNA cleavage assay

The Ire1 *in vitro* cleavage assay was previously described in detail [25]. Briefly, short RNA oligos (Dharmacon, Inc) was gel purified and 5' radio labeled. Then the RNA oligos were further cleaned (ssDNA/RNA Clean and Concentrator kit, Zymo Research D7010) and folded. In the Ire1 cleavage assays, the reaction samples contained 6 μM of *S. pombe* Ire1-KR. The cleavage reaction was performed at 30° C in reaction buffer (50 mM Tris/HCl pH 7.5, 200 mM NaCl, 2 mM $Mg(OAc)_2$, 2 mM TCEP, 10% Glycerol). At each time point, 0.75 μl

of the reaction sample was transferred to 5 µl STOP buffer (10 M urea, 0.1% SDS, 1 mM EDTA, 0.05% xylene cyanol, 0.05% bromophenol blue). Next, the cleaved and uncleaved RNAs were separated with electrophoresis using denaturing 15% urea-PAGE gels. Gels were imaged with a Phosphorimager (Typhoon FLA 9500, GE Health) and quantified using imageJ. The cleaved portion was the amount of the cleaved RNA divided by the sum of uncleaved and cleaved RNA. The hairpin RNA substrate was derived from the Ire1 cleavage site on *S. pombe BIP1* mRNA, and has the sequence of CGCGAGAUAACUGGUGCUUUGUUAUCUCGCG.

Immunoblotting

Proteins were extracted from *E. coli* cells using BugBuster protein extraction reagent (Novagen). Samples were boiled, separated using electrophoresis and then transferred to nitrocellulose. Anti-GST antibody (abcam ab19256) was used to probe the GST-Ire1-KR.

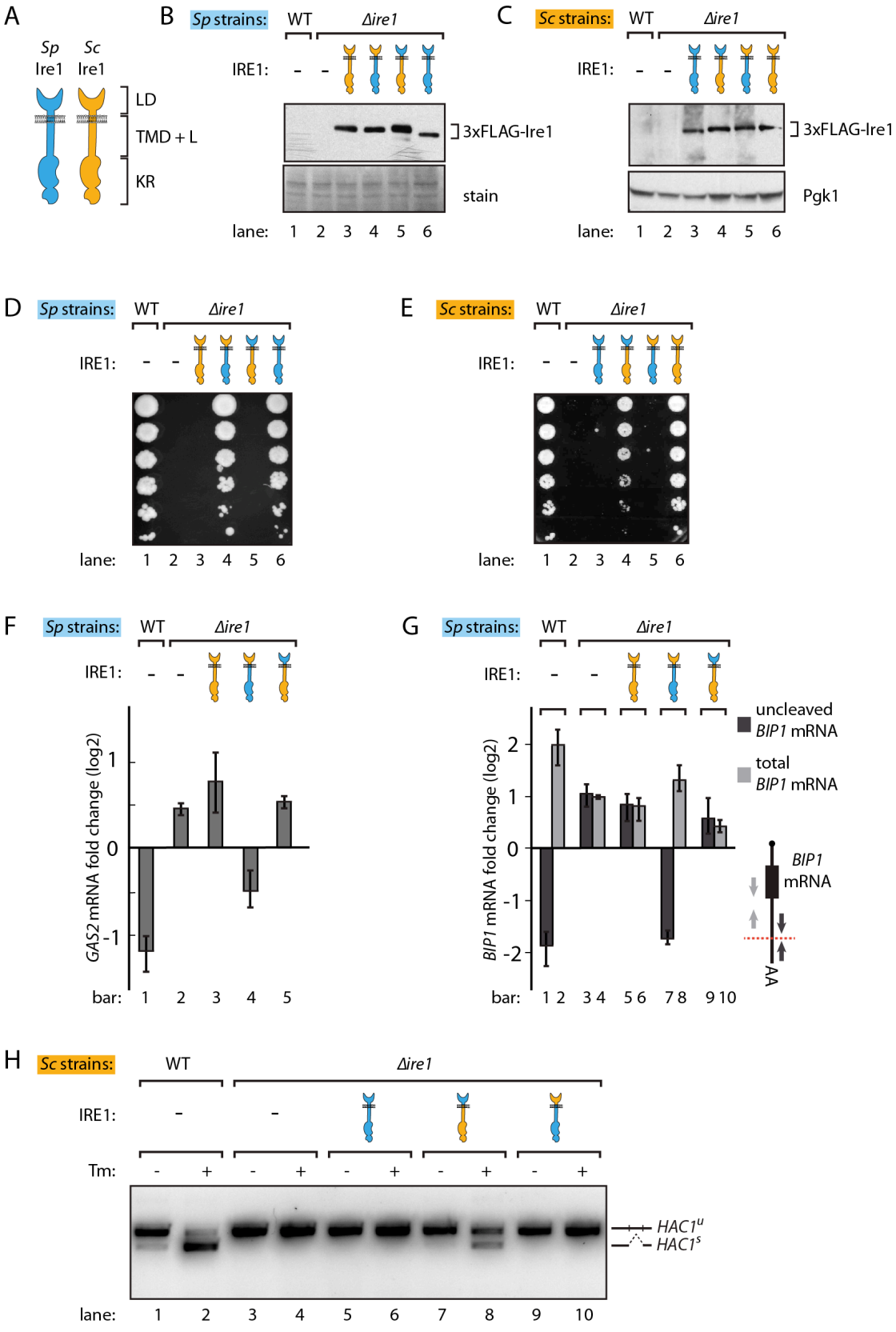
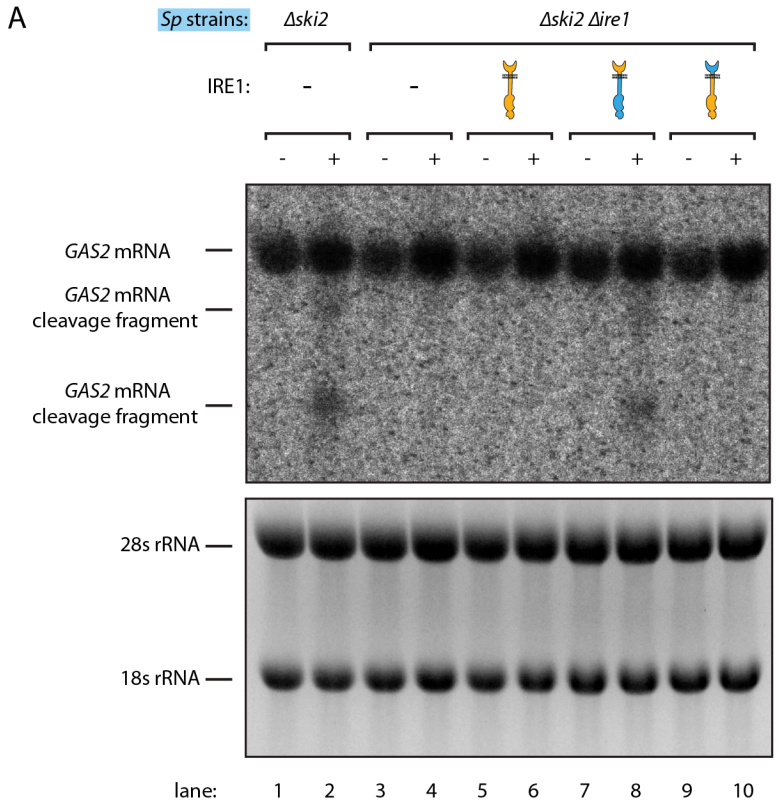

