PRINCIPE FELIPE

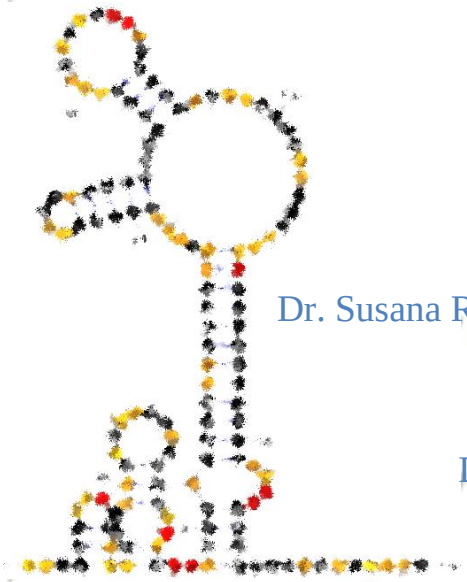
CENTRO DE INVESTIGACION

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Universidad Católica de Valencia San Vicente Mártir

Biotechnology Department

IDENTIFICATION OF RNA STRUCTURES MODULATING THE EXPRESSION OF THE mRNA BIOGENESIS FACTOR *SUS1*



Doctoral Thesis

Ali NA ABUQATTAM

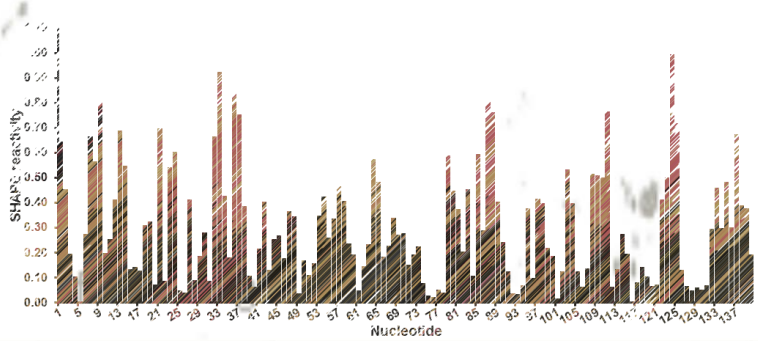
Directors

Dr. Susana Rodríguez Navarro and Dr. José Gallego Sala

Tutor

Dr. José Ramón Murguía

Valencia, JUNE 2017



La Dra. Susana Rodríguez Navarro, investigador jefe del centro de Investigación Príncipe Felipe (CIPF) y Dr. José Gallego Sala investigador jefe de la Universidad Católica de Valencia (UCV) (como directores de la tesis) y el Dr. José Ramón Murguía Subdirector del Área de Biotecnología del Dpto. de Biotecnología de la escuela técnica superior de ingenieros agrónomos de la Universidad politécnica de Valencia (UPV).

CERTIFICAN que el Alumno graduado en Biología Ali NA AbuQattam ha realizado bajo su dirección en CIPF y UCV y bajo su tutela en UPV respectivamente, el trabajo que lleva por el título "IDENTIFICATION OF RNA STRUCTURES MODULATING THE EXPRESSION OF THE mRNA BIOGENESIS FACTOR SUS1" y autorizan su presentación para optar el grado de doctor en Biotecnología.

Y para que así conste, expiden y firman el presente certificado en Valencia, Junio de 2017.

Dra. Susana Rodríguez Navarro Dr. José Gallego Sala Dr. José Ramón Murguía

ACKNOWLEDGEMENTS

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., and Santiago Grisolía fellowship to A.AQ), and by Universidad Católica de Valencia.

I would like to express my special appreciation and thanks to my advisors Dr. Susana Rodríguez Navarro (my second mother) and Dr.José Gallego Sala (who introduce me to investigation world); you have been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist and as a good building block of the society. Your advice on both researches as well as on my career has been priceless.

A special thanks to my family. Words cannot express how grateful I am to my mother, father and for all of the sacrifices that you've made on my behalf. Your prayer for me was what sustained me thus far. I would also like to thank all of my friends and colleagues whose supported me (Salamah Alwahsh, Mohamad Ajdid, Ali Abad, Ismael Otri and Moahmad Baker) , and encourage me to strive towards my goal. A special thanks to my colleagues in CIPF, UCV and UPV; David Dufour, Luis González, Angel Camacho, Silvia Prado, Cristian Medina, Jesus Castello, Isaias Sanmartin, Varinia García, Encarnacíon García, Manuel Martine, María Eugenia Gas, Carmen Nuño, Juan Serrano, Eloisa Barber and Amparo López. At the end I would like express appreciation to my beloved wife Israa E´mar who spent sleepless nights with me and was always my support in the moments when there was no one to answer my queries.

ABSTRACT

Sus1 is a conserved protein involved in chromatin remodeling and mRNA biogenesis. The *SUS1* gene of *Saccharomyces cerevisiae* is unusual, as it contains two introns and is alternatively spliced, retaining one or both introns in response to changes in environmental conditions. *SUS1* splicing may allow the cell to control Sus1 expression, but the mechanisms that regulate this process remain unknown. In this thesis project, we have investigated whether the structure adopted by *SUS1* RNA sequences contributes to regulate the splicing of this gene. Using *in silico* analyses together with NMR spectroscopy, gel electrophoresis and UV thermal denaturation experiments, we first show that the downstream intron (I2) of *SUS1* forms a weakly-stable, 37-nucleotide stem-loop structure containing the branch site near its apical loop and the 3' splice site after the stem terminus. A cellular assay revealed that two of four mutants containing altered I2 structures had significantly impaired *SUS1* expression. Semi-quantitative RT-PCR experiments indicated that all mutants accumulated unspliced *SUS1* pre-mRNA and/or induced distorted levels of fully spliced mRNA relative to wild-type. Concomitantly, Sus1 cellular functions in histone H2B deubiquitination and mRNA export were affected in I2 hairpin mutants that inhibited splicing. The second part of the thesis project focuses on the exon located between the two introns of the *SUS1* gene. This middle exon (E2) can be skipped during splicing, is generated in circular form, and has been found to influence the splicing of the flanking introns, an unusual situation in budding yeast where splicing mainly relies on intron recognition. Using NMR spectroscopy, gel electrophoresis, UV thermal denaturation and ribose 2'-OH modification 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 closed by an unusually structured purine-rich loop. This loop contained two consecutive sheared G:A base pairs and was 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 experiments indicated that all mutants were capable of altering the quantities of unspliced and/or fully-spliced *SUS1* RNA transcripts relative to wild-type. Overall, the results gathered in this thesis project indicate that RNA structures formed by the middle exon and the second intron of the *S. cerevisiae SUS1* gene are relevant for splicing and also influence other processes of *SUS1* mRNA biogenesis.

RESUMEN

Sus1 es una proteína conservada implicada en remodelación de cromatina y biogénesis de moléculas de ARNm. El gen *SUS1* de *Saccharomyces cerevisiae* es peculiar, ya que contiene dos intrones y sufre un proceso de ayuste (corte y empalme) alternativo, reteniendo uno o ambos intrones en respuesta a cambios en las condiciones ambientales. El ayuste del ARNpre-m de *SUS1* puede permitir a la célula controlar la expresión de la proteína Sus1, pero los mecanismos que regulan este proceso son poco conocidos. En este proyecto de tesis hemos investigado si la estructura adoptada por secuencias de ARN de *SUS1* contribuye a regular el proceso de ayuste de este gen. Utilizando análisis *in silico* junto con espectroscopia de RMN, electroforesis en gel y experimentos de desnaturalización térmica monitorizados por UV, primero demostramos que el ARN del segundo intrón (I2) del gen *SUS1* forma una horquilla débilmente estable de 37 nucleótidos. Esta horquilla contiene nucleótidos del sitio de ramificación (*branch site*) en su bucle apical y nucleótidos del sitio 3' de empalme adyacentes al extremo inferior del tallo. A través de ensayos funcionales descubrimos que dos de cuatro mutantes que alteran la estructura de la horquilla I2 exhibían peor expresión de *SUS1*. Experimentos de RT-PCR semi-cuantitativos indicaron que todos los mutantes acumularon ARNpre-m *SUS1* no ayustado y/o indujeron cambios en los niveles de ARNm maduro con respecto a la secuencia silvestre. Además, las funciones celulares de Sus1 relativas a desubicuitinación de histona H2B y transporte de ARNm se vieron afectadas en los mutantes de la horquilla I2 que inhibían el proceso de ayuste. La segunda parte de la memoria de tesis se centra en el análisis del exón central (E2) situado entre los dos intrones del gen *SUS1*. Este exón puede eliminarse durante el proceso de ayuste, se genera en forma circular, e influye en el procesamiento de los intrones adyacentes, una situación inusual para las regiones exónicas de *S. cerevisiae*, donde el ayuste se basa principalmente en el reconocimiento de intrones. Utilizando experimentos de espectroscopía de RMN, electroforesis en gel, desnaturalización térmica y modificación química combinados con predicciones computacionales, demostramos que el ARN del exón E2 de *SUS1* forma un tallo conservado de doble hélice coronado por una intersección de tres hélices. Una de las horquillas que emergen de esta intersección presentó una estabilidad térmica significativa, así como un bucle apical rico en purinas inusualmente estructurado. Este bucle contiene dos pares de bases G:A consecutivos y está estructuralmente relacionado con el bucle de substrato de la ribozima VS. Ensayos celulares revelaron que tres mutantes con estructuras modificadas de E2 exhibían peor expresión de *SUS1*, y que una mutación compensatoria que restauraba el tallo conservado recuperaba la expresión a los niveles de la secuencia silvestre. Experimentos de RT-PCR semicuantitativos indicaron que todos los mutantes de E2 eran capaces de alterar las cantidades de transcritos ayustados y no ayustados de *SUS1* con respecto a la secuencia silvestre. En general, los resultados obtenidos en este proyecto de tesis indican que las

estructuras de ARN formadas por el exón central y el segundo intrón del gen *SUS1* de *S. cerevisiae* son relevantes para el ayuste y otros procesos implicados en la biogénesis del ARNm del gen *SUS1*.

