An exon three-way junction structure modulates splicing and degradation of the *SUS1* yeast pre-mRNA

Ali AbuQattam^{a,b,#}, Joan Serrano-Quílez^{c,#}, Susana Rodríguez-Navarro^{b,c*} and José Gallego^{a*}

^aFacultad de Medicina. Universidad Católica de Valencia, C/Quevedo 2, 46001 Valencia

^bGene Expression and RNA Metabolism Laboratory. Centro de Investigación Príncipe Felipe, C/ E. Primo Yúfera 3, 46012 Valencia

^cGene Expression and RNA Metabolism Laboratory. Instituto de Biomedicina de Valencia. Consejo Superior de Investigaciones Científicas (CSIC), C/ Jaime Roig 11, 46010 Valencia

[#] Equal contributors

*Correspondence to: srodriguez@ibv.csic.es and jose.gallego@ucv.es

Abstract

The *SUS1* gene of *Saccharomyces cerevisiae* is unusual as it contains two introns and undergoes alternative splicing, retaining one or both introns depending on growth conditions. The exon located between the two introns can be skipped during splicing and has been detected in circular form. This exon (E2) has also been found to influence the splicing of the flanking introns, an unusual situation in budding yeast where splicing mainly relies on intron recognition. Using SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension), NMR spectroscopy, gel electrophoresis and UV thermal denaturation experiments combined with computational predictions, we show that E2 of *SUS1* comprises a conserved double-helical stem topped by a three-way junction. One of the hairpins emerging from the junction exhibited significant thermal stability and was capped by a purine-rich loop structurally related to the substrate loop of the VS ribozyme. Cellular assays revealed that three mutants containing altered E2 structures had impaired *SUS1* expression, and that a compensatory mutation restoring the conserved stem recovered expression to wild-type levels. Semi-quantitative RT-PCR measurements paralleled these results, and revealed that mutations in E2 altered splicing and transcript degradation processes. Thus, exon structure plays an important role in *SUS1* RNA metabolism.

Keywords: exon; RNA structure; splicing; Sus1; three-way junction; yeast.

1. Introduction

In eukaryotic cells, introns are eliminated from pre-mRNA molecules by the spliceosome, a dynamic multicomponent machine [1-4]. The process of alternative splicing is widespread in higher eukaryotes, where it is considered to be a main source of protein diversity [5, 6]. The pervasiveness of this mechanism has been firmly supported by transcriptome sequencing data [7, 8]. Regardless of whether it leads to functional protein isoforms, alternative splicing contributes to modulate gene expression [9, 10]. For example, regulated splicing has been shown to control the levels of key transcripts encoding proteins involved in gene expression pathways, so that the cell can response to environmental changes via feed-back mechanisms [11]. In addition, the spliceosome generates circular RNA molecules (circRNA) through exon back-splicing. circRNAs are more commonly produced than initially thought, and may play important roles in gene regulation [12, 13].

Sus1 (ENY2 in mammals) is a small, evolutionary conserved 11-KDa protein involved in several processes of mRNA biogenesis [14]. In *Saccharomyces cerevisiae*, Sus1 interacts during transcription elongation with RNA polymerase II and the export factors Yra1 and Mex67 [15], and also accumulates at the nuclear pore, where it is part of the TREX-2 mRNA export complex [16, 17]. In addition, it participates in histone H2B deubiquitination as a component of the SAGA complex [16, 18].

Compared to higher eukaryotes, alternative splicing is rarer in *S. cerevisiae* cells. In this organism genes containing introns are scarce and typically contain only one intron with canonical 5' and 3' splice site (SS) and branch site (BS) sequences. In this context, the structure of the *SUS1* gene of *Saccharomyces cerevisiae* is remarkable: it contains two introns (Figure 1), and the first intron (I1) exhibits non-canonical 5'SS and BS sequences [16, 19]. The possible functional role of the two introns of *SUS1* has been explored. In agreement with the presence of non-canonical 5'SS and BS, I1 is retained

in more than 15% of the *SUS1* transcripts, and growth conditions affect the degree of I1 retention [20, 21]. On the other hand, the second intron (I2) is efficiently spliced and forms a weakly stable stem-loop structure that increases the accessibility of the BS and 3'SS nucleotides (nt). Changes in this I2 hairpin structure were found to alter the patterns of *Sus1* expression as well as *SUS1* splicing, giving rise to I1 retention and skipping of the second exon (E2) [22] (Figure 1). Altogether, these findings suggest that the functions of *Sus1* in mRNA biogenesis are modulated *via* splicing regulation.

During these studies it was detected that the sequence of *SUS1* E2 influenced the splicing of the flanking introns (I1 and I2) [20, 21]. This was also remarkable, as splicing in budding yeast relies on recognition by the spliceosome of conserved 5' and 3' SS and BS intronic sequences, and examples of modulation of this process by adjacent exon sequences are scarce [23]. In contrast, in higher eukaryotes the 5' and 3' SS are more degenerate, and splice site selection is carried out in conjunction with enhancer and silencer sequences located in introns and exons, which act as binding platforms for auxiliary proteins. The process of SS selection is also influenced by RNA structure in all eukaryotic organisms, which adds another layer of regulation. This is usually accomplished by modulation of the accessibility or spatial distribution of splicing signal sequences via base-pairing, but also through the presence of more complex folds like riboswitches and ribozymes [24-27]. In yeast, however, studies of the effect of RNA structure have focused on intronic sequences [28]. It was therefore relevant to evaluate whether a non-intronic sequence like *SUS1* exon 2 was structured and whether this structure had an impact on splicing, as this might compensate for the apparent absence of enhancer and silencer sequences in yeast pre-mRNA molecules. Our interest was also reinforced by the recent discovery that *SUS1* E2 of *Saccharomyces cerevisiae* is generated in circular form (E2c) [29] (Figure 1). In this

report we explore the role of E2 structure on *SUS1* metabolism by combining structural analyses with cellular assays evaluating the impact of mutant E2 sequences.

Figure 1. *S. cerevisiae* *SUS1* gene structure and RNA transcripts. (A) Schematic representation of the gene. (B) *S. cerevisiae* *SUS1* RNA species detected so far: pre-mRNA, partly-spliced pre-mRNA lacking I2 or I1, fully-spliced mRNA, fully-spliced mRNA lacking E2, and circular E2 (E2c).

2. Materials and Methods

2.1. Sequences

The sequences of the second exon of *SUS1* across different species of yeast were obtained from the Yeast Genome Database (SGD) and NCBI (<http://www.ncbi.nlm.nih.gov/>). E2 RNA and protein sequence identity calculations were carried out with the SIAS web tool (<http://imed.med.ucm.es/Tools/searches.html>). E2 RNA and protein sequence alignments were carried out with the T-COFFEE multiple sequence alignment web server [30]. Codon usage analyses were carried out with the CAIcal web server [31].

2.2. Secondary structure predictions

Secondary structure predictions were carried out with the Mfold [32], RNAfold [33], RNAstructure [34] and MC-Fold [35] web servers using default parameters. The structures were drawn using VARNA (<http://varna.lri.fr/>) [36]. We used the locARNA [37] web server to assess the conservation of the secondary structure formed by E2 across the seven yeast species containing two introns in the *SUS1* gene [20].

2.3. Preparation of RNA samples for experiments *in vitro*

The DNA template used for transcribing the full-length *S. cerevisiae* *SUS1* E2 RNA sequence used in SHAPE studies was obtained by PCR, using the primers indicated in Table S1 and a GPDp-*SUS1-CUP1* plasmid containing the wild-type *SUS1* sequence. The integrity of this DNA template was checked by sequencing. The 140-nt E2 RNA sequence was generated by T7-polymerase *in vitro* transcription and was flanked by 14- and 43-nt cassette sequences in its 5' and 3' sides, respectively [38]. The incorporation of these flanking sequences, designed to fold into stable hairpin structures, allowed evaluation of the reactivity of all 140 E2 nt positions. Likewise, the flanking 5' segment contained a binding site for a retrotranscription primer [38]. Folding calculations indicated that the presence of these cassettes did not interfere with the structure formed by the E2 RNA (data not shown). After transcription, the E2 RNA was purified on denaturing gels containing 8% 29:1 acrylamide:bisacrylamide and 8 M urea. The RNA product was extracted from the gel by passive diffusion, precipitated with ethanol and washed with 70% ethanol. The shorter E2s RNA sequence representing the capping purine-rich hairpin was similarly prepared by T7-polymerase *in vitro* transcription using a synthetic oligonucleotide DNA template. In addition to unlabeled E2s samples, we also generated ¹³C/¹⁵N-labelled transcripts using NTPs obtained from CortecNet. The E2s transcripts were purified on denaturing gels containing 20% 19:1 acrylamide:bisacrylamide and 8 M urea. After electroelution from the gel, the E2s RNA was ethanol-precipitated two times and desalted with Sephadex G-25 cartridges.

2.4. SHAPE experiments

SHAPE was performed with N-methylisatoic anhydride (NMIA) essentially as described [38]. Briefly, approximately 4 pmol of purified E2 RNA were denatured at 95 °C for 2 min, transferred to ice for 15 min, and renatured at 37 °C for 5 min in a 100 mM HEPES pH 8.0, 100 mM NaCl and 6 mM MgCl₂ folding buffer. 2'-acylation was

initiated by adding NMIA dissolved in DMSO to a final concentration of 6 mM, while only DMSO was added to an RNA control. The reactions proceeded for 15 min at 37 °C and were stopped by ethanol precipitation. After annealing the NMIA-treated and control RNAs with a DNA primer labeled with VIC fluorophore (Table S1), reverse transcription reactions were performed for 30 min at 52 °C with 100 U SuperScript III RT (Invitrogen) and 0.5 mM dNTPs. In parallel, 4 pmol of untreated RNA annealed to an analogous RT primer tagged with NED fluorophore (Table S1) were subjected to a sequencing reaction, which proceeded similarly except that ddTTP was added to the primer extension mix. The VIC- and NED-labeled DNA primers were acquired from Invitrogen and purified in denaturing polyacrylamide gels. In all cases, the resulting cDNAs were ethanol-precipitated, resuspended in deionized formamide and resolved by capillary electrophoresis in an ABI 3130 XL Genetic Analyzer DNA sequencer (Applied Biosystems). Electropherograms were analyzed using QuShape [39], which also normalized the reactivity values. Manual normalization of the reactivity data [40] gave rise to similar results. The SHAPE experiment was repeated with three different E2 RNA samples, and 3 to 6 replicas were performed for each sample. The average reactivity of each E2 nt was calculated from a total of 15 datasets.