A
Sp Ire1
Sc Ire1
LD
TMD + L
KR
B
Sp strains:
WT
Δire1
IRE1:
3xFLAG-Ire1
stain
lane: 1 2 3 4 5 6
C
Sc strains:
WT
Δire1
IRE1:
3xFLAG-Ire1
Pgk1
lane: 1 2 3 4 5 6
D
Sp strains:
WT
Δire1
IRE1:
lane: 1 2 3 4 5 6
E
Sc strains:
WT
Δire1
IRE1:
lane: 1 2 3 4 5 6
F
Sp strains:
WT
Δire1
IRE1:
GAS2 mRNA fold change (log2)
bar: 1 2 3 4 5
G
Sp strains:
WT
Δire1
IRE1:
BIP1 mRNA fold change (log2)
uncleaved BIP1 mRNA
total BIP1 mRNA
BIP1 mRNA
AA
bar: 1 2 3 4 5 6 7 8 9 10
H
Sc strains:
WT
Δire1
IRE1:
Tm:
HAC1u
HAC1s
lane: 1 2 3 4 5 6 7 8 9 10

Figure 1 *S. pombe* and *S. cerevisiae* Ire1 have functionally conserved stress sensing ER-lumenal domains and divergent cytosolic domains.

(A) Cartoon illustration of lumenal domain (LD), transmembrane/cytosolic linker domain (TMD+L) and kinase/RNase domain (KR) for *S. pombe* (*Sp*) (*blue*) and *S. cerevisiae* (*Sc*) Ire1 (*orange*). (B, C) Expression levels of *S. cerevisiae* Ire1 (128 kD), *S. cerevisiae* lumenal *S. pombe* cytosolic Ire1 (126 kD), *S. pombe* lumenal *S. cerevisiae* cytosolic Ire1 (125 kD) and *S. pombe* Ire1 (122 kD) in *S. pombe* (B) and *S. cerevisiae* cells (C). Extracts were immunoblotted for 3xFLAG-Ire1. Ponceau stain (B) or Pgk1 (C) was used as loading control. (D, E) Cell growth assay on tunicamycin (Tm) plates. Serial dilutions of *S. pombe* (D) or *S. cerevisiae* (E) cells, which expressed the indicated Ire1 constructs, were spotted onto plates containing 0.05 μg/ml (B) or 0.1 μg/ml (C) of Tm. Plates were photographed after incubation at 30° C for 4 days. (F, G) qPCR assay for *S. pombe GAS2* (F) or *BIP1* (G) mRNA fold change upon 1 μg/ml Tm treatment for 1 h. Experiments were done in triplicates. In (G), uncleaved (*dark grey*) or total (*light grey*) *BIP1* mRNA was detected using the corresponding PCR primers illustrated as arrows in the schematic insert. The red dashed line indicates the Ire1 cleavage position on *BIP1* mRNA. (H) Detection of *S. cerevisiae HAC1* mRNA splicing by RT-PCR across the splice junction. Cells were treated with or without 1 μg/ml of Tm for 1 h.


A
Sp strains:
Δski2
Δski2 Δire1
IRE1:
- + - + - + - + - +
GAS2 mRNA
GAS2 mRNA cleavage fragment
GAS2 mRNA cleavage fragment
28s rRNA
18s rRNA
lane: 1 2 3 4 5 6 7 8 9 10

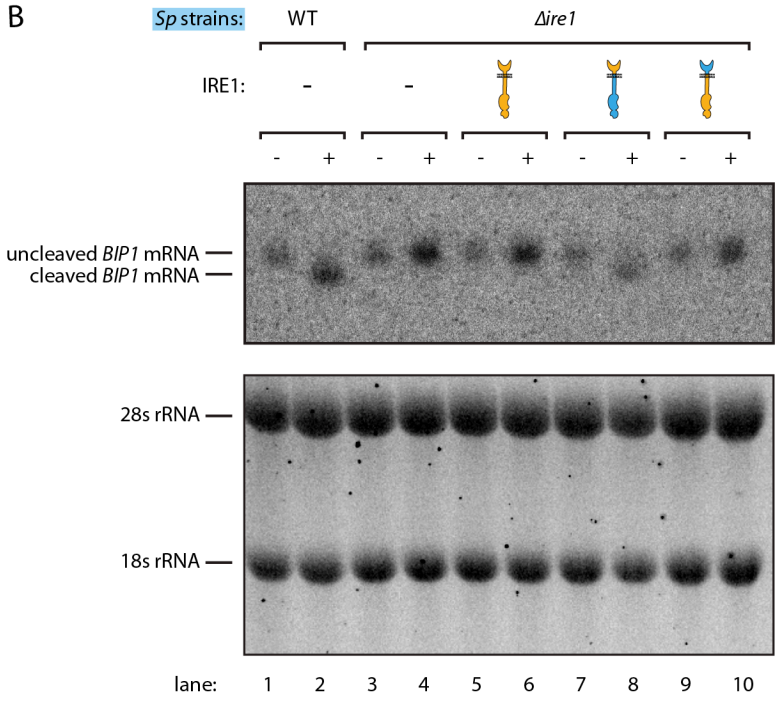

B
Sp strains:
WT
Δire1
IRE1:
- + - + - + - + - +
uncleaved BIP1 mRNA
cleaved BIP1 mRNA
28s rRNA
18s rRNA
lane: 1 2 3 4 5 6 7 8 9 10

Figure 1-figure supplement 1 Ire1 chimeras with *S. pombe* cytosolic domain cleave *BIP1* and *GAS2* mRNA in *S. pombe*.

Northern blots of *S. pombe GAS2* (A) and *BIP1* (B) mRNA. Cells were treated with 1 μg/ml of Tm for 1 h.