RESUM

Sus1 és una proteïna conservada implicada a la remodelació de la cromatina i la biogènesi de l'ARNm. El gen *SUS1* de *Saccharomyces cerevisiae* és inusual, ja que conté dos introns i s'empalma de manera alternativa, retenint un o ambdós introns en resposta a canvis en les condicions ambientals. L'empalmament de *SUS1* pot permetre a la cèl·lula controlar l'expressió de Sus1, però els mecanismes que regulen aquest procés són segueixen sent desconeguts. En aquest projecte de tesi investiguem si l'estructura adoptada per seqüències d'ARN de *SUS1* contribueix a regular l'empalmament d'aquest gen. Emprant anàlisi *in silico* juntament amb espectrometria de RMN, electroforesi en gel i experiments de desnaturalització tèrmica d'UV, es mostra primer que l'intró aigües a baix (I2) de *SUS1* forma una estructura de forqueta de 37 nucleòtids feblement estable que conté el lloc de la branca a prop del seu bucle apical; i el lloc d'empalmamnet 3´ després de l'extrem de la forqueta. Un assaig cel·lular va revelar que dos de quatre mutants que contenien estructures alterades de l'I2 havien modificat significativament l'expressió de *SUS1*. Els experiments semi-quantitatius de RT-PCR van indicar que tots els mutants acumulaven el pre-ARNm madur respecte al tipus salvatge. Concomitantment, les funcions cel·lulars de Sus1 a la desubiqüitinació de la histona H2B i l'exportació d'ARNm es van veure afectats als mutants de la forqueta d'I2 que inhibeixen l'empalmament. La segona part del projecte de tesi se centra a l'exó situat entre els dos introns del gen *SUS1*. Aquest exó (E2) es pot ometre durant l'empalmament, es genera amb forma circular, i s'ha trobat que influeix a l'empalmamet dels introns que flanquegen, una situació inusual al llevat on l'empalmament està basat principalment al reconeixement d'introns. Emprant espectroscòpia de RMN, electroforesi en gel, desnaturalització tèrmica d'UV i experiments de modificació de ribosa 2´-OH combinats amb prediccions computacionals, mostrem que E2 de *SUS1* comprén un tall conservat de doble hèlix corornat per una unió de tres vies. Una de les forquetes que emergeixen de la unió, va mostrar una estabilitat tèrmica significativa i va ser tapada per un bucle ric en purina inusualment estructurat. Aquest bucle contenia dos pars de bases G:A tallats consecutivament i estava estructuralment relacionat amb el bucle de substrat del ribozim VS. Els assajos cel·lulars van revelar que tres mutants que contenien estructures alterades de E2 havien alterat l'expressió de *SUS1* i que una mutació compensatòria que restaurava el tall conservat recuperava l'expressió a nivells del tipus salvatge. Els assajos cel·lulars van revelat que tres mutants que contenien estructures alterades d'E2 havien alterat l'expressió de *SUS1* i que una mutació compensatòria que restaurava el tall conservat recuperava l'expressió a nivell d'un tipus salvatge. Els experiments semi-quantitatius de RT-PCR van indicar que tots els mutants eren capaços d'alterar les quantitats de transcrits d'ARN de *SUS1* no empalmats i/o empalmats en relació amb el tipus salvatge. En general, els resultats obtinguts en aquesta investigació indiquen que les estructures d'ARN formades per l'exó mitjà i el segon intró de

SUS1 de *S. cerevisiae* són rellevants per l'empalmament i també influeixen a altres processos de biogènesi de l'ARN de *SUS1*.

Index

IDENTIFICATION OF RNA STRUCTURES MODULATING THE EXPRESSION OF THE mRNA BIOGENESIS FACTOR *SUS1*

OVERVIEW

In this Ph.D. project, I have studied the structure and functionality of different RNA sequences of the *Saccharomyces cerevisiae SUS1* gene. The thesis dissertation is organized as follows;

1) I first present a short Introduction, covering some of the essential topics related to the thesis's work. This section is general and concise, since the following chapters contain a specific Introduction that covers more precisely the results contained in them.

2) The Materials and Methods, Results and Discussion sections of the thesis are next presented in two chapters, each of them representing one manuscript in which I am first author. **Chapter I** was published in 2016 in the journal *RNA* [1] , and **Chapter II** will be submitted in the near future. Each chapter is structured as a complete manuscript and contains the following sections: Abstract, Introduction, Results, Discussion, Materials and Methods, Figures, Bibliography and Supplementary Material.

3) Last, I present a list of conclusions summarizing my work.

INTRODUCTION

RNA DISCOVERY AND EVOLUTION

RNA has gained enormous interest in the past two decades as an extremely versatile molecule, which is able to carry genetic information, function as an enzyme, nucleate and monitor the assembly of complex molecular machines and regulate a variety of cellular processes [2]. The nucleic acid was discovered in 1868 by the Swiss biologist Johannes Friedrich Miescher; he called it ´nuclein´ because of its presence in the nucleus. Until the beginning of 1900s, the RNA and DNA were named depending on the materials from which they were isolated; RNA was known as yeast nucleic acid and DNA was thymus nucleic acid [3]. Ludwig Kossel isolated and described the five nitrogenous bases of nucleic acids: adenine, cytosine, guanine, thymine, and uracil. These bases were later named nucleosides when they were covalently bonded to a ribose or deoxyribose sugar, and nucleotides when the nucleoside was linked to one, two or three phosphate groups [4]. RNA is distinguished from DNA by its sensitivity to alkaline conditions, because it has an additional hydroxyl group on the ribose.

After Crick's description of his "Central Dogma of Molecular Biology", which asserted that DNA led to the formation of RNA, a sophisticated genetic analysis of mutations in the lac operon of E. coli was instrumental in defining the nature of both messenger RNA and the genetic code [5]. The first direct experimental evidence for the existence of mRNA was provided by a hemoglobin synthesizing system [6].

RNA resembles DNA in many ways. It is a nucleic acid, it stores and transmits information. However, it differs from DNA in the key aspect that it has been shown to catalyze chemical reactions in the cell. The evolutionary history of RNA is rich and varied, and this dynamic molecule is turning up everywhere in the cell. RNA looks like half of a DNA double helix, and carries only a couple of ostensibly small chemical modifications, such as a missing methyl group in one of its nucleobases (uracil) and an extra hydroxyl group in its sugar rings [7].

There are many arguments supporting the idea that RNA molecules were the origin of life. These arguments are summarized as follows: i) protein biosynthesis can take place in the absence of DNA but not of RNA; ii) the existence of RNA viruses and RNA molecules with catalytic properties [8]. In the last two decades, with the identification of noncoding RNAs

(ncRNA), numerous new types of RNA have been discovered and described in different living organisms. A summary of the main types of RNA is presented in Table1.

Table 1: Different types of RNA and their functions. For a full list of newly discovered types see [9].

Name	Definition	Function
miRNA	MicroRNA	RNA that, in complex with AGO protein, uses seed sequences near its 5´end to base pair with a target mRNA to induce deadenylation and decay or translational regulation
rRNA	ribosomal RNA	Component of the small or large ribosomal subunit; the largest is a ribozyme
snRNA	small nuclear RNA	RNAs localized in the eukaryotic cell nucleus. Form ribonucleoproteins that are components of the spliceosome.
piRNA	PIWI-associated RNA	RNA that directs the modification of chromatin to repress transcription; best characterized in the male germline
snoRNA	small nucleolar RNA; in vertebrates, most noRNAs are processed intron fragments	essential for pre-rRNA processing
RNase P RNA	RNA component of ribonuclease P (processing)	catalytic subunit of the enzyme that removes 5´ -leaders from pre-tRNAs
CRISPR RNA	clusters of regularly interspersed short palindromic repeat RNA	targets Cas nuclease to cleave a specific DNA, such as a phage DNA, in bacteria or archaea
eRNA	transcriptional enhancer element RNA	binds Mediator to enhance transcription
gRNA	guide RNA	base pairs with an RNA or DNA target, orienting bound proteins to carry out a site-specific cleavage, ligation or modification reaction
Group I intron	a structural class of self-splicing RNAs	ribozyme that binds guanosine and uses it as nucleophile to catalyze RNA splicing
Group II intron	intron a structural class of self-splicing RNAs	ribozyme that catalyzes splicing via formation of a lariat intron
lncRNA	long noncoding RNA	autonomously transcribed RNA that does not encode a protein; often capped and polyadenylated; can be nuclear, cytoplasmic or both
Telomerase RNA (TR, TER or TERC)	telomerase RNA	provides template for telomeric DNA synthesis and a scaffold for protein assembly
U snRNA	U-rich small nuclear RNA	RNA subclass of snRNAs; many are building blocks of the major or minor spliceosome, serving to recognize intron boundaries and perhaps to catalyze intron removal.
mRNA	messenger RNA	contains a coding region that directs synthesis of a protein product; flanked by untranslated sequences

MESSENGER (mRNA) BIOGENESIS AND MODIFICATION

Among the various RNA species, mRNA is the most divergent in sequence, length, and structure [10]. mRNA biogenesis is a coordinated process that comprises different linked steps like transcription, RNA processing (5´capping, epigenetic modification, splicing and 3´polyadenalytion), export to the cytoplasm, translation and degradation. Regulation of gene expression is vital to establishing the metabolic activities in every cell in all organisms. A cross-talk between nuclear transcription process and cytoplasmic mRNA degradation has been demonstrated, suggesting that gene expression is a circular system [11].

Before the mRNA molecule can be translated into a protein, it must be processed and spliced into a mature transcript and then exported from sites of transcription to the cytoplasm [12]. In the following sections, a short overview of mRNA biogenesis will be presented as an important introduction to this dissertation.

A- Transcription.

As a general definition, transcription relies on the complementary pairing of bases. The process is catalyzed by the enzyme RNA polymerase II (RNAPII), which attaches and moves along the DNA adding ribonucleotides to the growing RNA chain [13].

mRNA transcription starts with the recruitment of the RNAPII complex to the gene promoter and is followed by a set of successive steps, such as initiation, elongation, splicing, termination, mRNA export, and translation or degradation. The early events of this process occur co-transcriptionally while the pre-mRNA is still associated with a transcribing RNA polymerase II [14]. Transcription initiation of protein-coding genes involves the coordinated assembly of RNAPII and general transcription factors (GTFs) to form a stable pre-initiation complex (PIC) [15]. In budding yeast the assembly of the transcription preinitiation complex is regulated by two pathways that use the coactivators TFIID or SAGA (Spt-Ada-Gcn5-Acetyl transferase) [16]–[19]. SAGA complex regulate TATA-containing promoters that are generally related to stress response [20]. Only ~19% of yeast promoters contain TATA elements, and many of these (~10% of all yeast promoters) are dependent on SAGA coactivator function [21].

B- Splicing.