The SHAPE-supported secondary structure of E2 was generated by incorporating the average SHAPE reactivities as constraints in the RNA folding programs RNAStructure version 5.6 [34] and MC-Fold [35]. We used $m = 1.8$ and $b = -0.6 \text{ Kcal mol}^{-1}$ in the SHAPE pseudo-free energy change term of RNAStructure [41]. Both algorithms generated similar E2 secondary structures with SHAPE restraints. Nt with normalized average reactivity values lower than 0.3 were considered to have low reactivity and were represented in black, nt with reactivities between 0.3 and 0.7 were intermediate

and were depicted in orange, and nt with reactivities greater than 0.7 were considered highly reactive and were colored red.

2.5. UV thermal denaturation

The thermal stability of the E2s RNA oligomer was monitored by measuring the UV absorbance at 260 nm as a function of temperature in a Varian Cary 100 UV/VIS spectrophotometer. The temperature was raised from 10 °C to 85 °C at a gradient of 0.5 °C min⁻¹ and subsequently decreased at the same rate to evaluate the reversibility of the process. The experiments were carried out using 0.4 ODU/mL RNA (1.2 µM). The thermal denaturing profiles were examined in the following solution conditions: 10 mM sodium phosphate (pH 6.0) and 0.1 mM EDTA with no other salts, or additionally containing either 100 mM NaCl or 2 mM MgCl₂. All melting experiments were repeated at least two times in each ionic condition. Before each experiment, RNA samples were heated at 95 °C for approximately 5 minutes and immediately placed on ice for 5 minutes.

2.6. Native gel electrophoresis

We used this technique to assess the number of conformations formed by the E2 and E2s RNA sequences in solution. The experiments were run at 4 °C for approximately 14 hours under constant voltage (90 V), and utilized either 8% 29:1 (for E2) or 20% 19:1 (E2s) acrylamide:bisacrylamide gels with 89 mM Tris-Borate (TB) as running buffer. They involved 8.9 µM E2s and 1.5 µM E2 samples, previously heated to 95 °C and either snap-cooled or cooled down slowly in the following ionic conditions: 10 mM sodium phosphate (pH 6.0) and 0.1mM EDTA with no added salts, or additionally containing either 100 mM NaCl or 2 mM MgCl₂. E2 was also analyzed in SHAPE folding buffer (100 mM HEPES pH 8.0, 100 mM NaCl and 6 mM MgCl₂). Gels were stained with methylene blue and destained with water.

2.7. NMR spectroscopy

The NMR samples contained 0.22-0.49 mM E2s RNA, previously microdialyzed in an aqueous solution containing 10 mM sodium phosphate (pH 6.0) and 0.1 mM EDTA. NMR spectra were acquired on 600 MHz (cryoprobe-equipped) and 500 MHz Bruker Avance III spectrometers, and analyzed using Topspin 3.5 (Bruker Biospin) and Sparky 3.113 (T. D. Goddard, D. G. Kneller, UCSF USA, 2004). The unlabeled E2s samples were studied with two-dimensional watergate-NOESY experiments (250 ms mixing time) recorded in 90% H₂O/10% D₂O at two temperatures (9 and 16 °C), as well as two-dimensional dqf-COSY, TOCSY (60 ms mixing time) and NOESY (120 and 250 ms mixing time) experiments acquired in D₂O at 25 and 36 °C. The recycle delay was 2 s for these experiments, which had between 600 and 800 points in the indirect dimension.

The ¹³C/¹⁵N-labelled E2s transcript was analyzed at 25 °C in D₂O with two-dimensional ¹H-¹³C HSQC and ¹H-¹H HCCH-TOCSY[42] and three-dimensional ¹³C-edited NOESY-HMQC (100 and 200 ms mixing times), HCCH-e-COSY and HCP experiments. Two-dimensional ¹H-¹⁵N HSQC experiments were also recorded in 90% H₂O/10% D₂O at 18 °C, as well as HNN COSY experiments allowing detection of hydrogen bonds between bases *via* two bond N-N couplings [43]. For ¹H-¹⁵N HSQC experiments we typically acquired 2048 and 256 complex points in the direct and indirect dimensions, respectively, and between 64 and 128 scans for each indirect experiment. For HNN-COSY experiments, the delay for evolution of the ²J_{NN} coupling was set to 15 ms, and we typically collected 2048 and 128 complex points in the t₂ and t₁ dimensions, respectively, with 300 scans for each t₁ increment. The recycle delays ranged between 1.0 and 1.3 s for all experiments with labelled E2s samples.

Assignments. E2s assignments involved exchangeable imino and amino signals as well as non-exchangeable aromatic and H1', H2' and H3' protons of all residues, and were based on standard analyses of NOESY, TOCSY and COSY spectra acquired in H₂O and D₂O at different temperatures, supported by the study of two-dimensional ¹H-¹H HCCH TOCSY, ¹H-¹³C HSQC, ¹H-¹⁵N HSQC and ¹H-¹⁵N HNN-COSY data obtained from ¹⁵N, ¹³C-labeled E2s samples dissolved in D₂O or H₂O.

2.8. E2s model building. A low-resolution three-dimensional model of stem-loop E2s was built to facilitate visual inspection of the most important structural features. The program RNAComposer [44] was used to generate an initial model, which was subsequently refined with restrained molecular dynamics at a temperature of 300 K followed by restrained energy-minimization. The list of NMR-based restraints included 62 hydrogen-bonding distance and planarity constraints supported by scalar couplings, NOEs and chemical shift patterns observed in HNN-COSY, HSQC and NOESY experiments; 164 distance constraints based on ¹H-¹H NOEs; and 32 δ backbone dihedral angle constraints supported by ¹H-¹H TOCSY crosspeak analyses. In addition, 59 canonical α , β , γ , ϵ and ζ backbone dihedral angle constraints were imposed to residues located in double-helical regions exhibiting A-conformational features as indicated by the NMR analyses. The calculations employed the ff10 force field of AMBER 8.0 [45] and a generalized Born model for solvent simulation. The quality of the final model was evaluated with MolProbity [46].

2.9. Generation of *SUSI* constructs

All *SUSI* gene constructs contained the last twenty nt of the 5'UTR. The E2 mutants (mut_s, mut1, mut2 and mut2r) were constructed by the fusion PCR method [47, 48], using the primers specified in Table S1. The constructs were cloned into a modified pRS425 vector containing the glyceraldehyde-3-phosphate dehydrogenase (GPD)

promoter and the *CUP1* gene as a reporter without ATG, followed by the first 200 nt of the *SUS1* gene 3'UTR (see the scheme in Figure S6). Taq DNA polymerase (Roche®) was used to amplify the WT, E2-muts, E2-mut1, E2mut2 and E2-mut2r constructs of *SUS1*.

2.10. Yeast strains and microbiological techniques

The yeast strains used in this report are listed in Table S2. Copper resistance assays were carried out by growing the transformed *cup1Δ* cells at 30 °C on synthetic selective medium (SC: glucose 2%, ammonium sulphate 0,5%, yeast nitrogen base 0,17% and supplements (Dropout)) lacking Leucine (Leu) to 0.4-0.5 OD₆₀₀. Subsequently, 10-fold serial dilutions of an equal number of cells were made and drops spotted onto SC-Leu plates containing different concentration of CuSO₄ [48]. Plates were photographed after 3-5 days of incubation at 30 °C. Yeast cell transformations were done by the LiAc/SS carrier DNA/PEG method [49].

2.11. RNA extraction, Reverse transcription PCRs and semi q-RT-PCRs

Total RNA was harvested from *sus1Δ* cells transformed with the CUP1 plasmids bearing SUS1g, SUS1-E2-muts, SUS1-E2-mut1, SUS1-E2-mut2 or SUS1-E2-mut2r by the Hot/Acid-phenol method [50], and quantified using a Nanodrop spectrophotometer. RNA quality was checked by 1% agarose gels dyed with ethidium bromide (EtBr). The cells were grown in 100 ml of SC-Leu at 30 °C until 0.4 OD₆₀₀ and then divided into two equal aliquots, in order to incubate the cell cultures under two different conditions: the cells of the first aliquot were grown 2 h more at 30 °C in SC-Leu and the cells of the second aliquot were collected by centrifugation, resuspended in equal volume of pre-heated 42 °C SC-Leu media and incubated for 20 m at 42 °C. A 500 ng of DNase I-treated RNA was used to perform the reverse transcription PCR in each case. Reverse transcription was carried out using standard procedures, with random hexamers and M-

MLV reverse transcriptase (Invitrogen®). For semi-quantitative RT-PCR a specific pair of primers located upstream exon1 and exon3 was used to amplify the transcripts of *SUS1* (the number of cycles was adjusted depending on the abundance of the transcripts) and *SCR1* (20 cycles). *SCR1* levels are commonly used as loading control. The amplified products were run in a REALSAFE® (Real laboratory) stained 3% agarose gel and visualized with a BioRad® UV CCD Camera. In all cases, negative controls that included all reagents except cDNA were included. The mRNA concentrations were normalized relative to *SCR1*, and the accumulation of mRNA is represented relative to wild-type. The mRNA bands were quantified with the ImageJ program (<http://rsbweb.nih.gov/ij/>)

3. Results

3.1. The RNA of the second exon of *SUS1* forms a three-way junction structure

The RNA-folding algorithms Mfold [32], RNAfold [33], RNAstructure [34] and MC-Fold [35] concurrently indicated that E2 of *S. cerevisiae SUS1* formed a 9- to 13-base pair stem (termed P1) topped by a three-way junction (Figure 2A). This structure was predicted to form regardless of the presence or absence of introns in the *SUS1* RNA sequence (Figure S1). Calculations using locARNA [37] and the E2 sequences of the seven yeast species containing two introns in the *SUS1* gene [20] indicated that the P1 double-helical stem of E2 was well conserved. The secondary structure of the junction region, in contrast, was more divergent across species (Figure S2). The unusual junction predicted for this exon and its partial conservation, together with the effect of E2 on the splicing of the flanking introns observed in *S. cerevisiae* [20, 21], prompted us to experimentally analyze the structure formed by this exon *in vitro*.