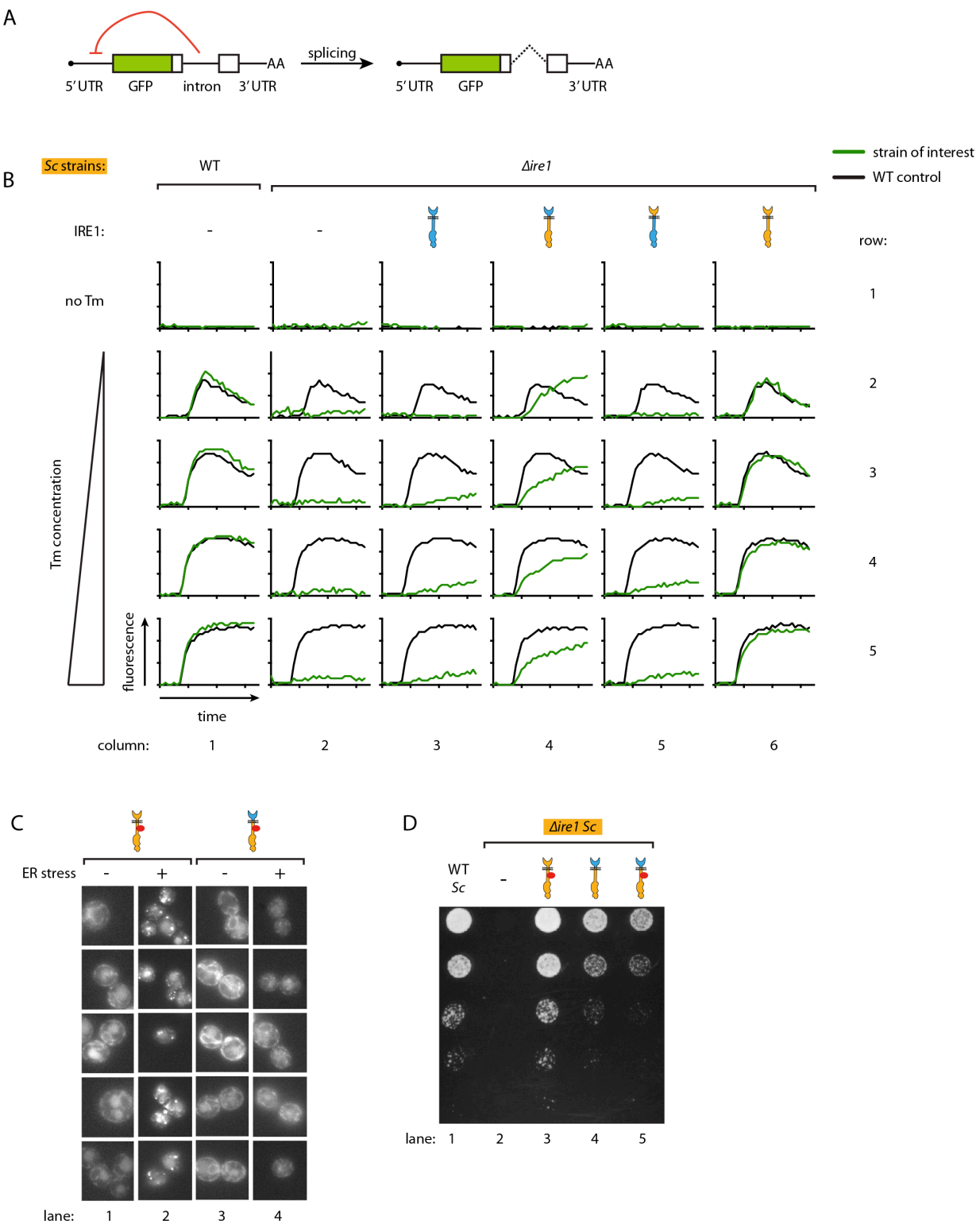


Figure 1-figure supplement 2 Ire1 oligomeric state determines the *HAC1* mRNA splicing dynamics in *S. cerevisiae* cells.

(A) Illustration of the *HAC1* mRNA derived splicing reporter. The splicing reporter contains a part of the *HAC1* mRNA 5' exon replaced by the green fluorescent protein (GFP) coding sequence. In the unspliced reporter mRNA, translation is inhibited by a translation block formed by the intron and 5'UTR (indicated by the red arrow). Upon splicing, intron is removed and translation begins. (B) Measuring the *HAC1* mRNA splicing dynamics in *S. cerevisiae* using automated flow cytometry. After 1.5 h of incubation, either no Tm or 0.25 μg/ml, 0.5 μg/ml, 1 μg/ml, 2 μg/ml of Tm was added. Then, we monitored the splicing dynamics for 10 h. The splicing dynamics under various conditions is plotted. *Green* lines represent the strains of interest, which expressed indicated Ire1 variants, and the *black* line represents WT control strain under the same condition. (C) Examining Ire1 foci formation in *S. cerevisiae* cells via fluorescence microscopy with or without 1 μg/ml Tm treatment for 20 min. (D) Growth assay on Tm plate for *S. cerevisiae* cells expressing Ire1 constructs with or without mCherry inserted into the cytosolic linker. The inserted mCherry does not affect Ire1's ability to alleviate ER stress.