Precursor messenger RNA (pre-mRNA) splicing is a critical step in the posttranscriptional regulation of gene expression, providing significant expansion of the functional proteome of eukaryotic organisms with limited gene numbers [22]. Splicing constitutes an essential step of gene expression in which noncoding sequences (introns) are removed from pre-mRNAs and coding sequences (exons) are ligated together [23]. Splicing is therefore, a required process for proper protein expression. Splicing can also have an impact on transcription since the presence of introns can also confer increased transcriptional efficiency, possibly through increased initiation rates [24].

In both yeast and mammals, splicing is catalyzed by the spliceosome, a huge macromolecular assembly consisting of the U1, U2, U4, U5 and U6 small nuclear ribonucleoproteins (snRNPs) in conjunction with a large number of additional proteins [25]. Notably, spliceosome assembly needs to occur repeatedly every time an intron is removed from a pre-mRNA in a eukaryotic nucleus [26]. In yeast, most introns are defined by specific features mainly present in their sequences. For instance, the 5′ splice site (5′ss), the 3′ splice site (3′ss), and the branch point sequence (BPS) together with the polypyrimidine tract are necessary signals for proper splicing reaction [23], [27]. The sequence of 5′ss, 3′ss and BPS are well conserved in *Saccharomyces cerevisiae* (Figure 1) [28]. This sequence conservation is correlated with the fact that the vast majority of yeast introns are constitutively spliced, with only a few examples of alternative splicing. However, in metazoans, degenerate consensus splice sites may be a key feature that allows the generation of diverse alternative splicing patterns and may also lead to a requirement for additional protein factors to stabilize or target specific splice sites in a given tissue or cell type [26].

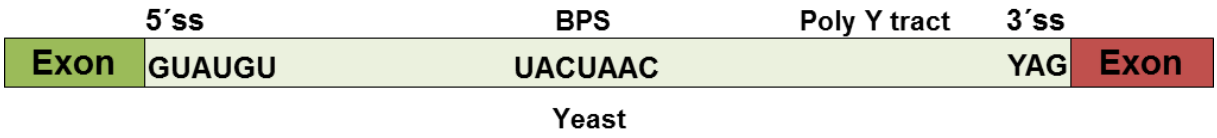


Figure 1: Pre-mRNA splicing recognition signals. Conserved sequences found at the 5′ and 3′ splice sites and branch site of U2-type pre-mRNA budding yeast (*S. cerevisiae*). Y = pyrimidine and R = purine. Figure adapted from [23].

Cells can generate multiple messenger RNAs (mRNAs) with different functions from a single genomic gene through alternative splicing [29], [30]. Types of alternative splicing include the use of alternative 5′ splice sites (alternative donor), alternative 3′ splice sites (alternative acceptor AA), exon skipping, intron retention and mutually exclusive exons (Figure 2) [31]. In

addition, new discoveries have shed light into the mechanisms responsible for the production of circular RNAs (circRNAs), which imply alternative splicing [32]–[34].

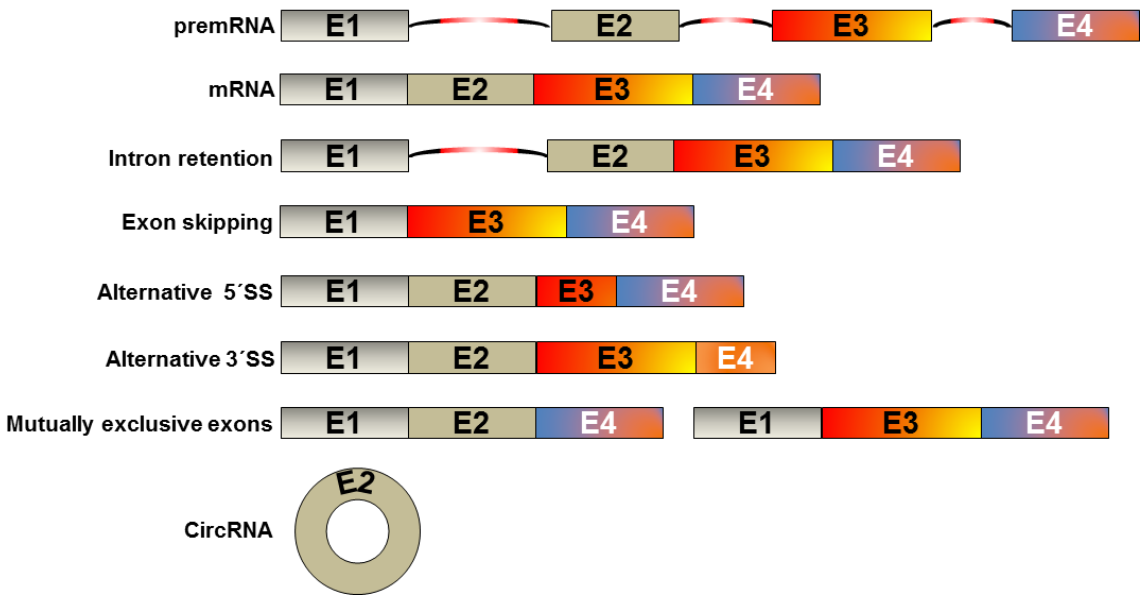


Figure 2: Species of mRNA produced as a consequence of alternative splicing.

C- mRNA export.

mRNA export is an indispensable and well-coordinated step during the complex process of eukaryotic gene expression. During the life of mRNA inside the nucleus it is packaged into a messenger ribonucleoprotein particle (mRNP), in order to transport it to the cytoplasm through the nuclear pore (NPC), which represents the link between the nucleus and cytoplasm [25], [35]. The transport process of mRNA is not restricted to the simple passage through the NPC, but also requires the association of the evolutionarily conserved heterodimeric transport factors Mex67-Mtr2 [36] with the concomitant formation of an export-competent mRNP [37].

One of the most important complexes in mRNP biogenesis and maturation is the TRanscription and EXport complex named TREX [38]. TREX is conserved across a wide range of organisms including *Saccharomyces cerevisiae* [39]–[41]. TREX regulates the specific transport of mRNA by Mex67-Mtr2 and guarantees the coupling of different steps during mRNA biogenesis [41], [42]. TREX complex in yeast consists of the THO subcomplex (Hpr1, Tho2, Mft1, Thp2 and Tex1), Sub2 and Yra1[10], [39], [43], [44].

A second complex also involved in transcription coupled to mRNA export in yeast and humans is the TREX-2 complex [15]. TREX-2 is composed by Sac3, Thp1, Sem1, Sus1 and Cdc31. It associates with the NPC through interaction with *NUP*1 and also with the transcription machinery via the SAGA transcriptional co-activator complex [45]. TREX-2 plays an important role in targeting a subset of transcriptionally active genes to NPCs trough a phenomenon known as "gene gating" [45]–[48].

D- mRNA Decay.

RNA Decay is the last station of mRNA life for transcripts that are not actively translated. The differential degradation of specific mRNAs can play an important role in setting the basal level of mRNA expression and how that mRNA level is modulated by environmental stimuli. This step is also essential, since RNA degradation removes the by-products of gene expression, including excised introns and other RNA pieces released during RNA processing. Different machineries are in charge of mRNA degradation. Some mRNA are degraded co-transcriptionally in the nucleus or once they are exported into the cytoplasm [49], [50].

SUS1: AN UNCOMMON TWO-INTRON YEAST GENE

In this dissertation, I have focused my studies on the *SUS1* gene due to its biological relevance and its peculiar genomic organization. Sus1 is an evolutionarily conserved 11-kDa nuclear protein present in all organisms ranging from yeast to human (hENY2). Sus1 is a component of two complexes described in this introduction: the SAGA coactivator and the nuclear pore-associated complex TREX-2 [45]. Moreover, Sus1 has been observed under specific circumstances also associated with stress granules and p-bodies [46], [51]. All these features make Sus1 study a good model to understand multiple cellular processes, including initiation of transcription, elongation, as well as mRNA export, biogenesis and decay (Figure 3). Furthermore, while most of *S. cerevisiae* genes are intronless, the *SUS1* gene consists of three exons (of 71, 140 and 77 nt) and two introns (of 80 and 70 nt), and its splicing and expression levels are regulated by different mechanisms [1], [52].

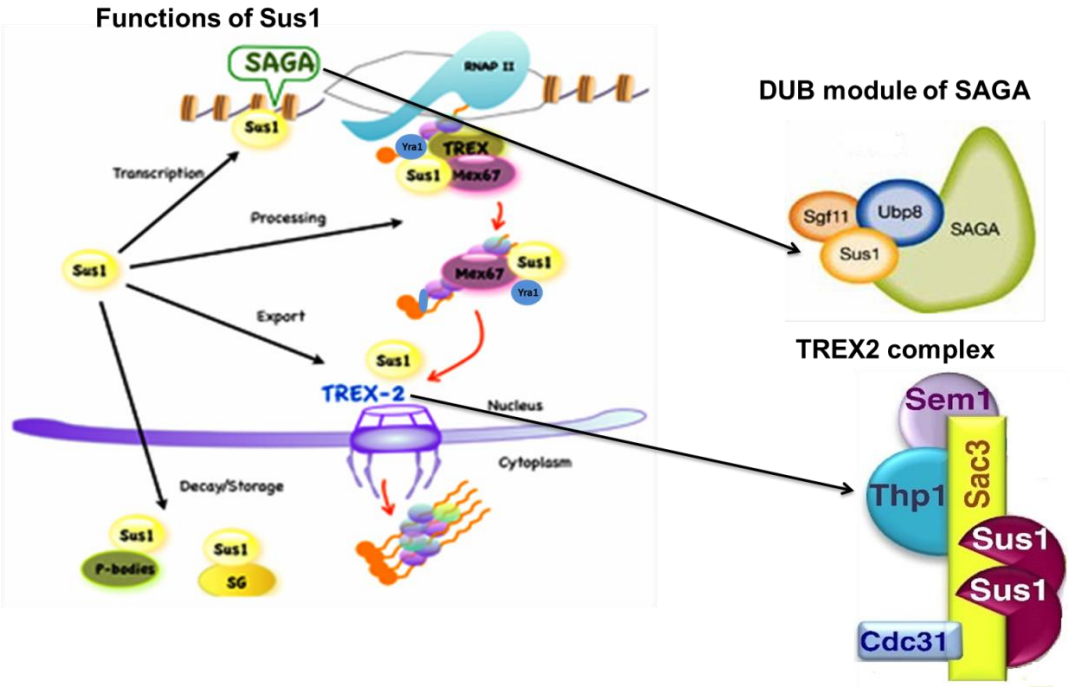


Figure 3: Functions of Sus1 (left), SAGA and TREX2 complexes (right). Figure adapted from [53], [54].