We first assessed the secondary structure adopted by the 140-nt E2 (full RNA sequence) of *S. cerevisiae* using selective 2'-hydroxyl acylation analyzed by primer extension

(SHAPE). This technique generates reactivity values for all RNA nt that depend on local backbone flexibility and can be used as restraints in an RNA folding algorithm [38, 40, 41, 51]. Using the average E2 reactivities (Figure 2B), a SHAPE-driven secondary structure of the E2 RNA was obtained (Figure 2C). With some rearrangements, this structure was consistent with those predicted by the unrestrained calculations (Figure 2A). The presence of the P1 stem conserved across different species was clearly confirmed by the experiments (Figure 2C). As predicted, this double-helical stem was connected to a three-way junction from which two hairpins (P2 and P3) emerged. The P3 hairpin (hereafter identified as the capping hairpin) was closed by a purine-rich loop comprising 12 nt (Figure 2C), and was predicted to have a longer stem by most folding algorithms (Figure 2A). The SHAPE-supported structure contained 21 single-stranded nt at the junction, most of them located in the J31 strand (Figure 2C). According to this nt distribution, the E2 junction would belong to topological family C of three-way intersections [52, 53]. In this type of junctions, helix P3 (the capping hairpin) usually bends towards P1 (conserved stem), stems P1 and P2 are typically coaxially stacked, and the J31 strand is frequently structured (Figure 2D). Electrophoretic experiments in native conditions indicated that E2 formed several conformations in solution (Figure S3A), likely reflecting the presence of different three-dimensional folds. In addition to the three-way junction, the SHAPE reactivities also supported the presence of two additional smaller hairpins preceding stem P1 at the 5' side of the exon (Figure 2C).

Figure 2. Analysis of the secondary structure adopted by the E2 RNA of *S. cerevisiae* *SUS1*.

(A) Theoretical minimum free-energy structure predicted for E2 by the RNAfold web server

[33] with default parameters. The RNAstructure [34] and MC-Fold [32] programs predicted approximately similar structures. **(B)** Normalized SHAPE reactivity of E2 RNA as a function of nt position, averaged from three independent assays. **(C)** Experimentally supported secondary structure of E2 RNA, generated by incorporating average SHAPE reactivities as constraints in the RNAstructure [34] folding algorithm. **(D)** Three-dimensional organization likely adopted by junction E2, which belongs to topological type C of three-way intersections. In these intersections, J31 contains most junction nt, P1 and P2 are coaxially stacked, and P3 bends towards P1. In **(B)** and **(C)**, nt with normalized average reactivity values <0.3 , $0.3-0.7$ and >0.7 are depicted in black, orange and red, respectively. E2 nt numbering was used in all panels.

3.2. The three-way junction formed by E2 is capped by an unusually structured hairpin

All folding algorithms together with the SHAPE reactivities supported the presence of the same P3 capping hairpin closed by a purine-rich loop (Figure 2). The sequence of this apical loop, G₇₉AACAUAAAUGA₉₀, had an unusually high percentage of purines (75%) and particularly adenine bases (58%), which have been shown to be frequently involved in tertiary interactions [54]. As part of our efforts to characterize the structure of E2, we next examined by NMR spectroscopy the conformation of the P3 hairpin using as a model a 32-nt oligomer (henceforth named E2s) encompassing E2 nt 71-98 (Figure 3A). The E2s hairpin included the A₇₃G₇₄:A₉₅C₉₆ internal loop and two Watson-Crick pairs at the base of the P3 stem that were predicted by most folding algorithms (Figure 2A), as well as two terminal G:C pairs inserted to increase transcription yields (Figure 3A). Electrophoretic experiments confirmed that E2s formed a single monomeric conformation in native conditions (Figure S3B), and the NMR data indicated that it adopted the stem-loop structure depicted in Figure 3.

In contrast to the opened loop depictions typically generated by RNA folding programs (Figure 2), the NMR analysis of the E2s sequence revealed that both the apical and the internal loops of the capping P3 hairpin were unusually structured (Figures 3A and 3B). Focusing first on the apical region, the spectral data indicated that nt G79, A80, G89 and A90 formed two tandem G:A pairs in the sheared (*trans* Hoogsteen-sugar edge) conformation. These G₇₉:A₉₀ and A₈₀:G₈₉ sheared pairs were stacked between U₇₈:A₉₁, the last canonical base pair of the stem, and a Watson-Crick A:U pair formed by nt A81 and U88 in the upper region of the loop (Figures 3A and 3B). This was revealed by the observation of sequential U78-G79-A80-A81 and U88-G89-A90-A91 sugar-aromatic and aromatic-aromatic NOE connectivities (Figure S4), and by sequential and cross-strand H2-H1' NOEs involving the H2 protons of A80, A81, A90 and A91 (Figure S4, crosspeaks i, j, k, l, s, t, u and v). The presence of the tandem G:A pairs sandwiched between Watson-Crick U₇₈:A₉₁ and A₈₁:U₈₈ pairs was likewise supported by the detection at 25 °C of a C3'-*endo* conformation for the U78, A80, A81, A90 and A91 sugars (data not shown); by the unusual chemical shift of the G79 and G89 HN1 imino protons (~10 ppm; Figure 3C), which are not engaged in hydrogen bonding interactions with the opposite adenines in the sheared pairs[55]; by the observation of two broad U HN3 iminos with chemical shifts typical of Watson-Crick U:A pairs (assigned to U78 and U88; Figure 3C); and finally by the strong upfield shift of A91 H1' (Figure S4), typical of a purine residue following a sheared G:A pair [56]. In this last respect, tandem sheared G:A pairs flanked by pyrimidine residues on their 5' ends, as in the E2s hairpin (Figure 3A), have been reported to give rise to inter-strand stacking of purine residues [56], and this is observed in the E2s model built from the NMR data (Figure 3B).

In the upper region of the loop, C82 and A87 were stacked on the A81:U88 pair and were probably base-paired. This was indicated by sequential A81-C82-A83 and A86-A87-U88 sugar-aromatic and aromatic-aromatic NOE connectivities (Figure S4), as well as by a strong sequential A87 H2-U88 H1' NOE (Figure S4, crosspeak r). A 1.5 ppm upfield shift of A87 C2 at pH 6.0 suggested partial protonation of A87 N1 [56, 57], further supporting C82:A87 pairing. The remaining A83, U84, A85 and A86 of the E2s apical region closed the terminal loop. They all exhibited strong TOCSY signals typical of C2'-*endo* sugars and had weaker or non-detectable sequential NOE interactions (Figure S4), particularly U84, whose base was likely exposed in the solvent (Figure 3B).

The A₇₃G₇₄:A₉₅C₉₆ internal loop in the middle of the E2s stem also adopted an unusual structure. Instead of the expected A₇₃:C₉₆ and G₇₄:A₉₅ pairs (Figure 2A), G74 established a Watson-Crick pair with C96 flanked by unpaired A73 and A95 bases (Figures 3A and 3B). The presence of the G₇₄:C₉₆ pair was unambiguously demonstrated by the detection of a hydrogen bond between G74 HN1 and C96 N3 with HNN-COSY experiments (Figure 3C). The unpaired A73 and A95 nt were stacked between G₇₄:C₉₆ and the neighboring canonical pairs of the stems (Figure 3B), as revealed by the detection of C72-A73-G74-A75 and U94-A95-C96-G97 sugar-aromatic and aromatic-aromatic NOEs (Figure S4), by sequential and cross-strand H2-H1' NOEs involving the H2 hydrogens of A73 and A95 (Figure S4, crosspeaks d, w and x), and by the C3'-*endo* character of the A73, A95 and C96 sugars at 25 °C (data not shown). Excluding loop and terminal nt, all of the residues forming the canonical stems of the hairpin exhibited hydrogen-bonding patterns, NOE connectivities and scalar signals typical of A-helices (Figures 3C and S4).

We also assessed the stability of the E2s stem-loop with UV thermal denaturation experiments. The E2s hairpin melted reversibly at 55 °C in an aqueous solution containing 100 mM NaCl, and at 66 °C in the presence of 2 mM MgCl₂ (Figures 3D and Figure S3C). Even considering the presence of the extra base pairs at the base of the stem, these results indicated that the P3 capping hairpin had significant thermal stability. This was in agreement with the detection by NMR of non-canonical pairs and stacked residues in the apical and internal loops (Figures 3A and 3B), which likely contributed to increase the melting temperature.