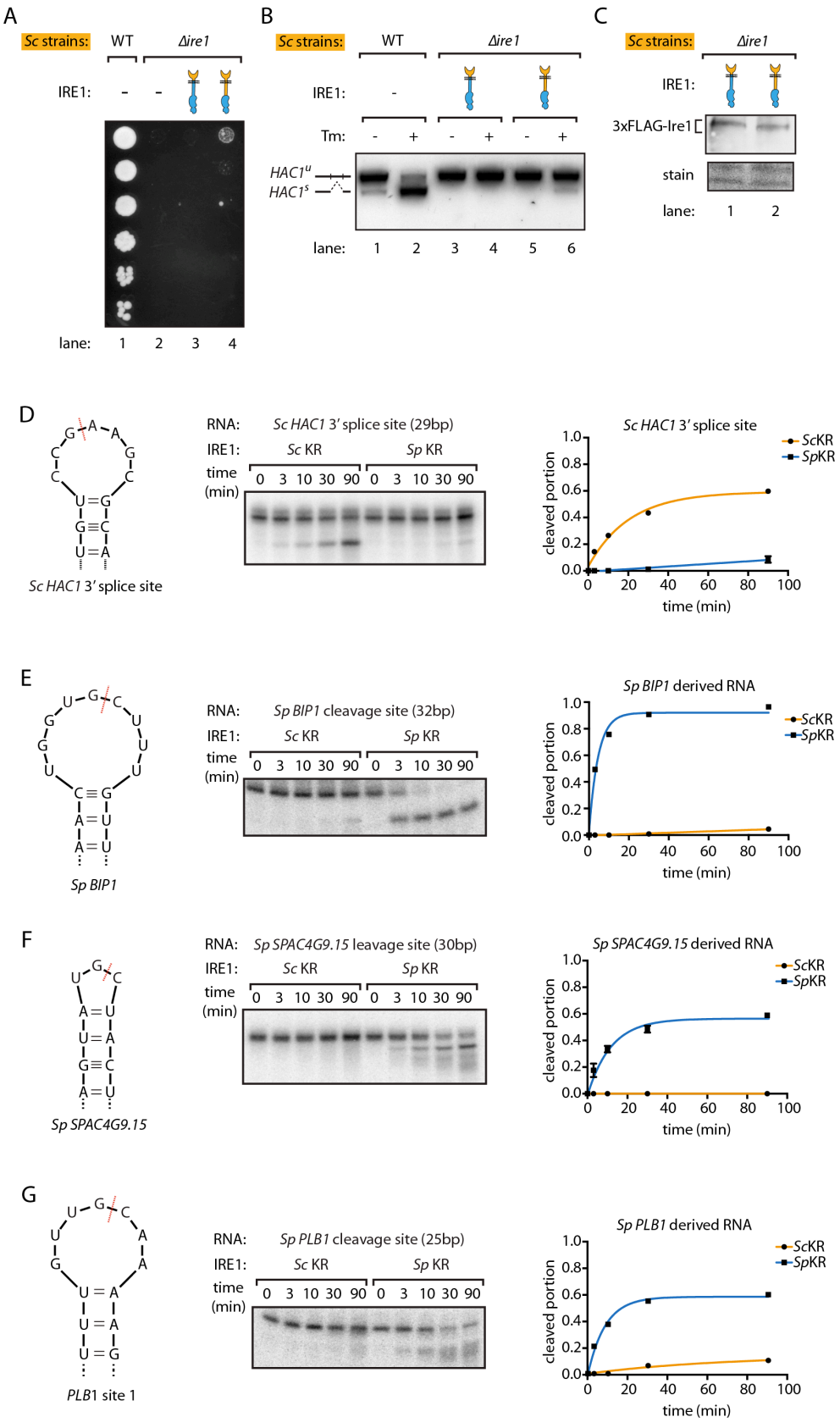

A
Sc strains:
WT
Δire1
IRE1:
lane: 1 2 3 4
B
Sc strains:
WT
Δire1
IRE1:
Tm:
HAC1u
HAC1s
lane: 1 2 3 4 5 6
C
Sc strains:
Δire1
IRE1:
3xFLAG-Ire1
stain
lane: 1 2
D
Sc HAC1 3′ splice site
RNA: Sc HAC1 3′ splice site (29bp)
IRE1: Sc KR Sp KR
time (min) 0 3 10 30 90 0 3 10 30 90
Sc HAC1 3′ splice site
ScKR
SpKR
cleaved portion
time (min)
E
Sp BIP1
RNA: Sp BIP1 cleavage site (32bp)
IRE1: Sc KR Sp KR
time (min) 0 3 10 30 90 0 3 10 30 90
Sp BIP1 derived RNA
ScKR
SpKR
cleaved portion
time (min)
F
Sp SPAC4G9.15
RNA: Sp SPAC4G9.15 leavage site (30bp)
IRE1: Sc KR Sp KR
time (min) 0 3 10 30 90 0 3 10 30 90
Sp SPAC4G9.15 derived RNA
ScKR
SpKR
cleaved portion
time (min)
G
PLB1 site 1
RNA: Sp PLB1 cleavage site (25bp)
IRE1: Sc KR Sp KR
time (min) 0 3 10 30 90 0 3 10 30 90
Sp PLB1 derived RNA
ScKR
SpKR
cleaved portion
time (min)

Figure 2 *S. pombe* and *S. cerevisiae* Ire1 have distinct RNase specificity.

(A) Growth assay for *S. cerevisiae* cells expressing indicated Ire1 constructs on Tm plates, as Figure 1E. (B) Measuring *HAC1* mRNA splicing, as Figure 1H. (C) Comparing the expression levels of the indicated 3xFLAG-tagged Ire1 chimeras using immunoblotting. Ponceau stain was used as loading control. (D, E, F, G) *In vitro* RNA cleavage assays. 5'-radiolabeled hairpin RNA substrates were incubated with 12.5 μM *S. cerevisiae* or *S. pombe* Ire1 kinase/RNase domains (KR) at 30° C for the indicated time. (D) Hairpin RNA substrate derived from the 3' splice site of *S. cerevisiae HAC1* mRNA. The calculated k_{obs} is 9.4 ± 0.9 × 10^{-4} s^{-1} for *S. cerevisiae* Ire1 KR and 0.15 ± 0.01 × 10^{-4} s^{-1} for *S. pombe* Ire1 KR. (E) Hairpin RNA substrate derived from the Ire1 cleavage site on *S. pombe BIP1* mRNA. The calculated k_{obs} is 0.079 ± 0.0006 × 10^{-4} s^{-1} for *S. cerevisiae* Ire1 KR and 37.3 ± 4.4 × 10^{-4} s^{-1} for *S. pombe* Ire1 KR. (F) Hairpin RNA substrate derived from the Ire1 cleavage site on *S. pombe SPAC4G9.15* mRNA, encoding a gene of unknown function. The calculated k_{obs} was below our detection limit for *S. cerevisiae* Ire1 KR and 15.6 ± 2.2 × 10^{-4} s^{-1} for *S. pombe* Ire1 KR. (G) Hairpin RNA substrate derived from the Ire1 cleavage site on *S. pombe PLB1* mRNA. The calculated k_{obs} is 0.2 ± 0.003 × 10^{-4} s^{-1} for *S. cerevisiae* Ire1 KR and 19.0 ± 2.5 × 10^{-4} s^{-1} for *S. pombe* Ire1 KR.

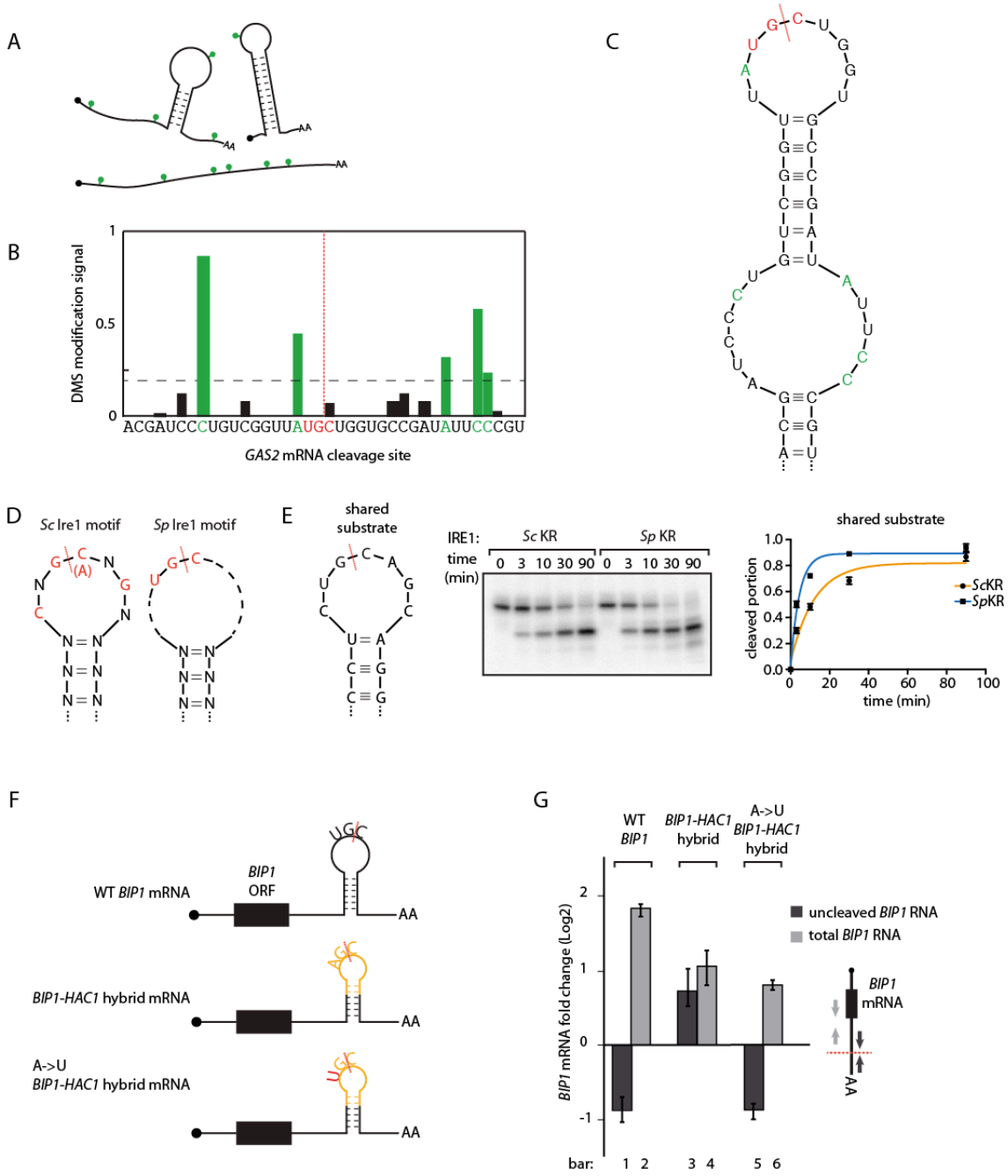


Figure 3 *S. pombe* and *S. cerevisiae* Ire1 recognize distinct RNA sequence and structural features.