PRINCIPLES OF RNA STRUCTURE

A- Primary structure.

RNA is a polymer composed of nucleotides. Each RNA nucleotide is formed by three chemical entities: (1) a nitrogenous base, usually adenine, guanine, cytosine or uracil (Figure 4A); (2) a D-ribose sugar containing a 2'-hydroxyl group; and (3) a phosphate group (Figure 4B). The ring formed by the ribose is not planar, but tends to adopt different conformations where one of the five atoms of the ring (usually C2' or C3') is out of the plane defined by the remaining atoms. The most common conformation adopted by RNA sugars is C3'-*endo*, with the C3' atom out of plane. In an RNA polymer, the phosphate group of a nucleotide is covalently attached to the 5' position of its own ribose and to the 3' position of the preceding nucleotide. An RNA strand has a primary structure defined by the sequence of its successive nucleotides, with 5' and 3' termini [55], [56].

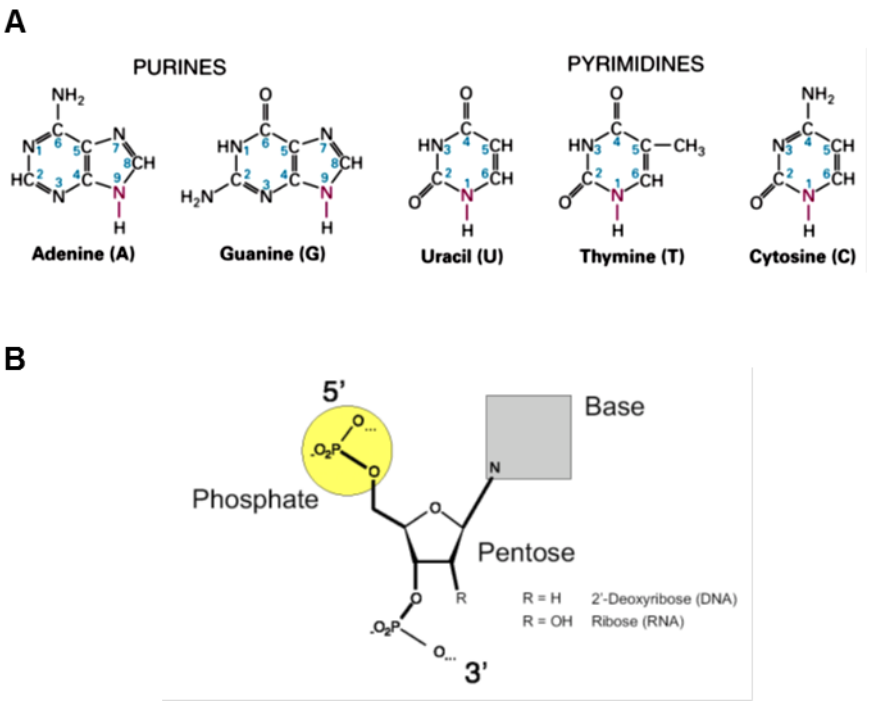


Figure 4: (A) The chemical structures of the principal bases of nucleic acids. (B) The components of the nucleotide. Figure adapted from [4], [57].

B- Secondary structure.

In contrast to DNA, an RNA polymer does not have a complementary strand but will contain complementary segments that have the potential to form double helices when the molecule folds back onto itself. These double helices represent the most important element of RNA secondary structure. A canonical RNA double helix is composed by Watson-Crick G:C and A:U base pairs and adopts an A-type conformation characterized by C3'-*endo* sugars, a deep and narrow major groove, and a wide and shallow minor groove [55], [56]. RNA double helices are interspersed with other secondary structure elements: apical loops, internal loops and helical intersections (Figure 5A). These secondary structure elements are usually stabilized by non-canonical base pairs and by base-base and sugar-base stacking interactions, and often contain nucleotides in a C2'-*endo* conformation. Some of the most abundant non-canonical RNA base pairs are wobble G:U pairs, sheared G:A pairs, and reverse Hoogsteen A:U pairs (Figure 5B) [58].

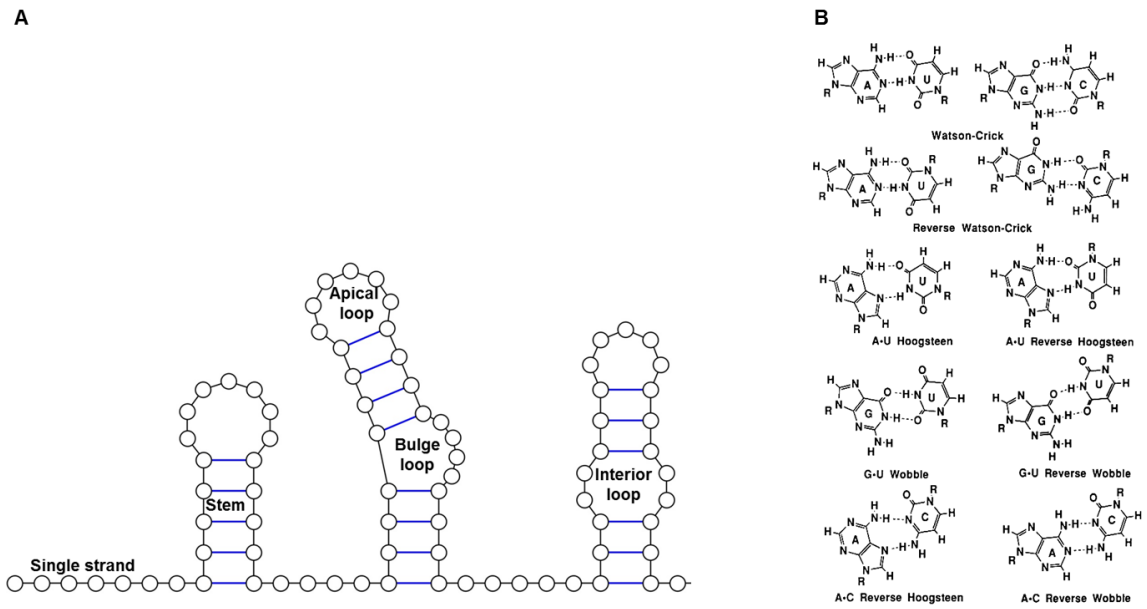


Figure 5: (A) RNA secondary structure elements. (B) Base pairs formed by RNA.

C- Tertiary structure.

Secondary structure elements may in turn be arranged in space to form three-dimensional tertiary structure (Figure 6), which is stabilized by noncovalent interactions including base stacking, van der Waals forces and hydrogen bonds. Some of the most frequent RNA tertiary structure motifs include coaxial stacking of double helices, loop-loop interactions, pseudoknots and triple helices. RNA tertiary structures can create binding sites for proteins, small ligands or metal ions, and can promote catalytic reactions. This leads to a variety of biochemical events that impact many aspects of gene expression processes [59]–[61].

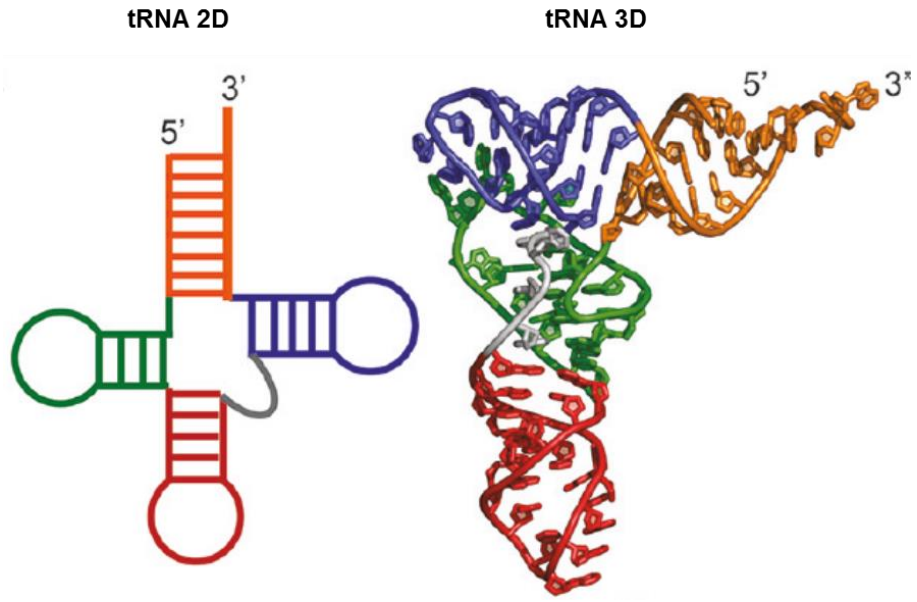


Figure 6: Secondary and tertiary structure of RNA. Secondary structure of a tRNA molecule (left), together with its tertiary structure (right). Figure adapted from [60].

RELATIONSHIP BETWEEN RNA STRUCTURE AND FUNCTION

The function exerted by RNA molecules is frequently determined by their structure. This is true not only for the "classical" mRNA, tRNA and rRNA molecules, but also for many of the recently identified non-coding RNAs [9].

Ribozymes provide important examples in this respect. They are catalytically active RNA molecules or RNA–protein complexes, in which solely the RNA provides catalytic activity. The term ribozyme refers to the enzymatic activity and ribonucleic acid nature at the same time. Ribozymes have been found in all kingdoms of life where they play crucial biological roles. Some well-established classes of natural ribozymes known to date are the hairpin, hammerhead, hepatitis delta virus (HDV), Varkud Satellite (VS), and *glmS* ribozymes, which are classified as small ribozymes. Group I and II introns, the ribosome, the spliceosome and RNase P are classified as large ribozymes [62]–[65].

Riboswitches provide another example of the relationship between RNA structure and function. They are highly structured RNA elements commonly found in the 5′-untranslated region (UTR) of mRNAs, which exert their regulatory control over the transcript in a *cis*-fashion by directly binding a small molecule ligand [66], [67].

pre-mRNA STRUCTURE AND GENE EXPRESSION

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 [68]. The effect of RNA secondary structure on the gene expression processes of S*accharomyces cerevisiae* has been addressed by several reports [69]–[75].

The process of splice site selection by the spliceosome has also been shown to be influenced by pre-mRNA structure in all eukaryotic organisms. This is usually accomplished by modulation of the accessibility or spatial distribution of splicing signal sequences via base-pairing [76] but also through the presence of more complex folds like riboswitches and ribozymes [77]–[80]. In yeast, for example, the *RP51B* gene contains a predicted stem-loop that brings the 5' and 3' splice sites into closer proximity. Splicing efficiency is reduced *in vivo* when the stem is mutated, and restored with compensatory mutations [81]–[83]. Another example is provided by the auto-regulated *YRA1* gene: shortening the intron improves splicing but abolishes autoregulation. Notably, proper regulation can be restored by inserting unrelated sequences into the shortened intron [84].