Figure 3. Structure and stability of the E2 capping hairpin P3, represented by the 32-nt sequence E2s. **(A)** NMR-supported secondary structure of the E2s stem-loop. Watson-Crick A:U and G:C pairs are represented with single and double lines, respectively, and sheared (*trans* Hoogsten-sugar edge) A:G pairs with open squares and triangles [55]. C82 and A87 are likely base-paired (discontinuous line), and the two terminal G:C pairs inserted to increase transcription yields are represented with lower case letters. All other residues follow full-length E2 nt numbering. **(B)** Three-dimensional model of the E2s hairpin supported by the NMR analyses, shown in two different orientations. Residues belonging to the 12-nt apical loop and to the 4-nt internal loop are depicted with orange and green carbon atoms, respectively. **(C)** Assignment of HSQC (top) and HNN-COSY (bottom) spectra of E2s acquired at 18 °C, based on NOESY and TOCSY analyses in H₂O and D₂O. Parentheses indicate tentative assignments. A broad imino resonance with a chemical shift similar to G79 was also detected and assigned to G89. **(D)** Representative UV-monitored thermal denaturation curve of E2s in 10 mM sodium phosphate (pH 6.0), 0.1 mM EDTA and 100 mM NaCl. The average melting temperature of E2s obtained under these ionic conditions is indicated in the graph.

3.3. Synonymous mutation of E2 impacts *SUS1* splicing and affects Sus1 expression

The long double helical stem topped by a complex junction structure formed by E2 of *S. cerevisiae SUS1* together with its partial conservation in seven yeast species suggested a possible functional relevance. To address this question we first studied a *SUS1* E2 mutant (E2-muts) in which the entire sequence of the exon was modified with synonymous mutations that kept invariant the amino acid sequence of the Sus1 protein and the length of the exon. Thus this mutant coded for the same protein but had a completely different E2 RNA sequence (Figure S5). The changes in RNA sequence increased GC content by 6% without significantly affecting codon usage (Table S3). As expected, the predicted structure of E2-muts changed relative to that formed by the wild-type sequence, becoming significantly more compact and thermodynamically stable (Figure 4A).

To evaluate the expression efficiency of this mutant *in vivo*, we used the *CUP1* reporter system [48]. For this assay, *cup1* Δ cells were transformed with one of the following reporter constructs: pACT1-CUP1, pSUS1g-CUP1 (containing a wild-type *SUS1* gene) and pSUS1-E2-muts-CUP1 (*SUS1* with synonymous E2 sequence). Splicing was monitored both by assessing copper tolerance and by semi-quantitative reverse transcription PCR (semi-qRT-PCR). Synonymous mutation of E2 led to a clear reduction in copper tolerance compared to wild-type E2 (Figure 4B), indicating that the sequence and/or structure of E2 were important for *SUS1* expression. Similarly, when the different *SUS1* splice forms generated from expressing SUS1g and SUS1-E2-muts in *sus1* Δ cells were measured by semi-qRT-PCR, we detected a clear reduction in the amount of fully spliced *SUS1* mRNA relative to wild-type, together with a concomitant enrichment of unspliced or partially spliced forms (Figure 4C, compare lanes 1 and 2).

Since it has been shown that unspliced or partially spliced *SUS1* species are degraded by quality control systems such as the nonsense-mediated decay (NMD) pathway or the nuclear exosome [20, 21, 58], we decided to evaluate the impact of E2 mutation in cells defective in Upf1 and Rrp6, which are components of the cytoplasmatic NMD and nuclear exosome systems, respectively [59, 60]. As expected from previous studies [20, 21, 58], absence of Upf1 or Rrp6 resulted in distinct phenotypes in terms of accumulation of processed and unprocessed *SUS1* transcripts. For the wild-type sequence, deletion of *UPF1* led to significant enrichment in intron-containing pre-mRNAs (Figure 4D, lane 3), while absence of *RRP6* showed no obvious phenotype (Figure 4E, lane 1). In contrast, *SUS1*-E2-muts exhibited a strong phenotype in both contexts. Elimination of *UPF1* led to a clear increase in unspliced or partially spliced mutant *SUS1* species (Figure 4D, lane 4), while deletion of *RRP6* resulted accumulation of E1-E3 transcripts generated by E2-muts skipping (Figure 4E, lane 2). In both cases, the amount of fully spliced mutant *SUS1* mRNA was significantly reduced (Figure 4D, lane 4 and 4E, lane 2).

Figure 4. Synonymous mutation of E2 impacts *SUS1* splicing and degradation and affects *Sus1* expression. (A) SHAPE-driven secondary structure of wild-type E2 RNA. The predicted secondary structure of E2-muts obtained with M-Fold [32] is shown on the right. For comparison, we obtained $\Delta G = -19.80 \text{ Kcal mol}^{-1}$ for the wild-type sequence with the same folding program. (B) Copper assay of *cup1* Δ cells transformed with plasmids containing *SUS1*g-CUP1 (E2-WT), *SUS1*-E2-muts-CUP1 (E2-muts) and ACT1-CUP1 (ACT1) as a control. Maximum copper tolerance (mM) is indicated. (C) Semi q-RT-PCR (25 cycles) to amplify *SUS1* transcripts from cells expressing *SUS1*g-CUP1 (WT; lane 1) and *SUS1*-E2-muts-CUP1 (E2-muts; lane 2), resolved in a 3% agarose gel. A PCR without cDNA (-) was included as control. The stick diagram on the right of the gel identifies the *SUS1* RNA species. (D) qPCR

showing mRNA accumulation from *upf1Δsus1Δ* cells expressing SUS1g-CUP1 (WT, lane 3) and SUS1-E2-muts-CUP1 (E2-muts, lane 4). A PCR without cDNA (-) was included as control. The stick diagram on the right of the gel identifies the *SUS1* RNA species. **(E)** As in **(D)** for *rrp6Δsus1Δ* cells expressing SUS1g-CUP1 (WT; lane 1) and SUS1-E2-muts-CUP1 (E2-muts; lane 2). A PCR without cDNA (-) was included as control. The stick diagram on the right of the gel identifies the *SUS1* RNA species.

3.4. The three-way junction structure formed by E2 affects Sus1 expression and modulates *SUS1* splicing and transcript degradation.

In the E2-muts mutant, both the sequence and the structure of E2 were extensively modified relative to wild-type. To more specifically evaluate if the P1 stem and the three-way junction detected by SHAPE experiments had a functional role, we prepared three additional mutants. In mutant E2-mut1, nt 15, 49-77 and 92-115 were eliminated (E2 nt numbering) (Figure 5A). In this way, the three-way junction was removed while keeping unchanged the purine-rich P3 loop and the peripheral regions of the exon (Figure 5A, right panel). This E2 mutant had 54 nt less than the wild-type RNA sequence and introduced an 18-amino acid deletion in the Sus1 protein sequence (Figure S5). On the other hand, mutant E2-mut2 was designed to disrupt the conserved stem (P1) at the base of the E2 junction. Thus, the G₄₃GUGG₄₇:C₁₁₇CACC₁₂₁ base pairs were replaced with C₄₃CUCC₄₇:C₁₁₇CUCC₁₂₁ (mutated nt are underlined) (Figure 5B, left panel). Mutant E2-mut2r was intended to restore this stem by additionally replacing U₄₅ with A, C₁₁₇C₁₁₈ with GG, and C₁₂₀C₁₂₁ with GG while keeping the E2-mut2 mutations, so that equivalent C₄₃CACC₄₇:G₁₁₇GUGG₁₂₁ base pairs would replace the wild-type pairings (Figure 5B, right panel). Mutants E2-mut2 and E2-mut2r introduced four and five Sus1 amino acid changes, respectively (Figure S5), but like E2-muts did

not affect the size of the exon. Folding calculations indicated that the E2-mut2 and E2-mut2r mutations had the expected changes in the structure of the exon (Figure 5). None of the mutations led to significant changes in codon usage (Table S3).

Figure 5. Predicted secondary structure of *SUS1* mutants affecting the three-way junction or the conserved P1 stem of E2. **(A)** SHAPE-supported secondary structure of wild-type E2. The shaded nt were removed to obtain the deletion mutant 1 (E2-mut1), whose predicted secondary structure, obtained with M-Fold [32], is shown on the right. In this mutant, the three-way junction was eliminated while keeping the P3 purine-rich loop unaltered (indicated with an arrow). **(B)** MFold-predicted secondary structure of the P1 disruption mutant (E2-mut2, left panel) and the P1 stem recovery mutant (E2-mut2r, right panel). The shaded and lower-case nt indicate mutations relative to the wild-type sequence. For comparison, $\Delta G = -19.80 \text{ Kcal mol}^{-1}$ was obtained for wild-type E2 with the same folding program.

Figure 6. The three-way junction and the conserved P1 stem of E2 modulate *Sus1* gene expression. **(A)** Copper assay of *cup1Δ* cells transformed with plasmids containing *SUS1g-CUP1* (E2-WT), *SUS1-E2-mut1-CUP1* (E2-mut1) and *ACT1-CUP1* (ACT1) as a control. **(B)** Copper assay of *cup1Δ* cells transformed with plasmids containing *SUS1g-CUP1*(E2-WT), *SUS1-E2-mut2-CUP1* (E2-mut2), *SUS1-E2-mut2r-CUP1* (E2-mut2r) and *ACT1-CUP1* (ACT1) as a control. Maximum copper tolerance (mM) is indicated in **(A)** and **(B)**.

We then evaluated the impact of each of these three mutants on *SUS1* expression as described above for E2-muts. The E2-mut1, E2-mut2 and E2-mut2r mutants were first cloned in the *CUP1* reporter system and transformed into *cup1Δ* cells. Cell growth was weakly impaired in E2-mut1 (copper tolerance up to 0.9 mM) relative to cells

transformed with wild-type (copper tolerance up to 1.1 mM) (Figure 6A), indicating that deletion of the E2 junction decreased *SUS1* expression. Likewise, the E2-mut2 mutant disrupting the P1 stem had a small deleterious effect (copper tolerance up to 0.9 mM) that was not observed in the E2-mut2r mutant, which restored copper resistance to wild-type levels (copper tolerance up to 1.1 mM) (Figure 6B).