(A) Illustration of RNA structural mapping by DMS modifications. Dimethyl sulfate (DMS) allows detection of unpaired adenine and cytosine RNA bases (*green* dots). (B) The normalized DMS modification signals near the Ire1 cleavage site on *S. pombe GAS2* mRNA (cleavage site is indicated by the *red* dashed line). The positions with high DMS

modification signals are labeled in *green* and the previously identified *S. pombe* Ire1 UG|C motif is labeled in *red*. (C) *In sillico* RNA secondary structure prediction of the Ire1 cleavage site on *GAS2* mRNA. Structure prediction was constrained by forcing the positions with high DMS modification signals (*green*) to be unpaired. (D) RNA sequence and structural motifs recognized by the *S. cerevisiae* and *S. pombe* Ire1. (E) *In vitro* cleavage assay using an RNA hairpin derived from human *XBP1* mRNA 3' splice site, which is predicted to be a shared substrate for *S. cerevisiae* and *S. pombe* Ire1 KR. The calculated k_{obs} is 16.7 ± 2.3 × 10^{-4} s^{-1} for *S. cerevisiae* Ire1 KR and 38.9 ± 4.0 × 10^{-4} s^{-1} for *S. pombe* Ire1 KR. (F) Illustrations of the *S. pombe BIP1* mRNA variants and (G) their uncleaved (*dark grey*) or total (*light grey*) mRNA fold change upon ER stress in *S. pombe* cells. Experiments were done in triplicates.

A

systematic ID	gene name	Ire1 cleavage site sequence
SPBC29A10.08	*GAS2*	ACGAUCCCUGUCGGUUAUGCUGGUGCCGAUAUUCCCGU
SPAC22A12.15c	*BIP1*	AACACCUAGUUAACUGGUGCUUUGUUAUCUUUGUAUUG
SPAC26A3.01	*SXA1*	AUUAUCGGUUUGUCAGUUGCCAUGACUAUUACUGGUAU
SPBC25H2.06c	*HRF1*	GAGCUAUUUGGCCUUCGUGCUAGUAAGGCUUGUGCUGU
SPAC1A6.04c	*PLB1*	UCCUUUGUGGCCUUUGUUGCAAAAGGGUCGUGAUGUCG
SPBP4G3.02	*PHO1*	AUUUGCAGUUAUGAAAUUGCCCUUCAAGACUAUAGCGA
SPCC970.03		CAUUUAUCGUUUCCGUCUGCCUCGCGCUAACGAUGUUC
SPBC29A10.08	*GAS2*	GCUGUCGCUUACGUUCGUGCUGCCGUUCGUGAUUCCAA
SPAC1A6.04c	*PLB1*	GUUGCCGAAAAGGCCAAUGCUGGCUUUAACACUCAGUCU
SPBP4G3.02	*PHO1*	AUCUUUGGAGGUGCCUAUGCUAAUAGCCUUGCAAAUUC
SPBC3D6.02	*BUT2*	CUUAGUGCUGACAUUCCUGCCAAGUCUAGCCGUCUCUU
SPAC56F8.07		CUUUUACUACGACAUGGUGCUGCAUGUUCGAGUUAUUU
SPCC830.08c	*YOP1*	GCCUUCUUUAGUAUCAAUGCUAUUGAAACUACUAACAA
SPAC4G9.15		GGAAUCGGUAAAGAGUAUGCUACUCAAUUAGCCAUGUC

B

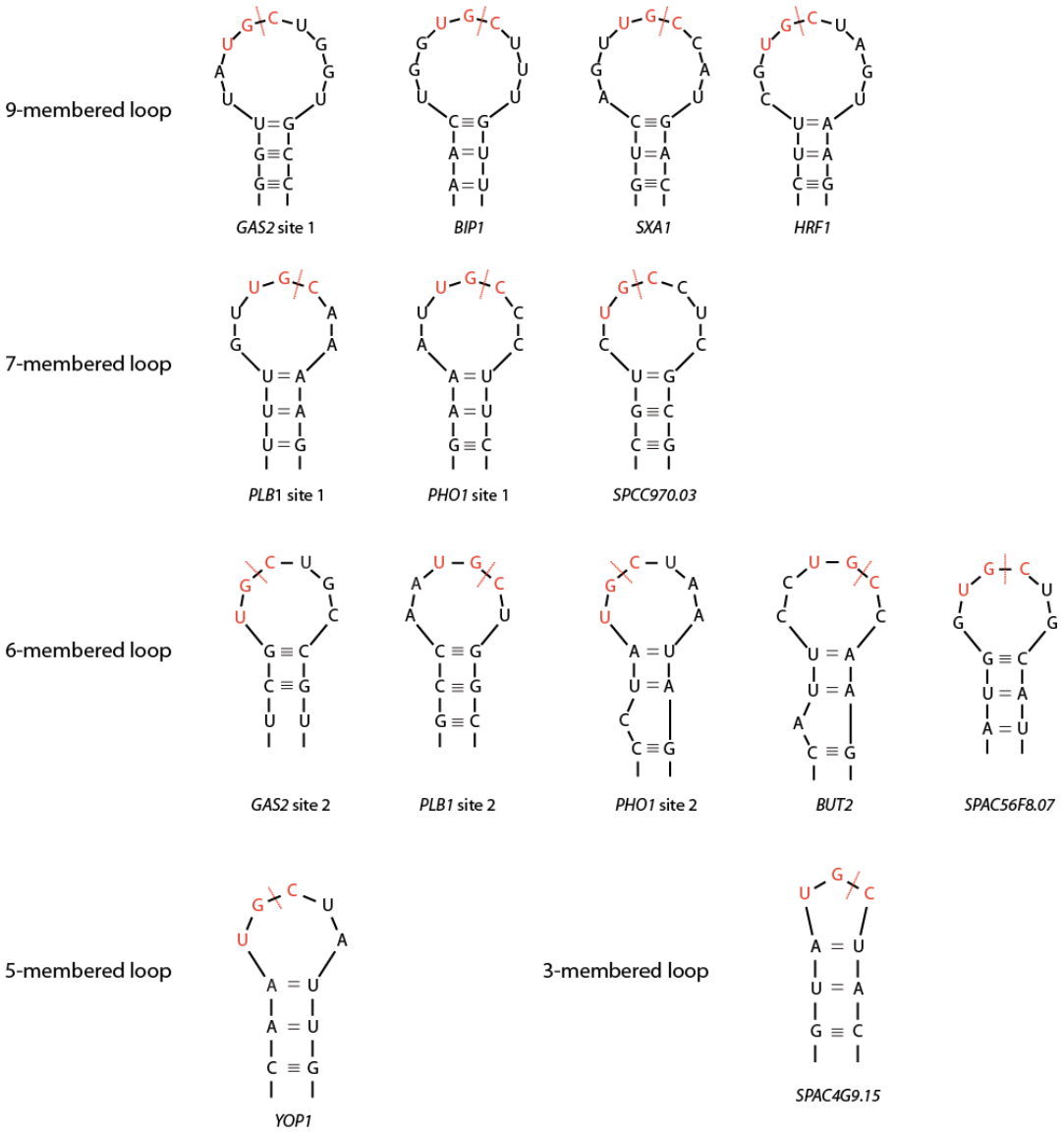


Figure 3-figure supplement 1 *S. pombe* Ire1 cleaves at UG|C positioned near the center of loops in RNA stem-loop structures.

(A) A list of all 14 *S. pombe* Ire1 mRNA cleavage sites, which were independently identified by both Kimmig, Diaz et al. and Guydosh et al.. The UG|C motifs are labeled in *red*. The positions with high DMS modification signals are labeled in *green*. (B) Predicted RNA secondary structures of *S. pombe* Ire1 cleavage sites. DMS modification signals were used to guide the secondary structure prediction of *S. pombe* Ire1 mRNA cleavage sites. The *red* dashed lines indicate the Ire1 cleavage sites.