EXPERIMENTAL METHODS TO STUDY RNA STRUCTURE

Structure determination of biological RNAs and their complexes with proteins and other ligands is necessary to reveal the structural basis of the multiple biological functions of RNA. Three methods allow high-resolution analyses of RNA structure: X-ray crystallography [64], NMR spectroscopy and cryo-electron microscopy [85]. There are also low-resolution techniques that have a significant impact in the field, including SHAPE, native gel electrophoresis [86] and small-angle X-ray scattering [87] among others. We will next briefly describe the most important techniques applied in this thesis project, NMR spectroscopy and SHAPE.

A- Nuclear Magnetic Resonance (NMR) Spectroscopy.

NMR spectroscopy has proven to be a powerful and versatile tool for the analysis of RNA structure, dynamics and interactions, in solution conditions. The underlying principle of nuclear magnetic resonance is the detection in a magnetic field of atomic nuclei in a molecule that have non-zero magnetic moment [88]. In natural RNA samples, these nuclei are hydrogen and phosphorus, which can absorb electromagnetic radiation at a frequency (chemical shift) that

depends on the chemical environment of each specific nucleus. The analysis by NMR of an RNA molecule starts by identifying (assigning) the hydrogen and phosphorus atoms belonging to each nucleotide in the RNA sequence. To facilitate this process, the RNA is commonly labeled with magnetically active ^{13}C and ^{15}N isotopes that are present at very low abundance in natural samples. The presence of these nuclei allows the acquisition of additional heteronuclear experiments and the generation of more spectral dimensions, facilitating in this way the assignment process [89]–[92].

The NMR experiments applied in this thesis project are based on detecting Nuclear Overhauser Effects (NOEs) and scalar couplings. The NOE occurs when an excited nucleus transfers magnetization through space to nearby nuclei. In addition to being an essential component of the assignment process, this effect (detected with NOESY experiments) can be used to determine the three-dimensional structure of the RNA under study. The scalar couplings, in contrast, are transmitted through bonding electrons. They are also essential for assignments and provide information about dihedral angles. Scalar couplings can be detected through homonuclear (*e.g.* ^{1}H-^{1}H TOCSY, ^{1}H-^{1}H COSY) or heteronuclear (*e.g.* ^{15}N-^{1}H HSQC, ^{13}C-^{1}H HSQC) experiments. ^{15}N-^{1}H HNN-COSY experiments are particularly useful for studying RNA systems because they allow detection of hydrogen bonds between bases *via* two bond N-N couplings [93].

There is a limitation regarding the size of the RNA molecules that can be studied by NMR spectroscopy, imposed by signal overlap and line broadening effects. Up to now, this technique has allowed the determination of the three-dimensional structure of the packaging signal of HIV-1, composed of 155 nucleotides [94].

NMR studies of RNA molecules require milligram quantities of pure material. RNA sequences aimed for NMR spectroscopy analyses are usually synthesized by in vitro transcription with a DNA-dependent RNA polymerase, such as T7 RNA polymerase [90], [95], [96].

B- Selective 2´-hydroxyl acylation analyzed by primer extension (SHAPE).

The nucleophilic reactivity of the ribose 2′-hydroxyl group in RNA is determined by local nucleotide flexibility. This group is reactive at conformationally flexible positions but unreactive at nucleotides constrained by base pairing in canonical helices. Sites of modification can be efficiently detected with single-nucleotide resolution using primer extension, with small sample quantities [97]. This relatively simple technique can in principle generate information about the secondary structure of long and flexible RNA molecules that are beyond the limits imposed by high-resolution methods.

SHAPE can be applied in vitro or in cellular environments [98]. The SHAPE technology involves covalently modifying the RNA by a SHAPE reagent, such as isatoic anhydride (IA), N-methyl isatoic anhydride (NMIA), 1-methyl-7-nitro-isatoic anhydride (1M7), or benzoyl cyanide (BzCN). Following selective 2′-hydroxyl modification, covalent SHAPE adducts are detected by reverse transcriptase-mediated primer extension. DNA synthesis by reverse transcriptase stops one nucleotide prior to the position of an adduct (Figure 7A) [99].

The SHAPE technology can be implemented in an efficient and high-throughput way with automated capillary electrophoresis, using DNA sequencing instruments (Figure 7B). The capillary electrophoresis data are analyzed using the software program QuShape [100]. After normalization, the reactivity values can be converted to ΔG_{SHAPE} pseudo-free energy terms and used as constraints in RNA folding algorithms such as RNAStructure [101] to generate an experimentally supported secondary structure model of the RNA molecule under study [97], [102].

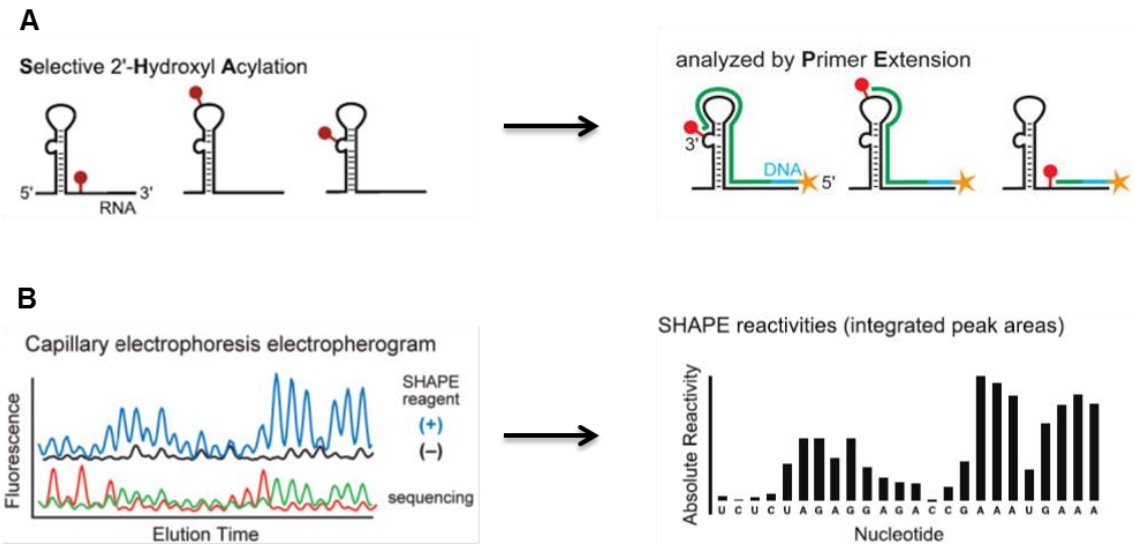


Figure 7: scheme showing: (A) RNA modication by a SHAPE reagent, followed by reverse transcriptase-mediated primer extension. (B) The raw capillary electrophoresis data from which values of SHAPE reactivity are obtained. Figure adapted from [99], [103].

RNA STRUCTURE PREDICTION METHODS

A large number of computational programs have been developed to predict both the secondary structure of RNA sequences and their three-dimensional folding [104].

The best known algorithms predict the secondary structure of a single input RNA sequence by finding the minimum free energy structure with respect to a given thermodynamic model. This is usually accomplished by maximizing the number of G:C, A:U and G:U base pairs [105], [106]. Examples of these algorithms include Mfold [107], RNAfold [108] and RNAStructure [101]. Due to deficiencies in the thermodynamic models or the folding parameters, the predicted structure can be different from the experimental one, particularly for longer sequences.

REFERENCES

[1] A. AbuQattam, J. Gallego, and S. Rodríguez-Navarro, "An intronic RNA structure modulates expression of the mRNA biogenesis factor Sus1.," *RNA*, vol. 22, no. 1, pp. 75–86, 2016.

[2] T. Carlomagno, "Present and future of NMR for RNA-protein complexes: A perspective of integrated structural biology," *J. Magn. Reson.*, vol. 241, no. 1, pp. 126–136, 2014.

[3] R. Cammack *et al.*, "Oxford Dictionary of Biochemistry and Molecular Biology." "Oxford University Press," 2008.

[4] A. Uzman, H. Lodish, A. Berk, L. Zipursky, and D. Baltimore, "Molecular Cell Biology (4th edition) New York, NY, 2000, ISBN 0-7167-3136-3," *Biochem. Mol. Biol. Educ.*, vol. 29, p. Section 1.2The Molecules of Life, 2000.

[5] E. P. Geiduschek and R. Haselkorn, "Messenger RNA," *Annu. Rev. Biochem.*, vol. 38, no. 1, pp. 647–676, Jun. 1969.

[6] H. Lamfrom, "Factors determining the specificity of hemoglobin synthesized in a cell-free system," *J. Mol. Biol.*, vol. 3, no. 3, pp. 241–252, 1961.

[7] N. Lehman, "RNA in evolution," *Wiley Interdiscip. Rev. RNA*, vol. 1, no. 2, pp. 202–213, 2010.

[8] A. Lazcano, R. Guerrero, L. Margulis, and J. Oró, "The evolutionary transition from RNA to DNA in early cells," *J. Mol. Evol.*, vol. 27, no. 4, pp. 283–290, 1988.

[9] T. R. Cech and J. A. Steitz, "The noncoding RNA revolution - Trashing old rules to forge new ones," *Cell*, vol. 157, no. 1. pp. 77–94, 2014.

[10] J. Katahira, "Nuclear export of messenger RNA," *Genes (Basel).*, vol. 6, no. 2, pp. 163–184, 2015.

[11] G. Haimovich *et al.*, “Gene Expression Is Circular: Factors for mRNA Degradation Also Foster mRNA Synthesis,” *Cell*, vol. 153, no. 5, pp. 1000–1011, May 2013.

[12] V. O. Wickramasinghe and R. A. Laskey, “Control of mammalian gene expression by selective mRNA export,” *Nat. Rev. Mol. Cell Biol.*, vol. 16, no. 7, pp. 431–442, 2015.

[13] A. Griffiths, J. Miller, and D. Suzuki, *Introduction to Genetic Analysis. 7th edition.*, vol. 2021. 2000.

[14] G. Haimovich, M. Choder, R. H. Singer, and T. Trcek Pulisic, “The fate of the messenger is pre-determined: A new model for regulation of gene expression.,” *Biochim. Biophys. Acta*, pp. 1–11, 2013.