To assess the production of RNA isoforms by these mutants, semi-q-RT-PCR was carried out to amplify *SUS1* transcripts in *sus1Δ* cells bearing wild-type and mutant constructs. In agreement with the copper assay observations, a slight decrement was detected in the levels of fully spliced E2-mut1 mRNA relative to wild-type, indicative of reduced splicing efficiency (Figure 7A). However, since the presence of unspliced or partially spliced species was not obviously detected in *sus1Δ* cells, we repeated the experiments in *upf1Δsus1Δ* and *rrp6Δsus1Δ* cells as described above for mutant E2-muts. Accumulation of unspliced or partially spliced *SUS1* E2-mut1 transcripts was detected in *upf1Δ* cells, together with a concomitant reduction of fully spliced mRNA (Figure 7B, compare lanes 3 and 2). Deletion of *RRP6*, in contrast, did not give rise to significant changes (Figure 7B, lanes 4 and 2).

With respect to the mutants affecting the P1 stem, a reduction of copper tolerance was observed for the destabilizing E2-mut2 mutant, which was canceled by the E2-mut2r mutant restoring the stem. However, semi-q-RT-PCR did not reveal significant differences of fully-spliced *SUS1* mRNA levels for these two mutants relative to wild-type (Figure 7C, left panel). We previously detected a reduction in *SUS1* expression when cells are incubated at higher temperatures (for instance 20 min at 42 °C), accompanied by accumulation of unspliced *SUS1* transcripts and decrement of fully spliced mRNA [20]. Moreover, we recently showed that *SUS1* I2 structural mutants affected the ratio of the different species in these conditions [22]. We thus tested the

effect of destabilizing and restoring the P1 stem at 42 °C. As shown in Figure 7C, (right panel), accumulation of unspliced forms of *SUS1* transcripts became evident in E2-mut2. Notably, restoring the structure of the P1 stem abolished this accumulation. Furthermore, analyses of E2-mut2 and E2-mut2r splicing in *upf1Δsus1Δ* and *rrp6Δsus1Δ* cells revealed that destabilization of the P1 stem exacerbated the accumulation of unspliced or partially spliced *SUS1* transcripts, particularly in the absence of *RRP6*; this effect was canceled by reconstitution of the P1 stem in the E2-mut2r mutant (Figure 7D). The different results obtained with E2-mut1 and E2-mut2 in *upf1Δ* and *rrp6Δ* cells indicated that, in addition to affecting the process of splicing, the nature of the structural changes implemented in the exon contributes to target unspliced or partially spliced transcripts to either the NMD or nuclear degradation systems (compare Figures 7B and 7D).

Figure 7. *SUS1* splicing and degradation is affected by the three-way junction and the conserved P1 stem of E2. **(A)** qPCR showing mRNA accumulation from cells expressing *SUS1g-CUP1* (WT) and *SUS1-E2-mut1-CUP1* (E2-mut1), normalized to the amount of *SCR1*. Error bars represent SE for at least three independent experiments. **(B)** Semi q-RT-PCR (25 cycles) to amplify *SUS1* transcripts resolved in a 3% agarose gel from cells expressing *SUS1-E2-mut1-CUP1* (E2-mut1) in *sus1Δ* (lane 2), *upf1Δsus1Δ* (lane 3) and *rrp6Δsus1Δ* (lane 4) cells. In **(A)** and **(B)**, a PCR without cDNA (-) was included as control, and the stick diagrams on the right of the gel identify the *SUS1* RNA transcripts. **(C)** As in **(A)**, from cells expressing *SUS1g-CUP1* (WT), *SUS1-E2-mut2-CUP1* (E2-mut2) and *SUS1-E2-mut2r-CUP1* (E2-mut2r) at 30 °C (left panel) or 42 °C (right panel). **(D)** As in **(C)**, from *upf1Δsus1Δ* and *rrp6Δsus1Δ* cells at 30 °C.

4. Discussion

The metabolism of mRNA molecules is regulated at different stages of the gene expression process, and RNA structure has been shown to contribute to these regulatory mechanisms [61]. However, little is known about how exon RNA folding influences these processes in yeast organisms. To shed light in this area, we have studied the second exon of the *SUS1* gene of *S. cerevisiae*, whose sequence has been previously shown to influence the splicing of its two flanking introns [20, 21].

SHAPE experiments indicated that this exon formed a 16-base pair discontinuous stem (P1; Figure 2), and bioinformatics analyses indicated that this stem was conserved across the seven yeast species containing two introns in the *SUS1* gene (Figure S2). A three-way junction of type C [52] connected stem P1 with hairpins P2 and P3 (Figure 2). In this topological family of RNA intersections, strand J31 contains most junction nt and is usually structured as a hairpin, helix P3 is bent towards P1, and stems P1 and P2 are coaxially stacked [52] (Figure 2D). There are numerous examples of three-dimensional structures adopted by type-C RNA junctions. They recurrently adopt a “parallel-Y” shape where helix P3 establishes tertiary contacts with helix P1 several base pairs away from the junction. These parallel-Y junctions have been observed to play essential functional and architectural roles in riboswitches, ribozymes and ribonucleoprotein systems [53]. If the E2 junction adopted this three-dimensional shape, stem P1 would occupy with the capping hairpin P3 the upper branches of the parallel “Y”. The low SHAPE reactivities exhibited by many junction nt support the possibility that the E2 junction strands form a complex structure. However, electrophoretic experiments also indicated that the E2 structure was flexible (Figure S3A) and likely adopted several three-dimensional folds in solution.

An NMR spectroscopy analysis of the purine-rich P3 stem-loop indicated that both the apical and the internal loops of this hairpin were unusually structured. The apical loop comprised two consecutive sheared G:A pairs flanked by Watson-Crick U₇₈:A₉₁ and A₈₁:U₈₈ pairs. The spectral data were also consistent with an additional C:A pair stacked on A₈₁:U₈₈, and indicated that the loop was closed by the remaining A83, U84, A85 and A86 residues. On the other hand, the non-opposing G74 and C96 bases of the internal loop predicted to be formed in the middle of the P3 stem established a Watson-Crick pair flanked by unpaired but stacked adenine bases (Figure 3). Not surprisingly, UV-monitored denaturation experiments indicated significant thermal stability for this hairpin (Figures 3C and S3C), in agreement with the detection by NMR of non-canonical base pairs and stacked residues in the apical and internal loops.

The base pairing organization of the P3 apical loop has some striking similarities with that of the internal loop present in domain I of the *Neurospora* Varkud Satellite (VS) ribozyme. This internal loop is called the substrate loop because it contains the scissile phosphate [62]. Like the P3 loop, the substrate loop of the VS ribozyme contains two tandem sheared G:A pairs flanked by pyrimidine residues on their 5' ends [56]. The tandem G:A pairs of the VS ribozyme are flanked by a Watson-Crick A:U pair and an A+:C pair, while the tandem G:A pairs of the P3 loop are flanked by two Watson-Crick A:U pairs, and an A:C opposition is stacked on the upper-most A:U pair (Figures 3A and 3B). The active site of the VS ribozyme is formed by the tertiary association of the substrate loop containing the scissile phosphate with another loop [62]. Based on this analogies, E2 might contain a scissile phosphate in loop P3 that could be cleaved as a result of the tertiary association of that loop with nucleotides of domain P1 or a different RNA molecule. Alternatively, the unusual structure of loop P3 might serve as a binding

platform for a regulatory trans-acting protein factor. Further structural and biochemical analyses will be necessary to test these hypotheses.

To assess whether the RNA structure of E2 played a role in *Sus1* expression, we studied four structural mutants. The most drastic effects were observed with mutant E2-muts, where all of the E2 sequence was modified with synonymous mutations that increased GC content by 6% and considerably compacted the structure of the exon (Figure 4) with little impact on codon usage (Table S3). The maximum copper tolerance for this construct was 0.6 mM CuSO₄, which is consistent with the reduced *SUS1* copper tolerance observed when substituting E2 with a 140-nt sequence derived from *TAF14* (0.5 mM CuSO₄) [20]. Cells expressing E2-muts showed a clear reduction in the abundance of fully spliced *SUS1* transcript that was accompanied by an increase of unspliced forms (Figure 4C). These effects were significantly amplified in the absence of components of the nuclear and cytoplasmatic RNA surveillance systems (*rrp6Δ* and *upf1Δ*, respectively). In both situations, the reduction of fully spliced *SUS1* levels became more evident, and accumulation of unspliced or partially spliced transcripts was observed in *upf1Δ* cells with a defective NMD pathway (Figure 4D). In contrast, E1-E3 species generated by E2-muts skipping became abundant in cells lacking the nuclear exosome Rrp6 component (Figure 4D). This result was consistent with the previous observation in *rrp6Δ* cells of higher levels of E1-E3 species resulting from E2 skipping in *SUS1* and other yeast genes containing two introns [58]. Thus, the E2-muts mutant altered *SUS1* splicing, leading to increased intron retention and exon skipping. This latter effect likely resulted from the masking of splice sites by secondary structure or from a reduction in the effective length of the exon, consistent with secondary structure predictions (Figure 4). According to our results, the intron-containing species were

exported to the cytoplasm and mostly targeted by the NMD pathway, whereas E1-E3 transcripts were preferentially degraded by the nuclear exosome (Figure 8).

Figure 8. Splicing and degradation routes of *SUS1* transcripts, and effect of mutations in exon 2. The E2 mutants modulated in different ways the splicing process and the degradation route of mis-spliced RNAs. E2-muts generated intron-containing species and E1-E3 transcripts, which were targeted by the cytoplasmatic NMD (Upf1) and nuclear exosome (Rrp6) systems, respectively. In contrast, E2-mut1 and E2-mut2 primarily led to intron-containing transcripts, which were degraded by the cytoplasmic (E2-mut1) or nuclear (E2-mut2) pathways.