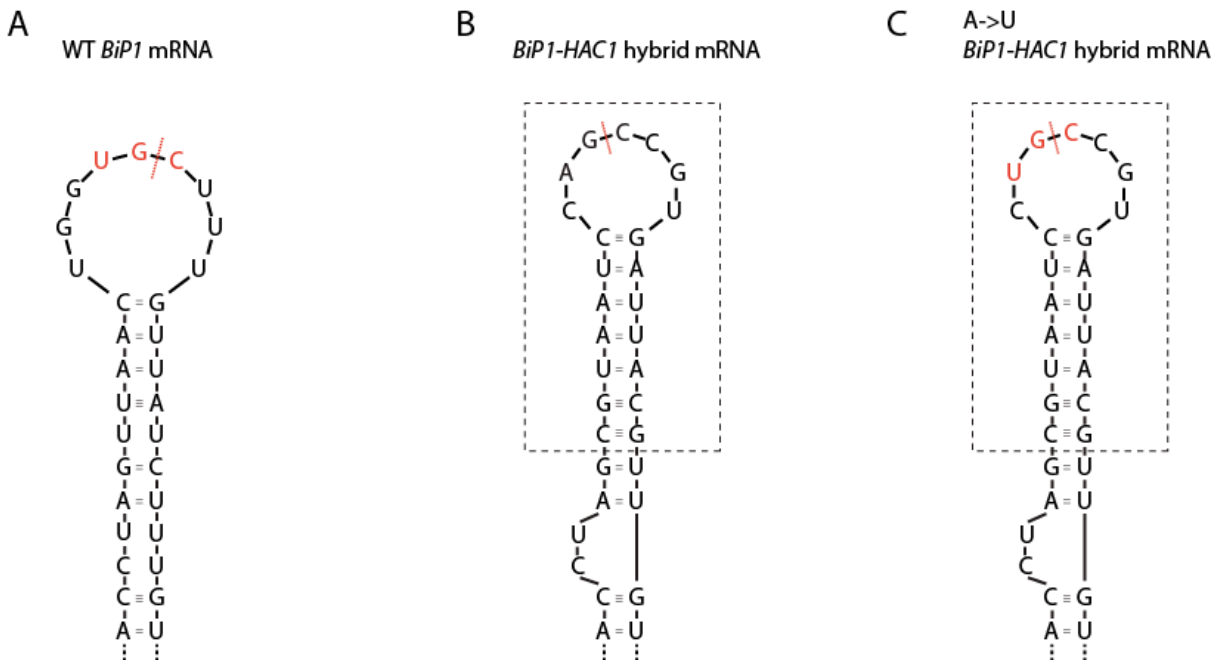


Figure 3-figure supplement 2 Ire1 cleavage sites on *BIP1* mRNA variants.

Sequence and predicted RNA secondary structures of Ire1 cleavage sites on (A) *S. pombe BIP1* mRNA, (B) *BIP1-HAC1* hybrid mRNA and (C) *BIP1-HAC1* hybrid mRNA with an A to U mutation. The part included in the dashed box is derived from *S. cerevisiae HAC1* mRNA 5' splice site.

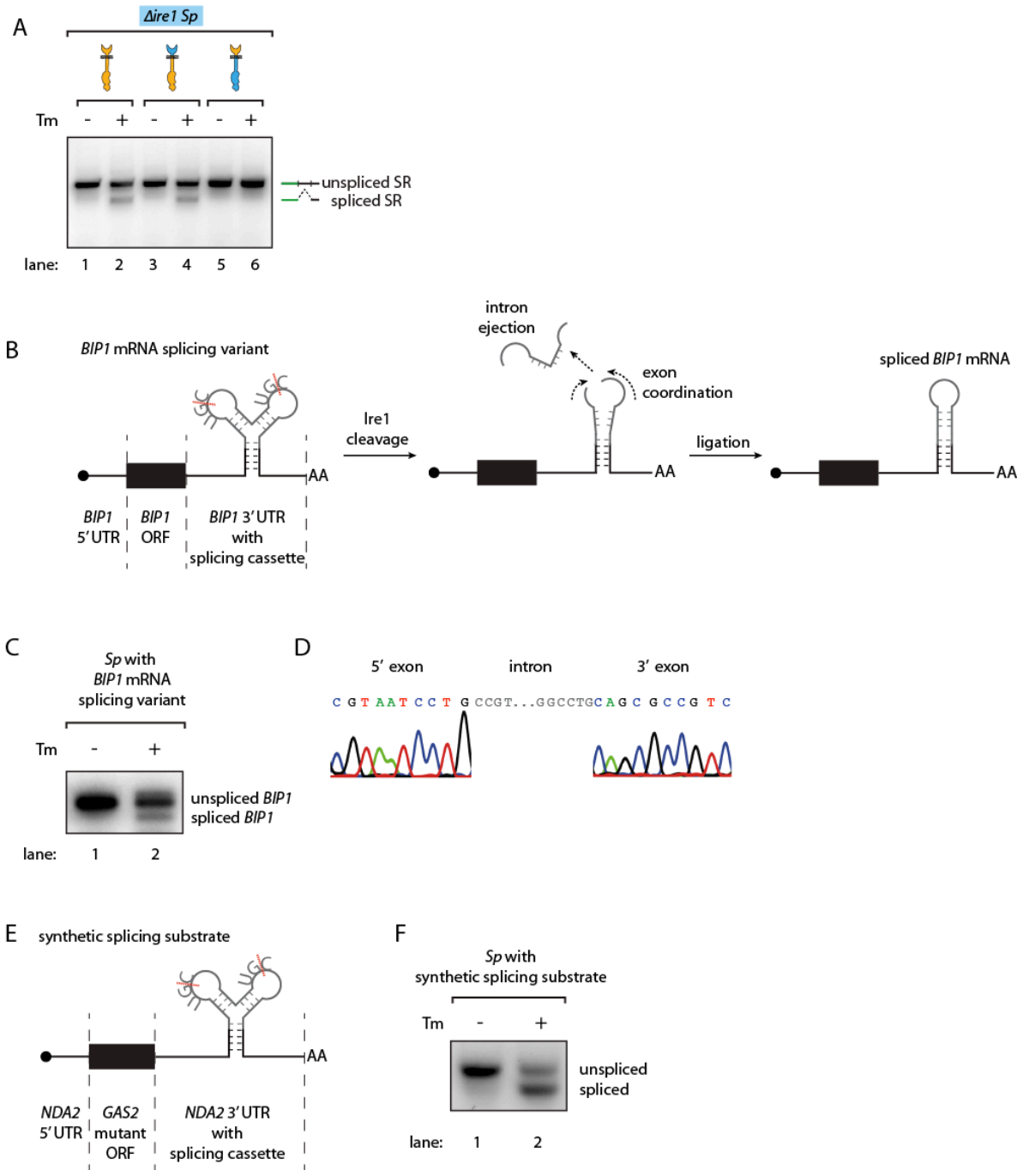


Figure 4 Engineering the Ire1-mediated non-conventional mRNA splicing in *S. pombe* cells.