[15] E. García-Oliver, V. García-Molinero, and S. Rodríguez-Navarro, “MRNA export and gene expression: The SAGA-TREX-2 connection,” *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, vol. 1819, no. 6. pp. 555–565, 2012.

[16] L. Kuras, P. Kosa, M. Mencia, and K. Struhl, “TAF-Containing and TAF-independent forms of transcriptionally active TBP in vivo.,” *Science*, vol. 288, pp. 1244–1248, 2000.

[17] X. Y. Li, S. R. Bhaumik, and M. R. Green, “Distinct classes of yeast promoters revealed by differential TAF recruitment.,” *Science*, vol. 288, no. May, pp. 1242–1244, 2000.

[18] H. Qiu, C. Hu, S. Yoon, K. Natarajan, M. J. Swanson, and A. G. Hinnebusch, “An array of coactivators is required for optimal recruitment of TATA binding protein and RNA polymerase II by promoter-bound Gcn4p.,” *Mol. Cell. Biol.*, vol. 24, no. 10, pp. 4104–17, 2004.

[19] S. Hahn and E. T. Young, “Transcriptional regulation in saccharomyces cerevisiae: Transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators,” *Genetics*, vol. 189, no. 3, pp. 705–736, 2011.

[20] K. L. Huisinga and B. F. Pugh, “A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae,” *Mol. Cell*, vol. 13, no. 4, pp. 573–585, 2004.

[21] S.-J. Wei *et al.*, “Identification of a specific motif of the DSS1 protein required for proteasome interaction and p53 protein degradation.,” *J. Mol. Biol.*, vol. 383, no. 3, pp. 693–712, Nov. 2008.

[22] Y. Lee, D. C. Rio, S. Biology, and C. Biology, “Mechanisms and Regulation of Alternative Pre-mRNA Splicing,” *Annu.Rev.Biochem.*, no. 258, pp. 291–323, 2015.

[23] C. L. Will and R. Lührmann, “Spliceosome structure and function,” *Cold Spring Harb. Perspect. Biol.*, vol. 3, no. 7, pp. 1–2, 2011.

[24] R. L. Brinster, J. M. Allen, R. R. Behringer, R. E. Gelinas, and R. D. Palmiter, “Introns increase transcriptional efficiency in transgenic mice.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 85, no. 3, pp. 836–40, 1988.

[25] S. Hocine, R. H. Singer, and D. Grünwald, “RNA processing and export.,” *Cold Spring Harbor perspectives in biology*, vol. 2, no. 12. 2010.

[26] Y. Lee and D. C. Rio, “Mechanisms and Regulation of Alternative Pre-mRNA Splicing,” *Annu. Rev. Biochem.*, vol. 84, no. 1, pp. 291–312, 2014.

[27] Z. Wang and C. B. Burge, “Splicing regulation : From a parts list of regulatory elements to an integrated splicing code Splicing regulation : From a parts list of regulatory elements to an integrated splicing code,” *Rna*, vol. 14, no. 617, pp. 802–813, 2008.

[28] E. Bon *et al.*, “Molecular evolution of eukaryotic genomes: Hemiascomycetous yeast spliceosomal introns,” *Nucleic Acids Research*, vol. 31, no. 4. pp. 1121–1135, 2003.

[29] A. Kalsotra and T. A. Cooper, “Functional consequences of developmentally regulated alternative splicing,” *Nat Rev Genet*, vol. 12, no. 10, pp. 715–729, 2011.

[30] Z. Xiping, W. Qingshan, Z. Shuai, Y. Hongjian, and D. Xiaowen, “A summary of relationships between alternative splicing and breast cancer,” 2017.

[31] L. P. Iñiguez and G. Hernández, “The evolutionary relationship between alternative splicing and gene duplication,” *Frontiers in Genetics*, vol. 8, no. FEB. 2017.

[32] S. P. Barrett, P. L. Wang, and J. Salzman, “Circular RNA biogenesis can proceed through an exon-containing lariat precursor,” *Elife*, vol. 4, p. e07540, 2015.

[33] L.-L. Chen and L. Yang, “Regulation of circRNA biogenesis,” *RNA Biol.*, vol. 12, no. 4, pp. 381–388, 2015.

[34] a Surono, Y. Takeshima, T. Wibawa, M. Ikezawa, I. Nonaka, and M. Matsuo, “Circular dystrophin RNAs consisting of exons that were skipped by alternative splicing.,” *Hum. Mol. Genet.*, vol. 8, no. 3, pp. 493–500, 1999.

[35] M. Stewart, “Nuclear export of mRNA,” *Trends in Biochemical Sciences*, vol. 35, no. 11. pp. 609–617, 2010.

[36] S. Rother and K. Sträßer, “mRNA Export-An integrative component of gene

expression," in *Nuclear Transport*, 2009, pp. 101–117.

[37] S. R. Carmody and S. R. Wente, "mRNA nuclear export at a glance," *J. Cell Sci.*, vol. 122, no. 12, pp. 1933–1937, Jun. 2009.

[38] K. Strässer *et al.*, "TREX is a conserved complex coupling transcription with messenger RNA export.," *Nature*, vol. 417, no. 6886, pp. 304–308, 2002.

[39] S. Rodríguez-Navarro and E. Hurt, "Linking gene regulation to mRNA production and export," *Current Opinion in Cell Biology*, vol. 23, no. 3. pp. 302–309, 2011.

[40] A. Segref *et al.*, "Mex67p, a novel factor for nuclear mRNA export. Binds to both poly(A)+ RNA and nuclear pores," *EMBO J.*, vol. 16, no. 11, pp. 3256–3271, 1997.

[41] J. Katahira, "MRNA export and the TREX complex," *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, vol. 1819, no. 6. pp. 507–513, 2012.

[42] C. G. Heath, N. Viphakone, and S. A. Wilson, "The role of TREX in gene expression and disease," *Biochem. J.*, vol. 473, no. 19, pp. 2911–2935, 2016.

[43] A. Köhler and E. Hurt, "Exporting RNA from the nucleus to the cytoplasm," *Nat. Rev. Mol. Cell Biol.*, vol. 8, no. 10, pp. 761–773, 2007.

[44] R. Reed and H. Cheng, "TREX, SR proteins and export of mRNA," *Current Opinion in Cell Biology*, vol. 17, no. 3. pp. 269–273, 2005.

[45] S. Rodríguez-Navarro *et al.*, "Sus1, a Functional Component of the SAGA Histone Acetylase Complex and the Nuclear Pore-Associated mRNA Export Machinery," *Cell*, vol. 116, no. 1, pp. 75–86, 2004.

[46] P. Pascual-García *et al.*, "Sus1 is recruited to coding regions and functions during transcription elongation in association with SAGA and TREX2," *Genes Dev.*, vol. 22, no. 20, pp. 2811–2822, 2008.

[47] A. Köhler, M. Schneider, G. G. Cabal, U. Nehrbass, and E. Hurt, "Yeast Ataxin-7 links histone deubiquitination with gene gating and mRNA export," *Nat. Cell Biol.*, vol. 10, no. 6, pp. 707–715, 2008.

[48] D. Jani *et al.*, "Sus1, Cdc31, and the Sac3 CID region form a conserved interaction platform that promotes nuclear pore association and mRNA export.," *Mol. Cell*, vol. 33, no. 6, pp. 727–37, Mar. 2009.

[49] R. Parker, "RNA degradation in Saccharomyces cerevisae," *Genetics*, vol. 191, no. 3, pp.

671–702, 2012.

[50] W. Hu, T. J. Sweet, S. Chamnongpol, K. E. Baker, and J. Coller, “Co-translational mRNA decay in Saccharomyces cerevisiae,” *Nature*, vol. 461, no. 7261, pp. 225–229, 2009.

[51] B. Cuenca-Bono, V. García-Molinero, P. Pascual-García, E. García-Oliver, A. Llopis, and S. Rodríguez-Navarro, “A novel link between Sus1 and the cytoplasmic mRNA decay machinery suggests a broad role in mRNA metabolism,” *BMC Cell Biol.*, vol. 11, no. 1, p. 19, 2010.

[52] B. Cuenca-Bono *et al.*, “SUS1 introns are required for efficient mRNA nuclear export in yeast,” *Nucleic Acids Res.*, vol. 39, no. 19, pp. 8599–8611, 2011.

[53] A. Galán and S. Rodríguez-Navarro, “Sus1/ENY2: a multitasking protein in eukaryotic gene expression,” *Crit. Rev. Biochem. Mol. Biol.*, vol. 47, no. 6, pp. 556–568, 2012.

[54] S. Rodríguez-Navarro, “Insights into SAGA function during gene expression,” *EMBO Rep.*, vol. 10, no. 8, pp. 843–850, 2009.

[55] J. H. Parish, “Principles of nucleic acid structure: By W Saenger. pp 556. Springer-Verlag, New York. 1984. DM 79. ISBN 3-540-90761-0,” *Biochem. Educ.*, vol. 13, no. 2, p. 92, 1985.

[56] S. Neidle, “RNA Structures and Their Diversity BT - Principles of Nucleic Acid Structure,” New York: Academic Press, 2008, pp. 204–248.

[57] T. Carlomagno, “NMR as a tool for structure determination of nucleic acids.”

[58] N. B. Leontis, “The non-Watson-Crick base pairs and their associated isostericity matrices,” *Nucleic Acids Res.*, vol. 30, no. 16, pp. 3497–3531, 2002.

[59] P. A. Sharp, “The Centrality of RNA,” *Cell*, vol. 136, no. 4. pp. 577–580, 2009.

[60] S. E. Butcher and A. M. Pyle, “The Molecular Interactions That Stabilize RNA Tertiary Structure: RNA Motifs, Patterns, and Networks,” *Acc. Chem. Res.*, vol. 44, no. 12, pp. 1302–1311, Dec. 2011.

[61] R. T. Batey, R. P. Rambo, and J. A. Doudna, “Tertiary motifs in RNA structure and folding,” *Angewandte Chemie - International Edition*, vol. 38, no. 16. pp. 2326–2343, 1999.

[62] M. Skilandat and R. K. O. Sigel, “Ribozymes A2 - Maloy, Stanley,” K. B. T.-B. E. of G.

(Second E. Hughes, Ed. San Diego: Academic Press, 2013, pp. 254–258.

[63] C. R. Burke and A. Lupták, “Catalytic RNA,” in *eLS*, 2012.

[64] T. R. Cech, “Structural biology. The ribosome is a ribozyme.,” *Science*, vol. 289, no. 5481, pp. 878–9, 2000.