In mutant E2-mut1 the three-way junction was removed without changing the P3 loop, while in mutant E2-mut2 the conserved P1 stem was disrupted (Figure 5). The study of these mutants helped us to gain insight into the specific role of these structural features in splicing and transcript stability. A small reduction of *SUS1* copper tolerance relative to wild-type was observed in both constructs (Figure 6). In cells with intact RNA quality control systems, E2-mut1 reduced the amount of fully spliced mRNA, and E2-mut2 increased the levels of unspliced or partially spliced forms at 42 °C (Figures 7A and 7C). The effect of both structural elements, however, became more evident in cells lacking components of the NMD and nuclear exosome pathways. E2-mut1 resulted in increased abundance of unspliced or partially spliced transcripts in *upf1Δ* cells (Figure 7B), suggesting that NMD was important for degrading these *SUS1* species lacking the three-way junction. On the other hand, disruption of the P1 stem contributed to targeting *SUS1* transcripts for nuclear RNA degradation, since in this case the increment of unspliced or partially spliced forms was most evident in the absence of Rrp6 (Figure 7D). Importantly, the E2-mut2r mutant restoring the P1 stem cancelled the effects

observed with E2-mut2 in both CUP1 and RT-PCR experiments (see Figures 6 and 7). Thus, E2-mut1 and E2-mut2 increased intron retention, but did not induce exon skipping as observed for E2-muts. Moreover, the E2-mut1 and E2-mut2 mutations targeted intron-containing transcripts to different degradation pathways.

A comparison of the results obtained with the different mutants indicate that the sequence and the structure of E2 play a role in modulating the splicing and degradation of *SUS1* transcripts, as summarized in Figure 8. On the other hand, splicing of *SUS1* has also been shown to generate E2 in circular form (E2c), which can be amplified by PCR under normal growth conditions ([29] and our unpublished observations). Whether E2c has any functional role in *Sus1* expression or function, or even more interesting, in any other cellular process, is still unknown. The work presented here will serve as a starting point to address new questions including the role of RNA structure in E2c generation and E2c function, as well as the possibility that E2c plays a regulatory role *in trans* in *Sus1* expression.

5. Conclusions

In this work we have shown that E2 adopts a three-way junction structure that includes a conserved double-helical stem and an unusually structured purine-rich loop. The functional data indicated that this structure exerted a role in *SUS1* transcript metabolism that included splicing and transcript degradation.

Acknowledgements

The authors wish to thank Álvaro Simba, Amparo López-Carrasco, and Ricardo Flores for help with the SHAPE experiments. This work has been supported by MINECO of Spain and FEDER funds (BFU2011-23418 and BFU2014-57636-P to S.R.-N, and BFU-2012-30770 and BFU2015-65103-R to J.G.), by Generalitat Valenciana of Spain

(PROMETEO/2013/061, PROMETEO/2016/091, ACOMP/2014/061 and ACOMP/2015/096 to S.R-N., ACOMP/2014/056 to J.G., Santiago Grisolia fellowship to A.AQ and FPU15/03862 fellowship to J.S.-Q), and by Universidad Católica de Valencia (2017-114-001 to J.G.).

References

- [1] M.C. Wahl, C.L. Will, R. Lührmann, The spliceosome: design principles of a dynamic RNP machine, *Cell*, 136 (2009) 701-718.
- [2] J.H. Cate, STRUCTURE. A Big Bang in spliceosome structural biology, *Science*, 351 (2016) 1390-1392.
- [3] R. Sperling, The nuts and bolts of the endogenous spliceosome, *Wiley Interdiscip Rev RNA*, 8 (2017).
- [4] P. Papasaïkas, J. Valcárcel, The Spliceosome: The Ultimate RNA Chaperone and Sculptor, *Trends Biochem Sci*, 41 (2016) 33-45.
- [5] T.W. Nilsen, B.R. Graveley, Expansion of the eukaryotic proteome by alternative splicing, *Nature*, 463 (2010) 457-463.
- [6] M.L. Tress, F. Abascal, A. Valencia, Alternative Splicing May Not Be the Key to Proteome Complexity, *Trends Biochem Sci*, 42 (2017) 98-110.
- [7] C. Ben-Dov, B. Hartmann, J. Lundgren, J. Valcárcel, Genome-wide analysis of alternative pre-mRNA splicing, *J Biol Chem*, 283 (2008) 1229-1233.
- [8] N. Stepankiw, M. Raghavan, E.A. Fogarty, A. Grimson, J.A. Pleiss, Widespread alternative and aberrant splicing revealed by lariat sequencing, *Nucleic Acids Res*, 43 (2015) 8488-8501.
- [9] J.J. Wong, A.Y. Au, W. Ritchie, J.E. Rasko, Intron retention in mRNA: No longer nonsense: Known and putative roles of intron retention in normal and disease biology, *Bioessays*, 38 (2016) 41-49.
- [10] U. Braunschweig, S. Gueroussov, A.M. Plocik, B.R. Graveley, B.J. Blencowe, Dynamic integration of splicing within gene regulatory pathways, *Cell*, 152 (2013) 1252-1269.
- [11] T.L. Johnson, J. Vilardeïl, Regulated pre-mRNA splicing: the ghostwriter of the eukaryotic genome, *Biochim Biophys Acta*, 1819 (2012) 538-545.
- [12] L.L. Chen, The biogenesis and emerging roles of circular RNAs, *Nat Rev Mol Cell Biol*, 17 (2016) 205-211.
- [13] J. Salzman, Circular RNA Expression: Its Potential Regulation and Function, *Trends Genet*, 32 (2016) 309-316.
- [14] A. Galán, S. Rodríguez-Navarro, Sus1/ENY2: a multitasking protein in eukaryotic gene expression, *Crit Rev Biochem Mol Biol*, 47 (2012) 556-568.
- [15] P. Pascual-García, C.K. Govind, E. Queralt, B. Cuenca-Bono, A. Llopis, S. Chavez, A.G. Hinnebusch, S. Rodríguez-Navarro, Sus1 is recruited to coding regions and functions during transcription elongation in association with SAGA and TREX2, *Genes Dev*, 22 (2008) 2811-2822.
- [16] S. Rodríguez-Navarro, T. Fischer, M.J. Luo, O. Antúnez, S. Brettschneider, J. Lechner, J.E. Pérez-Ortín, R. Reed, E. Hurt, Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery, *Cell*, 116 (2004) 75-86.

- [17] D. Jani, S. Lutz, N.J. Marshall, T. Fischer, A. Köhler, A.M. Ellisdon, E. Hurt, M. Stewart, Sus1, Cdc31, and the Sac3 CID region form a conserved interaction platform that promotes nuclear pore association and mRNA export, *Mol Cell*, 33 (2009) 727-737.
- [18] A. Köhler, P. Pascual-García, A. Llopis, M. Zapater, F. Posas, E. Hurt, S. Rodríguez-Navarro, The mRNA export factor Sus1 is involved in Spt/Ada/Gen5 acetyltransferase-mediated H2B deubiquitinylation through its interaction with Ubp8 and Sgf11, *Mol Biol Cell*, 17 (2006) 4228-4236.
- [19] P. Pascual-García, S. Rodríguez-Navarro, A tale of coupling, Sus1 function in transcription and mRNA export, *RNA Biol*, 6 (2009) 141-144.
- [20] B. Cuenca-Bono, V. García-Molinero, P. Pascual-García, H. Dopazo, A. Llopis, J. Vilardell, S. Rodríguez-Navarro, SUS1 introns are required for efficient mRNA nuclear export in yeast, *Nucleic Acids Res*, 39 (2011) 8599-8611.
- [21] M.A. Hossain, C.M. Rodriguez, T.L. Johnson, Key features of the two-intron *Saccharomyces cerevisiae* gene SUS1 contribute to its alternative splicing, *Nucleic Acids Res*, 39 (2011) 8612-8627.
- [22] A. AbuQattam, J. Gallego, S. Rodríguez-Navarro, An intronic RNA structure modulates expression of the mRNA biogenesis factor Sus1, *RNA*, 22 (2016) 75-86.
- [23] B.J. Fair, J.A. Pleiss, The power of fission: yeast as a tool for understanding complex splicing, *Curr Genet*, 63 (2017) 375-380.
- [24] C.J. McManus, B.R. Graveley, RNA structure and the mechanisms of alternative splicing, *Curr Opin Genet Dev*, 21 (2011) 373-379.
- [25] M.B. Warf, J.A. Berglund, Role of RNA structure in regulating pre-mRNA splicing, *Trends Biochem Sci*, 35 (2010) 169-178.
- [26] Y. Jin, Y. Yang, P. Zhang, New insights into RNA secondary structure in the alternative splicing of pre-mRNAs, *RNA Biol*, 8 (2011) 450-457.
- [27] I. García-Robles, J. Sánchez-Navarro, M. de la Peña, Intronic hammerhead ribozymes in mRNA biogenesis, *Biol Chem*, 393 (2012) 1317-1326.
- [28] J. Pérez-Valle, J. Vilardell, Intronic features that determine the selection of the 3' splice site, *Wiley Interdiscip Rev RNA*, 3 (2012) 707-717.
- [29] P.L. Wang, Y. Bao, M.C. Yee, S.P. Barrett, G.J. Hogan, M.N. Olsen, J.R. Dinneny, P.O. Brown, J. Salzman, Circular RNA is expressed across the eukaryotic tree of life, *PLoS One*, 9 (2014) e90859.
- [30] C. Notredame, D.G. Higgins, J. Heringa, T-Coffee: A novel method for fast and accurate multiple sequence alignment, *J Mol Biol*, 302 (2000) 205-217.
- [31] P. Puigbò, I.G. Bravo, S. Garcia-Vallve, CAIcal: a combined set of tools to assess codon usage adaptation, *Biol Direct*, 3 (2008) 38.
- [32] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Res*, 31 (2003) 3406-3415.
- [33] R. Lorenz, S.H. Bernhart, C. Höner Zu Siederdissen, H. Tafer, C. Flamm, P.F. Stadler, I.L. Hofacker, ViennaRNA Package 2.0, *Algorithms Mol Biol*, 6 (2011) 26.
- [34] J.S. Reuter, D.H. Mathews, RNAstructure: software for RNA secondary structure prediction and analysis, *BMC Bioinformatics*, 11 (2010) 129.
- [35] M. Parisien, F. Major, The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data, *Nature*, 452 (2008) 51-55.
- [36] K. Darty, A. Denise, Y. Ponty, VARNA: Interactive drawing and editing of the RNA secondary structure, *Bioinformatics*, 25 (2009) 1974-1975.
- [37] S. Will, T. Joshi, I.L. Hofacker, P.F. Stadler, R. Backofen, LocARNA-P: accurate boundary prediction and improved detection of structural RNAs, *RNA*, 18 (2012) 900-914.