(A) Measuring the non-conventional mRNA splicing in *S. pombe* cells, which were transformed with the *S. cerevisiae HAC1* mRNA splicing reporter and the indicated Ire1 constructs. Cells were treated with 1 μg/ml Tm for 1 h. (B) Illustration of the engineered *S. pombe BIP1* mRNA splicing variant. (C) Measuring the nonconventional mRNA splicing of the engineered *S. pombe BIP1* mRNA splicing variant. Experimental conditions are the same

as those for Figure 4A. (D) Sequencing reads of the spliced *BIP1* mRNA. The schematic illustrations (E) and the splicing assays (F) of the synthetic splicing substrates in *S. pombe*. Cells were treated with 1 μg/ml Tm for 1 h.

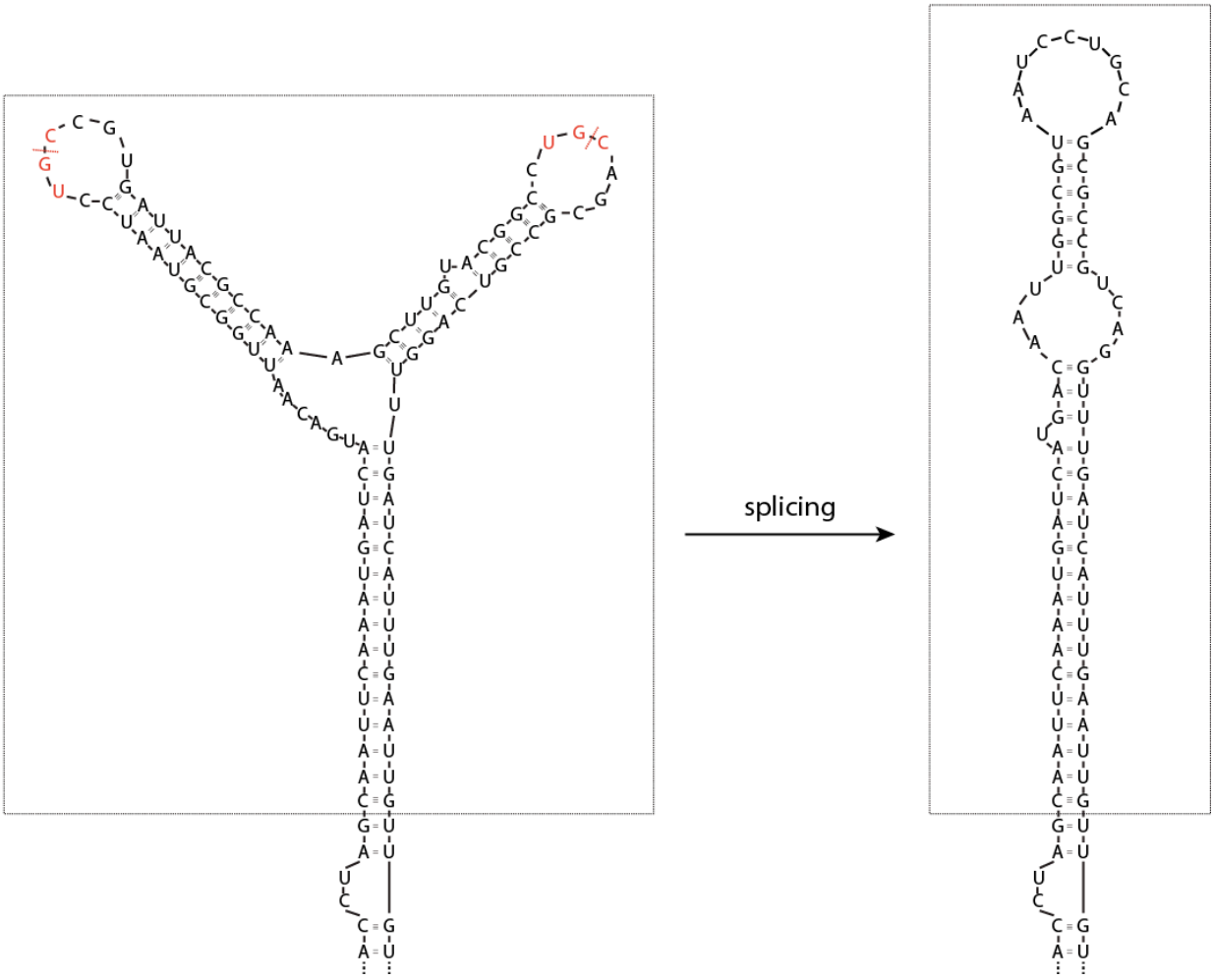


Figure 4-figure supplement 1 The splicing cassette in the engineered *S. pombe BIP1* mRNA splicing variant.

The part included in the dashed box is the inserted synthetic splicing cassette. The *red* dashed lines indicate the Ire1 cleavage sites. The *S. pombe* Ire1 UG|C motifs are labeled in red.

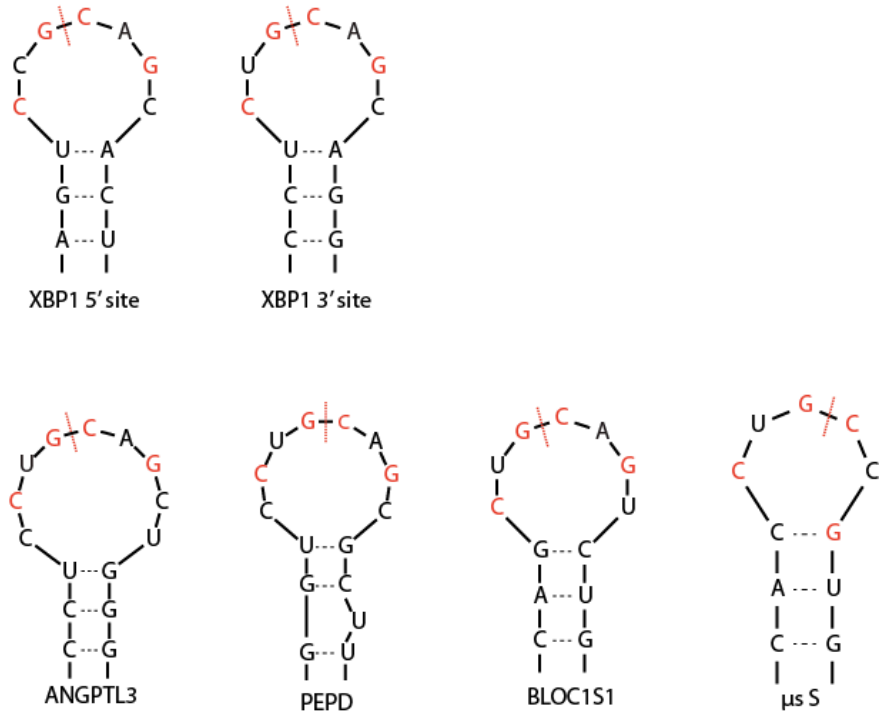


Figure 4-figure supplement 2 The Ire1 α cleavage sites on *XBP1* mRNA and a RIDD targets.

Red dashed lines mark the cleavage sites and the red letters indicate the previously identified sequence motif.