[65] S. M. Fica *et al.*, “RNA catalyses nuclear pre-mRNA splicing,” *Nature*, 2013.

[66] A. Serganov and D. J. Patel, “Ribozymes, riboswitches and beyond: regulation of gene expression without proteins,” *Nat. Rev. Genet.*, vol. 8, no. 10, pp. 776–790, 2007.

[67] A. D. Garst, A. L. Edwards, and R. T. Batey, “Riboswitches : Structures and Mechanisms,” pp. 1–14, 2017.

[68] Y. Wan *et al.*, “Landscape and variation of RNA secondary structure across the human transcriptome,” *Nature*, vol. 505, no. 7485, pp. 706–709, 2014.

[69] E. Buratti and F. E. Baralle, “Influence of RNA Secondary Structure on the Pre-mRNA Splicing Process,” *Mol. Cell. Biol.*, vol. 24, no. 24, pp. 10505–10514, 2004.

[70] A. Mougin *et al.*, “Secondary structure of the yeast Saccharomyces cerevisiae pre-U3A snoRNA and its implication for splicing efficiency.,” *RNA*, vol. 2, no. 11, pp. 1079–1093, 1996.

[71] V. Goguel, Y. Wang, and M. Rosbash, “Short artificial hairpins sequester splicing signals and inhibit yeast pre-mRNA splicing.,” *Mol. Cell. Biol.*, vol. 13, no. 11, pp. 6841–8, 1993.

[72] H. Hieronymus and P. a Silver, “A systems view of mRNP biology.,” *Genes Dev.*, vol. 18, no. 23, pp. 2845–60, Dec. 2004.

[73] X. Wang, P. Li, and R. N. Gutenkunst, “Systematic effects of mRNA secondary structure on gene expression and molecular function in budding yeast,” 2017.

[74] M. Kertesz *et al.*, “Genome-wide measurement of RNA secondary structure in yeast,” *Nature*, vol. 467, no. 7311, pp. 103–107, 2010.

[75] Y. Arava, Y. Wang, J. D. Storey, C. L. Liu, P. O. Brown, and D. Herschlag, “Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 7, pp. 3889–3894, 2003.

[76] J. Pérez-Valle and J. Vilardell, “Intronic features that determine the selection of the 3′

splice site," *Wiley Interdisciplinary Reviews: RNA*, vol. 3, no. 5. pp. 707–717, 2012.

[77] M. B. Warf and J. A. Berglund, "The role of RNA structure in regulating pre-mRNA splicing," *Trends Biochem. Sci.*, vol. 35, no. 3, pp. 169–178, Mar. 2010.

[78] C. J. McManus and B. R. Graveley, "RNA structure and the mechanisms of alternative splicing," *Curr. Opin. Genet. Dev.*, vol. 21, no. 4, pp. 373–379, Aug. 2011.

[79] Y. Jin, Y. Yang, and P. Zhang, "New insights into RNA secondary structure in the alternative splicing of pre-mRNAs," *RNA Biol.*, vol. 8, no. 3, pp. 450–457, 2011.

[80] I. Garcia-Robles, J. Sanchez-Navarro, and M. D. La Pena, "Intronic hammerhead ribozymes in mRNA biogenesis," *Biol. Chem.*, vol. 393, no. 11, pp. 1317–1326, 2012.

[81] D. Libri, F. Stutz, T. McCarthy, and M. Rosbash, "RNA structural patterns and splicing: molecular basis for an RNA-based enhancer.," *RNA*, vol. 1, no. 4, pp. 425–36, 1995.

[82] V. Goguel and M. Rosbash, "Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast," *Cell*, vol. 72, no. 6, pp. 893–901, 1993.

[83] B. Charpentier and M. Rosbash, "Intramolecular structure in yeast introns aids the early steps of in vitro spliceosome assembly," *RNA*, vol. 2, no. 6, pp. 509–522, 1996.

[84] P. J. Preker and C. Guthrie, "Autoregulation of the mRNA export factor Yra1p requires inefficient splicing of its pre-mRNA.," *RNA*, vol. 12, no. 6, pp. 994–1006, 2006.

[85] E. Callaway, "The revolution will not be crystallized: a new method sweeps through structural biology," *Nature*, vol. 525, no. 7568, pp. 172–174, 2015.

[86] D. M. J. Lilley, A. Bhattacharyya, S. P. McAteer, and D. R. Duckett, "Gel electrophoretic analysis of the structure of nucleic acids," *Biochem. Soc. Trans.*, vol. 21, no. 1, p. 111 LP-116, Feb. 1993.

[87] X. Fang, J. R. Stagno, Y. R. Bhandari, X. Zuo, and Y. X. Wang, "Small-angle X-ray scattering: A bridge between RNA secondary structures and three-dimensional topological structures," *Current Opinion in Structural Biology*, vol. 30. pp. 147–160, 2015.

[88] R. J. Abraham, "J. K. M. Sanders and B. K. Hunter. Modern NMR spectroscopy, a guide for chemists. Oxford University Press, Oxford, England; pp 302," *Magn. Reson. Chem.*, vol. 25, no. 12, pp. 1092–1092, Dec. 1987.

[89] B. Fürtig, C. Richter, J. Wöhnert, and H. Schwalbe, “NMR spectroscopy of RNA,” *ChemBioChem*, vol. 4, no. 10. pp. 936–962, 2003.

[90] M. P. Latham, D. J. Brown, S. a McCallum, and A. Pardi, “NMR methods for studying the structure and dynamics of RNA.,” *Chembiochem*, vol. 6, no. 9, pp. 1492–505, 2005.

[91] G. Varani, F. Aboul-ela, and F. H.-T. Allain, “NMR investigation of RNA structure,” *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 29, no. 1–2, pp. 51–127, Jun. 1996.

[92] G. Varani and I. Tinoco, “RNA structure and NMR spectroscopy.,” *Q. Rev. Biophys.*, vol. 24, no. 4, pp. 479–532, 1991.

[93] A. J. Dingley and S. Grzesiek, “Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide 2J(NN) couplings,” *J. Am. Chem. Soc.*, vol. 120, no. 33, pp. 8293–8297, 1998.

[94] S. C. Keane *et al.*, “Structure of the HIV-1 RNA Packaging Signal,” *Science*, vol. 348, no. 6237, pp. 917–921, May 2015.

[95] B. Beckert and B. Masquida, “Synthesis of RNA by In Vitro Transcription BT - RNA: Methods and Protocols,” H. Nielsen, Ed. Totowa, NJ: Humana Press, 2011, pp. 29–41.

[96] J. F. Milligan, D. R. Groebe, G. W. Witherell, and O. C. Uhlenbeck, “Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates.,” *Nucleic Acids Res.*, vol. 15, no. 21, pp. 8783–8798, Nov. 1987.

[97] K. M. Weeks and D. M. Mauger, “Exploring RNA structural codes with SHAPE chemistry,” *Acc. Chem. Res.*, vol. 44, no. 12, pp. 1280–1291, 2011.

[98] A. López-Carrasco, S. Gago-Zachert, G. Mileti, S. Minoia, R. Flores, and S. Delgado, “The transcription initiation sites of eggplant latent viroid strands map within distinct motifs in their in vivo RNA conformations,” *RNA Biol.*, vol. 13, no. 1, pp. 83–97, Jan. 2016.

[99] E. J. Merino, K. A. Wilkinson, J. L. Coughlan, and K. M. Weeks, “RNA structure analysis at single nucleotide resolution by Selective 2′-Hydroxyl Acylation and Primer Extension (SHAPE),” *J. Am. Chem. Soc.*, vol. 127, no. 12, pp. 4223–4231, 2005.

[100] F. Karabiber, J. L. McGinnis, O. V. Favorov, and K. M. Weeks, “QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis,” *RNA*, vol. 19, no. 1, pp. 63–73, 2013.

[101] S. Bellaousov, J. S. Reuter, M. G. Seetin, and D. H. Mathews, “RNAstructure: Web servers for RNA secondary structure prediction and analysis.,” *Nucleic Acids Res.*, vol. 41, no. Web Server issue, 2013.

[102] J. T. Low and K. M. Weeks, “SHAPE-directed RNA secondary structure prediction,” *Methods*, vol. 52, no. 2. pp. 150–158, 2010.

[103] R. C. Spitale, R. A. Flynn, E. A. Torre, E. T. Kool, and H. Y. Chang, “RNA structural analysis by evolving SHAPE chemistry,” *Wiley interdisciplinary reviews. RNA*, vol. 5, no. 6. pp. 867–881, 2014.

[104] “RNA structure prediction,” *RNA Structure and Folding Biophysical Techniques and Prediction Methods*. De Gruyter, Berlin, Boston, 2013.

[105] A. Condon, “Problems on RNA Secondary Structure Prediction and Design,” *Autom. Lang. Program.*, vol. 2719, pp. 22–32, 2003.

[106] F. Major and R. Griffey, “Computational methods for RNA structure determination,” *Curr. Opin. Struct. Biol.*, vol. 11, no. 3, pp. 282–286, 2001.

[107] M. Zuker, “Mfold web server for nucleic acid folding and hybridization prediction,” *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3406–3415, 2003.

[108] A. R. Gruber, R. Lorenz, S. H. Bernhart, R. Neuböck, and I. L. Hofacker, “The Vienna RNA websuite.,” *Nucleic Acids Res.*, vol. 36, no. Web Server issue, 2008.

OBJECTIVES

In this thesis project I have studied RNA structural elements important for *Saccharomyces cerevisiae SUS1* mRNA metabolism, expression and function. In particular, my research has focused on identifying and characterizing structures formed by the *SUS1* pre-mRNA, and studying their functional relevance. The main objectives of this thesis were the following ones:

- Structural and functional characterization of the downstream intron of the *S. cerevisiae SUS1* gene (Chapter 1);
- Structural and Functional characterization of the middle exon of the same gene (Chapter 2).

RESULTS AND DISCUSSION

CHAPTER 1

An intronic RNA structure modulates expression of the mRNA biogenesis factor Sus1

These results were published in the ***RNA Journal*** RNA 2016. 22: 75-86

The full text of the article is presented in this chapter.