- [38] K.A. Wilkinson, E.J. Merino, K.M. Weeks, Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution, *Nat Protoc*, 1 (2006) 1610-1616.
- [39] F. Karabiber, J.L. McGinnis, O.V. Favorov, K.M. Weeks, QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis, *RNA*, 19 (2013) 63-73.
- [40] K.E. Deigan, T.W. Li, D.H. Mathews, K.M. Weeks, Accurate SHAPE-directed RNA structure determination, *Proc Natl Acad Sci U S A*, 106 (2009) 97-102.
- [41] C.E. Hajdin, S. Bellaousov, W. Huggins, C.W. Leonard, D.H. Mathews, K.M. Weeks, Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots, *Proc Natl Acad Sci U S A*, 110 (2013) 5498-5503.
- [42] J.P. Marino, J.H. Prestegard, D.M. Crothers, Correlation of Adenine H2/H8 Resonances in Uniformly ^{13}C Labeled RNAs by 2D HCCH-TOCSY: A New Tool for ^1H Assignment., *Journal of the American Chemical Society*, 116 (1994) 2205-2206.
- [43] A.J. Dingley, S. Grzesiek, Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $(^2\text{J}(\text{NN}))$ couplings, *Journal of the American Chemical Society*, 120 (1998) 8293-8297.
- [44] M. Popenda, M. Szachniuk, M. Antczak, K.J. Purzycka, P. Lukasiak, N. Bartol, J. Blazewicz, R.W. Adamiak, Automated 3D structure composition for large RNAs, *Nucleic Acids Res*, 40 (2012) e112.
- [45] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, *J Comput Chem*, 26 (2005) 1668-1688.
- [46] I.W. Davis, A. Leaver-Fay, V.B. Chen, J.N. Block, G.J. Kapral, X. Wang, L.W. Murray, W.B. Arendall, J. Snoeyink, J.S. Richardson, D.C. Richardson, MolProbity: all-atom contacts and structure validation for proteins and nucleic acids, *Nucleic Acids Res*, 35 (2007) W375-383.
- [47] J. Yon, M. Fried, Precise gene fusion by PCR, *Nucleic Acids Res*, 17 (1989) 4895.
- [48] C.F. Lesser, C. Guthrie, Mutational analysis of pre-mRNA splicing in *Saccharomyces cerevisiae* using a sensitive new reporter gene, CUP1, *Genetics*, 133 (1993) 851-863.
- [49] R.D. Gietz, R.H. Schiestl, Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method, *Nat Protoc*, 2 (2007) 35-37.
- [50] M.E. Schmitt, T.A. Brown, B.L. Trumpower, A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*, *Nucleic Acids Res*, 18 (1990) 3091-3092.
- [51] K.M. Weeks, D.M. Mauger, Exploring RNA structural codes with SHAPE chemistry, *Acc Chem Res*, 44 (2011) 1280-1291.
- [52] A. Lescoute, E. Westhof, Topology of three-way junctions in folded RNAs, *RNA*, 12 (2006) 83-93.
- [53] M. de la Pena, D. Dufour, J. Gallego, Three-way RNA junctions with remote tertiary contacts: a recurrent and highly versatile fold, *Rna*, 15 (2009) 1949-1964.
- [54] S.E. Butcher, A.M. Pyle, The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks, *Acc Chem Res*, 44 (2011) 1302-1311.
- [55] N.B. Leontis, J. Stombaugh, E. Westhof, The non-Watson-Crick base pairs and their associated isostericity matrices, *Nucleic Acids Res*, 30 (2002) 3497-3531.
- [56] J. Flinders, T. Dieckmann, A pH controlled conformational switch in the cleavage site of the VS ribozyme substrate RNA, *J Mol Biol*, 308 (2001) 665-679.

- [57] P.J. Michiels, C.H. Schouten, C.W. Hilbers, H.A. Heus, Structure of the ribozyme substrate hairpin of *Neurospora* VS RNA: a close look at the cleavage site, *RNA*, 6 (2000) 1821-1832.
- [58] D.E. Eggecioglu, T.R. Kawashima, G.F. Chanfreau, Quality control of MATa1 splicing and exon skipping by nuclear RNA degradation, *Nucleic Acids Res*, 40 (2012) 1787-1796.
- [59] S. Sayani, M. Janis, C.Y. Lee, I. Toesca, G.F. Chanfreau, Widespread impact of nonsense-mediated mRNA decay on the yeast intronome, *Mol Cell*, 31 (2008) 360-370.
- [60] C. Bousquet-Antonelli, C. Presutti, D. Tollervy, Identification of a regulated pathway for nuclear pre-mRNA turnover, *Cell*, 102 (2000) 765-775.
- [61] Y. Wan, M. Kertesz, R.C. Spitale, E. Segal, H.Y. Chang, Understanding the transcriptome through RNA structure, *Nat Rev Genet*, 12 (2011) 641-655.
- [62] P. Dagenais, N. Girard, E. Bonneau, P. Legault, Insights into RNA structure and dynamics from recent NMR and X-ray studies of the *Neurospora* Varkud satellite ribozyme, *Wiley Interdiscip Rev RNA*, (2017).

An exon three-way junction structure modulates splicing and degradation of the SUS1 yeast pre-mRNA

Ali AbuQattam, Joan Serrano-Quílez, Susana Rodríguez-Navarro and José Gallego

SUPPLEMENTARY MATERIAL

Supplementary Figure legends

Figure S1. Predicted secondary structure of different RNA species of *S. cerevisiae* *SUS1*: unspliced pre-mRNA, partly-spliced pre-mRNA lacking I1, partly-spliced pre-mRNA lacking I2, and fully-spliced mRNA. E2 (green-shaded nt) and the previously studied I2 (red-shaded nt) [1] form the same structures regardless of sequence context. The E1, E3 and I1 components of the gene are represented in cyan, blue and gold, respectively. The predictions were carried out with Mfold [2] with default parameters, and the minimum free-energy structure is shown in all cases.

Figure S2. Conservation of E2 in yeast species containing two introns in the *SUS1* gene. (A) Sequence identity of *SUS1* E2 across seven yeast species containing two introns in this gene [3]: *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, *S.*

bayanus, *S. Castellii*, and *K. thermotolerans*. The sequence identities were obtained with the SIAS web tool (<http://imed.med.ucm.es/Tools/searches.html>) and are expressed relative to *S. cerevisiae*. Note that the sequence of E2 is more conserved than that of E1 or E3. **(B)** *SUS1* E2 sequence alignment across the same species, carried out with the T-COFFEE multiple sequence alignment web server [4]. **(C)** Secondary structure conservation of E2 RNA across the same species, carried out with locARNA [5]. The top panel shows the consensus secondary structure in dot-bracket notation above the sequence alignment, whereas the bottom panel represents the corresponding consensus secondary structure folding. In both cases, the color of the conserved base pairs indicates the number of different compatible pairings (up to six: C:G, G:C, A:U, U:A, G:U or U:G) contained in a given base pair position, and thus represents sequence conservation of the base pair. Color saturation decreases with the number of incompatible pairings, indicating structural conservation of the base pair [5].

Figure S3. *In vitro* analysis of the structure and stability of the E2 and E2s sequences of *S. cerevisiae* *SUS1* RNA. **(A)** Native gel comparing the electrophoretic mobility of E2 dissolved in SHAPE buffer: 100 mM HEPES (pH 8.0), 100 mM NaCl and 6 mM MgCl₂. The last lane contained 98-nt and 196-nt RNA controls. E2 exhibited similar electrophoretic mobility when other ionic conditions were tested (not shown). **(B)** Native gel comparing the electrophoretic mobility of E2s samples in different ionic conditions: (1) 10 mM sodium phosphate (pH 6.0) and 0.1 mM EDTA, (2) 10 mM sodium phosphate (pH 6.0), 0.1 mM EDTA and 100 mM NaCl, (3) 10 mM sodium phosphate (pH 6.0), 0.1 mM EDTA and 2 mM MgCl₂. The last two lanes contained 26-nt and 55-nt RNA controls. In (A) and (B), the E2 and E2s samples were previously heated to 95 °C and cooled-down slowly (annealing) or rapidly (snap-cooling). **(C)** UV-

monitored thermal denaturation curves of E2s. The curves show representative E2s melting curves in aqueous buffers containing 10 mM sodium phosphate (pH 6.0) and 0.1 mM EDTA with no added salts, or additionally containing 2 mM MgCl₂. The average melting temperatures measured for E2s under each ionic condition are indicated in the graphs.