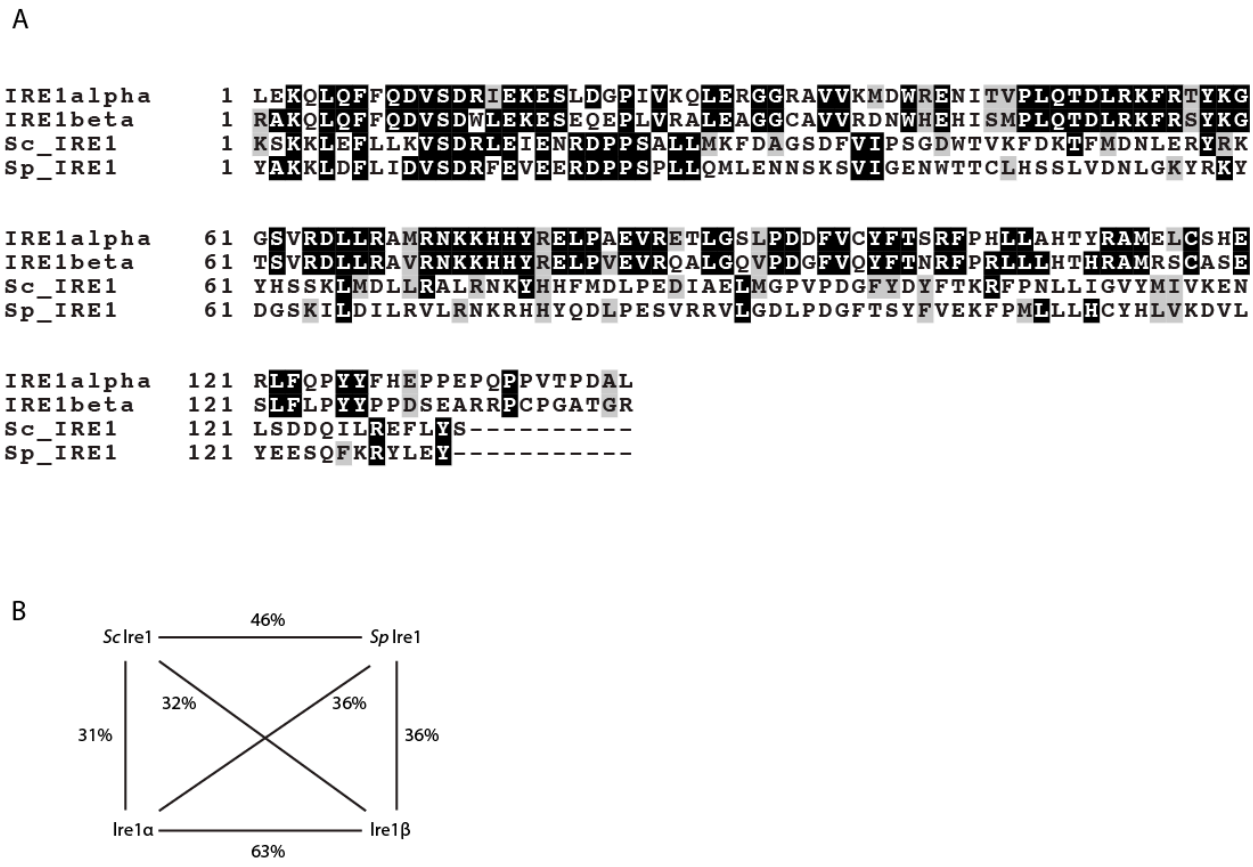


Figure 4-figure supplement 3 The sequence alignment of the kinase/RNase domains of Ire1 α , Ire1β, the *S. cerevisiae* Ire1 and the *S. pombe* Ire1.

(A) The sequence alignment and colored with BoxShade Server. (B) The sequence identities between the indicated pairs of Ire1 constructs.

References

1. Walter, P. and D. Ron, *The unfolded protein response: from stress pathway to homeostatic regulation.* Science, 2011. **334**(6059): p. 1081-6.

2. Ron, D. and P. Walter, *Signal integration in the endoplasmic reticulum unfolded protein response.* Nat Rev Mol Cell Biol, 2007. **8**(7): p. 519-29.

3. Gardner, B.M. and P. Walter, *Unfolded proteins are Ire1-activating ligands that directly induce the unfolded protein response.* Science, 2011. **333**(6051): p. 1891-4.

4. Karagoz, G.E., et al., *An unfolded protein-induced conformational switch activates mammalian IRE1.* Elife, 2017. **6**.

5. Credle, J.J., et al., *On the mechanism of sensing unfolded protein in the endoplasmic reticulum.* Proc Natl Acad Sci U S A, 2005. **102**(52): p. 18773-84.

6. Aragon, T., et al., *Messenger RNA targeting to endoplasmic reticulum stress signalling sites.* Nature, 2009. **457**(7230): p. 736-40.

7. Lee, K.P., et al., *Structure of the dual enzyme Ire1 reveals the basis for catalysis and regulation in nonconventional RNA splicing.* Cell, 2008. **132**(1): p. 89-100.

8. Korennykh, A.V., et al., *The unfolded protein response signals through high-order assembly of Ire1.* Nature, 2009. **457**(7230): p. 687-93.

9. Cox, J.S., C.E. Shamu, and P. Walter, *Transcriptional induction of genes encoding endoplasmic reticulum resident proteins requires a transmembrane protein kinase.* Cell, 1993. **73**(6): p. 1197-206.

10. Mori, K., et al., *A transmembrane protein with a cdc2+/CDC28-related kinase activity is required for signaling from the ER to the nucleus.* Cell, 1993. **74**(4): p. 743-56.

11. Yoshida, H., et al., *XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor.* Cell, 2001. **107**(7): p. 881-91.

12. Calfon, M., et al., *IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA.* Nature, 2002. **415**(6867): p. 92-6.

13. Sidrauski, C., J.S. Cox, and P. Walter, *tRNA ligase is required for regulated mRNA splicing in the unfolded protein response.* Cell, 1996. **87**(3): p. 405-13.

14. Hollien, J. and J.S. Weissman, *Decay of endoplasmic reticulum-localized mRNAs during the unfolded protein response.* Science, 2006. **313**(5783): p. 104-7.

15. Hollien, J., et al., *Regulated Ire1-dependent decay of messenger RNAs in mammalian cells.* J Cell Biol, 2009. **186**(3): p. 323-31.

16. Kimmig, P., et al., *The unfolded protein response in fission yeast modulates stability of select mRNAs to maintain protein homeostasis.* Elife, 2012. **1**: p. e00048.

17. Moore, K. and J. Hollien, *Ire1-mediated decay in mammalian cells relies on mRNA sequence, structure, and translational status.* Mol Biol Cell, 2015. **26**(16): p. 2873-84.

18. Niwa, M., et al., *Genome-scale approaches for discovering novel nonconventional splicing substrates of the Ire1 nuclease.* Genome Biol, 2005. **6**(1): p. R3.

19. Arnau, J., et al., *Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins.* Protein Expr Purif, 2006. **48**(1): p. 1-13.

20. Sambrook, J. and D.W. Russell, *Molecular cloning : a laboratory manual*. 3rd ed. 2001, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

21. Rosano, G.L. and E.A. Ceccarelli, *Recombinant protein expression in Escherichia coli: advances and challenges.* Front Microbiol, 2014. **5**: p. 172.

22. Korennykh, A.V., et al., *Cofactor-mediated conformational control in the bifunctional kinase/RNase Ire1.* BMC Biol, 2011. **9**: p. 48.

23. Sivashanmugam, A., et al., *Practical protocols for production of very high yields of recombinant proteins using Escherichia coli.* Protein Sci, 2009. **18**(5): p. 936-48.

24. Structural Genomics, C., et al., *Protein production and purification.* Nat Methods, 2008. **5**(2): p. 135-46.

25. Peschek, J., et al., *A conformational RNA zipper promotes intron ejection during non-conventional XBP1 mRNA splicing.* EMBO Rep, 2015. **16**(12): p. 1688-98.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Weihan Li
Author Signature

11/20/2018
Date