Figure S4. NMR spectroscopy analysis of the purine-rich P3 RNA hairpin formed by the second exon of the *SUS1* gene of *S. cerevisiae*, represented by the E2s sequence. The image shows the assignment of the H2/H6/H8-H1'/H5 region of an E2s NOESY spectrum (D₂O, 250 ms mixing time, 25 °C). Intraresidue H1'-H6/H8 cross-peaks are labelled with residue name and number, intraresidue H5-H6 crosspeaks are labelled with residue number, and sequential NOE connectivities are indicated with horizontal arrows. The A73 H1'-G74 H8 and A87 H1'-U88 H6 sequential crosspeaks were weak but visible at a lower contour level. Crosspeaks (a) to (c) are assigned as follows: a, A77 H8-U78 H5; b, A95 H8-C96 H5 (overlapped with C82 H1'-A83 H8); c, A98 H8-c31 H5. The g2 H8-U71 H5, A87 H8-U88 H5 and U92 H6-C93 H5 NOEs were weak but visible at a lower contour level. Crosspeaks (d) to (z) involve adenine H2 protons and were assigned as indicated: d, A73 H2-G74 H1' and A73 H2-G97 H1'; e A75 H2-G76 H1'; f, A77 H2-H1'; g, A77 H2-C93 H1'; h, A77 H2-U78 H1'; i, A80 H2-A90 H1'; j, A80 H2-A81 H1'; k A81 H2-G89 H1'; l, A81 H2-C82 H1'; m, A85 H2-A86 H1'; n, A85 H2-H1'; o, A86 H2-A85 H1'; p, A86 H2-A87 H1'; q, A87 H2-A83 H1'; r, A87 H2-U88 H1'; s, A90 H2-A80 H1'; t, A90 H2-A91 H1'; u, A91 H2-G79 H1'; v, A91 H2-U92 H1'; w, A95 H2-A75 H1'; x, A95 H2-C96 H1'; y, A98 H2-C72 H1'; z, A98 H2-c31 H1'. Crosspeak aa corresponds to g1 H8-H2'. Dynamics of A₇₃G₇₄:A₉₅C₉₆ internal loop nucleotides likely caused broadening of some G74 and C96 resonances as

well as the doubling of several resonances of the lower g₁-C₇₂:G₉₇-c₃₂ stem (extra resonances are labelled with asterisks), and G76 H1' was also broadened. All of the assignments were supported by analyses of NOESY and TOCSY spectra at 36 °C as well as by studies of two-dimensional ¹H-¹H HCCH TOCSY and ¹H-¹³C HSQC data obtained from a ¹⁵N/¹³C-labeled E2s sample in the same temperature and solution conditions.

Figure S5. Impact of *SUSI* E2 mutations on RNA and protein sequence. **(A)** Alignment (top) and sequence identity (bottom) of wild-type and mutant E2 RNA sequences. **(B)** Alignment (left) and sequence identity (right) of the amino acid sequences coded by wild-type and mutant E2 RNAs. None of the E2 mutants introduced stop codons in the RNA sequence. In **(A)** and **(B)**, sequence identities were obtained with the SIAS web tool (<http://imed.med.ucm.es/Tools/searches.html>) and are expressed relative to wild-type, whereas sequence alignments were carried out with the T-COFFEE web server [4].

Figure S7. Schematic representation of the pGSC1t-E2c-ms2 plasmid used for expressing E2c. The representation was generated by the ApE plasmid editor.

Supplementary Tables

Table S1. List of primers used in this study. Mutated nucleotides are indicated with lower case bold letters, SHAPE cassette sequences [6] are labeled with lower case letters.

E2-mut1 Forward	GAAGGTTGGGTGGATGAACATAAATGAATCCACCGTAGAACCCA AAGCATTAGGTATGT
E2-mut1 Reverse	TCATCCACCCAACCTTCTTGAAGTAGTCTGGCTTTTGTTCGTTTGA AATTCTAATA
E2-mut2 Forward	CCAAATCAGAGATGAACATAAATGAATCTACGAACTTTACCCAAA TTTTATCC t CCGTAGAACCC
E2-mut2 Reverse	GTTTCATCTCTGATTTGGTCAAATCTTTGACTTTAT ggAgg CAACCTT CTTGAAGTAG
E2-mut2r Forward	GATTTGACCAAATCAGAGATGAACATAAATGAATCTACGAACTTT ACCCAAATTTTAT ggtgg GTAGAACCCAAAGC
E2-mut2r Reverse	CATCTCTGATTTGGTCAAATCTTTGACTTTAT ggtgg CAACCTTCTTG AAG
Extreme Forward 5' primer for all constructs	AAAAAA aagctt <u>CAATTCTGGCCTTCACTCCA</u> ATGACTATGGATACT GC
Extreme Reverse 3' primer for all	AAAAAA Actgcag TTGTGTATCTACAATCTC
<i>SCR1</i> Forward	AACCGTCTTTCCTCCGTCGTAA
<i>SCR1</i> Reverse	AGAACTACCTTGCCGCACCA
SHAPE cassette Forward	TAATACGACTCACTATAggccttcgggccaaAATTTCAAACGAACTAAAA GCC
SHAPE cassette Reverse	gaaccggaccgaagcccgatttgatccggcgaaccggaatgaCTAATGCTTTGGGTTCTA CGG
VIC-labelled SHAPE	GAACCGGACCGAAGCCCG
NED-labelled SHAPE	GAACCGGACCGAAGCCCG
<i>SUS1</i> Forward	ATGACTATGGATACTGCGC
<i>SUS1</i> Reverse	TCATTGTGTATCTACAATCTCTTC

Table S2. List of yeast strains used in this report.

Yeast strain	Genotype	Reference
<i>sus1</i> Δ	<i>MAT a leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 sus1::KANMX4</i>	this study
<i>cup1</i> Δ	<i>MAT a leu2-Δ0 ura3-Δ0 trp1-Δ0 lys2-Δ0 ade2-Δ0 cup1::URA3</i>	[7]
<i>sus1</i> Δ <i>upf1</i> Δ	<i>MAT a leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 sus1::KANMX4 upf1::URA3</i>	this study
<i>sus1</i> Δ <i>cup1</i> Δ	<i>MAT a leu2-Δ0 ura3-Δ0 trp1-Δ0 lys2-Δ0 ade2-Δ0 cup1::URA3 sus1::KANMX4</i>	[3]
<i>sus1</i> Δ <i>rrp6</i> Δ	<i>MAT a leu2-Δ0 his3-Δ1 met15-Δ0 ura3-Δ0 sus1::KANMX4 rrp6::KANMX4</i>	this study

Table S3. Codon usage analysis for wild-type and mutant *SUS1* mRNA sequences, carried out with the CAIcal web server [8]. The codon adaptation indexes (CAI) of all sequences were within the range observed for *Saccharomyces cerevisiae* genes with higher expression levels [9].

mRNA	Codon adaptation index (CAI) [9]	GC content (%)
SUS1 (wild-type)	0.797	34.4
SUS1 (E2-muts)	0.702	40.5
SUS1 (E2-mut1)	0.773	35.4
SUS1 (E2-mut2)	0.793	34.8
SUS1 (E2-mut2r)	0.801	34.4

Supplementary References

- [1] A. AbuQattam, J. Gallego, S. Rodríguez-Navarro, An intronic RNA structure modulates expression of the mRNA biogenesis factor Sus1, *RNA*, 22 (2016) 75-86.
- [2] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Res*, 31 (2003) 3406-3415.
- [3] B. Cuenca-Bono, V. García-Molinero, P. Pascual-García, H. Dopazo, A. Llopis, J. Vilardell, S. Rodríguez-Navarro, SUS1 introns are required for efficient mRNA nuclear export in yeast, *Nucleic Acids Res*, 39 (2011) 8599-8611.
- [4] C. Notredame, D.G. Higgins, J. Heringa, T-Coffee: A novel method for fast and accurate multiple sequence alignment, *J Mol Biol*, 302 (2000) 205-217.
- [5] S. Will, T. Joshi, I.L. Hofacker, P.F. Stadler, R. Backofen, LocARNA-P: accurate boundary prediction and improved detection of structural RNAs, *RNA*, 18 (2012) 900-914.
- [6] K.A. Wilkinson, E.J. Merino, K.M. Weeks, Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution, *Nat Protoc*, 1 (2006) 1610-1616.
- [7] C.F. Lesser, C. Guthrie, Mutational analysis of pre-mRNA splicing in *Saccharomyces cerevisiae* using a sensitive new reporter gene, CUP1, *Genetics*, 133 (1993) 851-863.
- [8] P. Puigbò, I.G. Bravo, S. Garcia-Vallve, CAIcal: a combined set of tools to assess codon usage adaptation, *Biol Direct*, 3 (2008) 38.
- [9] P.M. Sharp, W.H. Li, The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications, *Nucleic Acids Res*, 15 (1987) 1281-1295.

A

B

Fig. 1

A

B

C

D

Fig. 2

Fig. 3

Fig. 4

A

B

Fig. 5

A

B

Fig. 6

Fig. 7

Fig. 8

Supplementary Material

Fig. S1

Exon1Scerevisae	100%	Exon2S.cerevisae	100%	Exon3Scerevisae	100%
Exon1Sparadoxus	94.36%	Exon2S.paradoxus	93.57%	Exon3Sparadoxus	96.1%
Exon1Sbayanus	90.14%	Exon2S.bayanus	85.71%	Exon3Sbayanus	85%
Exon1Skudriavzevii	88.73%	Exon2S.kudriavzevii	89.28%	Exon3Skudriavzevii	87.34%
Exon1Kthermotolerans	36.25%	Exon2K.thermotolerans	62.85%	Exon3Kthermotolerans	58.75%
Exon1Smikatae	33.84%	Exon2S.mikatae	90%	Exon3Smikatae	85.36%
Exon1Scastellii	27.94%	Exon2S.castellii	62.14%	Exon3Scastellii	62.5%

Consensus

S.cerevisiae
 S.paradoxus
 S.mikatae
 S.kudriavzevii
 S.bayanus
 S.castellii
 K.thermotolerans

10 20 30 40 50

S.cerevisiae
 S.paradoxus
 S.mikatae
 S.kudriavzevii
 S.bayanus
 S.castellii
 K.thermotolerans

70 80 90 100 110

S.cerevisiae
 S.paradoxus
 S.mikatae
 S.kudriavzevii
 S.bayanus
 S.castellii
 K.thermotolerans

130 140

A

Slow cool, shape buffer

Snap cool, shape buffer

B

C

Fig. S3

Fig. S4

A

E2-WT	100%
E2-muts	62.14%
E2-mut1	48.83%
E2-mut2	96.42%
E2-mutr	92.85%

B

Fig. S5

Fig. S6