An intronic RNA structure modulates expression of the mRNA biogenesis factor Sus1

ALI ABUQATTAM,[1,2] JOSÉ GALLEGO,[2] and SUSANA RODRÍGUEZ-NAVARRO[1]
[1]Gene Expression and RNA Metabolism Laboratory, Centro de Investigación Príncipe Felipe, Valencia 46012, Spain
[2]Facultad de Medicina, Universidad Católica de Valencia, Valencia 46001, Spain

An intronic RNA structure modulates expression of the mRNA biogenesis factor Sus1

Ali AbuQattam[a,b], José Gallego[b*] and Susana Rodríguez-Navarro[a*]

[a] Gene Expression and RNA Metabolism Laboratory. Centro de Investigación Príncipe Felipe, C/ E. Primo Yúfera 3, 46012 Valencia

[b] Facultad de Medicina. Universidad Católica de Valencia, C/Quevedo 2, 46001 Valencia

[*] Correspondence to: srodriguez@cipf.es and jose.gallego@ucv.es

Phone 0034963289680

FAX 0034963289701

Abstract. Sus1 is a conserved protein involved in chromatin remodelling and mRNA biogenesis. Unlike most yeast genes, the *SUS1* pre-mRNA of *Saccharomyces cerevisiae* contains two introns and is alternatively spliced, retaining one or both introns in response to changes in environmental conditions. *SUS1* splicing may allow the cell to control Sus1 expression, but the mechanisms that regulate this process remain unknown. Using *in silico* analyses together with NMR spectroscopy, gel electrophoresis and UV thermal denaturation experiments, we show that the downstream intron (I2) of *SUS1* forms a weakly-stable, 37-nucleotide stem-loop structure containing the branch site near its apical loop and the 3' splice site after the stem terminus. A cellular assay revealed that two of four mutants containing altered I2 structures had significantly impaired *SUS1* expression. Semi-quantitative RT-PCR experiments indicated that all mutants accumulated unspliced *SUS1* pre-mRNA and/or induced distorted levels of fully spliced mRNA relative to wild-type. Concomitantly, Sus1 cellular functions in histone H2B deubiquitination and mRNA export were affected in I2 hairpin mutants that inhibited splicing. This work demonstrates that I2 structure is relevant for *SUS1* expression, and that this effect is likely exerted through modulation of splicing.

Keywords: Sus1, gene expression, splicing, yeast, NMR, RNA, structure, thermal stability

Introduction

Introns are removed from pre-mRNAs to create a mature transcript by the action of the splicing machinery (Fica et al. 2013; Wahl et al. 2009). Recent reports have shown that splicing contributes to the control of gene expression (Acuña and Kornblihtt 2014; Moehle et al. 2014), and can *fine-tune* different stages of the gene expression process by regulating the expression levels of key pathway components involved in RNP and ribosome biogenesis, transcription and RNA transport (Johnson and Vilardell 2012). Some of these studies have focused on *Saccharomyces cerevisiae* Sus1, a factor involved in transcription and mRNA export (Rodriguez-Navarro et al. 2004; Galan and Rodríguez-Navarro 2012). Sus1 is part of the SAGA (Spt-Ada-Gcn5-Acetyltransferase) complex where it participates as a modulator of the deubiquitinase activity of Ubp8 (Kohler et al. 2006). It also interacts at the nuclear pore with TREX-2 (Transcription Export complex 2) and participates in mRNA export and genome stability (Gonzalez-Aguilera et al. 2008). In addition, Sus1 transiently interacts during transcription elongation with the RNA Polymerase II and the export factors Yra1 and Mex67 (Pascual-García et al. 2008). An interesting feature of *SUS1* gene -highly infrequent in the *S. cerevisiae* genome- is the fact that it contains two introns (Rodriguez-Navarro et al. 2004). The presence of two introns in *SUS1* gene opens up the possibility of a feedback mechanism involving splicing. A similar situation was found in the *YRA1* gene encoding another essential mRNA export factor, whose expression was found to be controlled by different mRNA biogenesis steps (Rodríguez-Navarro et al. 2002; Dong et al. 2007; Preker and Guthrie 2006).

The functional relevance of the presence of two introns in the *SUS1* pre-mRNA has been studied previously (Cuenca-Bono et al. 2011; Hossain et al. 2011). While intron 2 (I2) has consensus splice sequences, intron 1 (I1) is characterized by the presence of non-consensus 5' splice site (SS) (GUAUGA instead of canonical GUAUGU) and Branch site (BS) (UACUGAC instead of UACUAAC). These sequences influence *SUS1* splicing and as a consequence intron 1 is retained in more than 15% of the *SUS1* transcripts. Moreover, retention of I1 is affected by

growth conditions, suggesting a potential role for splicing in regulating Sus1 cellular function (Johnson and Vilardell 2012).

RNA structure is emerging as an important element in the regulation of gene expression, via modulation of interactions with proteins and microRNAs, or by arranging the spatial distribution of important sequences (Wan et al. 2015). For example, in the case of splicing intron RNA folding has been shown to influence 3'SS selection through the masking of AG sequences inside double-helical stems, or by bringing BS and 3' SS sequences into functional proximity (Warf and Berglund 2010; Pérez-Valle and Vilardell 2012; Meyer et al. 2011; Deshler and Rossi 1991; Rogic et al. 2008). In *S. cerevisiae*, intronic RNA structures have been reported to play a role in autoregulating *RPL30* and *YRA1* expression at different levels (Meyer et al. 2011; Dong et al. 2010). Gaining insight into the role of RNA structure in controlling mRNA biogenesis at different stages is a major challenge in the field. In this report, we explore the regulatory role of a stem-loop structure formed by I2 of *SUS1*.

Results

Intron 2 of SUS1 forms a weakly-stable hairpin structure in solution containing BS nucleotides in its apical loop

The RNA-folding algorithm Mfold (Zuker 2003) predicted that I2 of *S. cerevisiae SUS1* formed a 37-nucleotide (nt) stem-loop structure with BS nt lying in the apical loop and 3'SS nt immediately after the 3' stem terminus (Fig. 1 and Hossain et al., 2011). The pre-mRNAs of several yeast species, selected from our previous evolutionary analysis of *SUS1* (Cuenca-Bono et al. 2011), were also subjected to secondary structure predictions. The analyses revealed comparable putative hairpins formed by the second intron of *S. pastorianus, K. thermotolerans, S. kudriavzevii* and *S. mikatae*. In these hairpins the BS nt similarly localised close to the apical loop, and the 3'SS near the 3' terminus of the base-paired stem (Fig. 1). These results were

generally consistent with previous predictions *in silico* (Hossain et al. 2011).

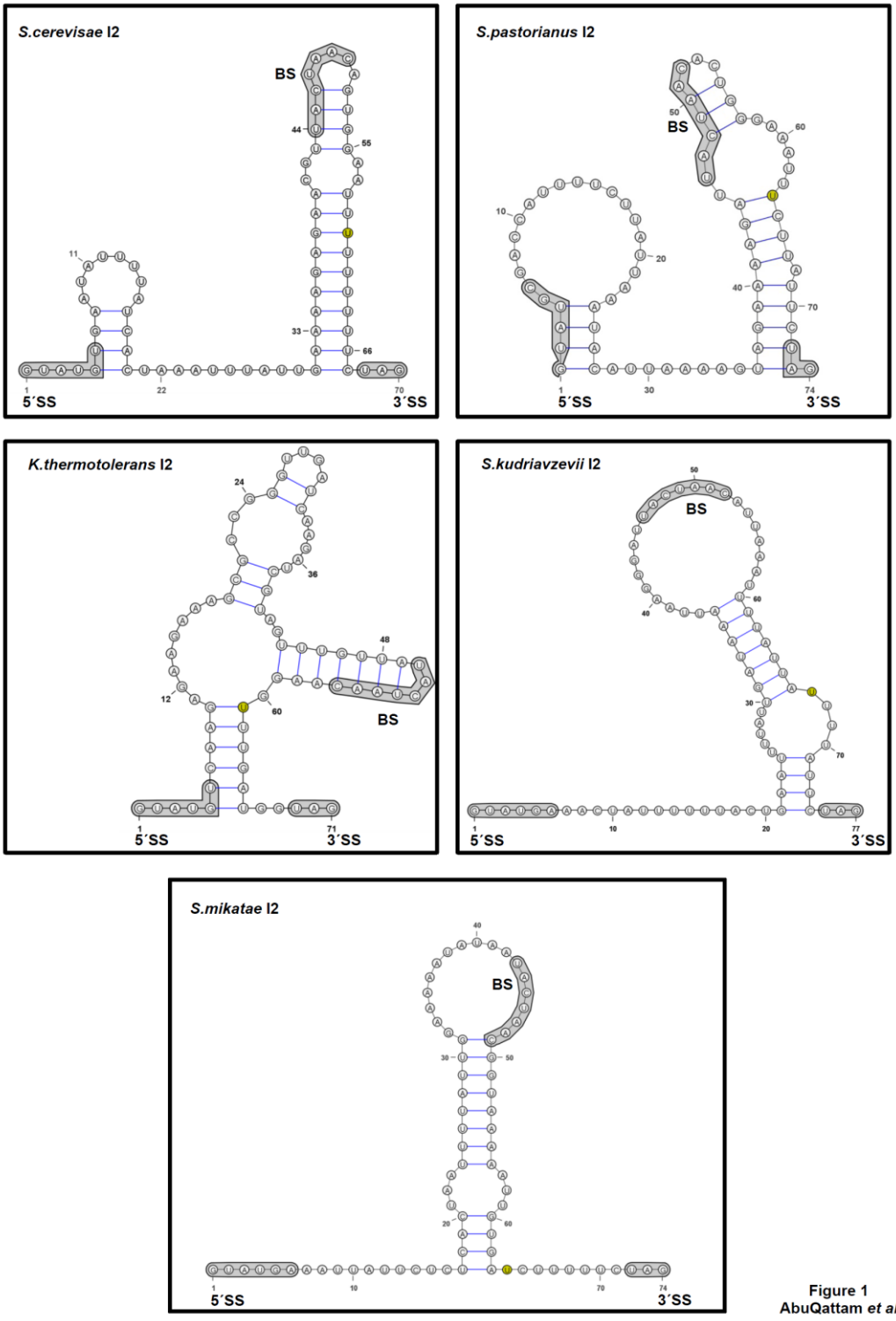


FIGURE 1. Predicted secondary structure of *SUS1* intron 2 RNA. Minimum free-energy I2 structures in *S. cerevisae, S. pastorianus, K. thermotolerans, S. kudriavzevii and S. mikatae*. The 5´SS, 3´SS, BS sequences and the uridine at -9 position preceding the AG are highlighted. The predictions were carried out with Mfold.

To confirm the computational predictions, we next examined by NMR spectroscopy and gel electrophoresis the conformation of *S. cerevisiae* I2, represented by a 37-nt oligomer (hereafter named I2s) encompassing the hairpin predicted to be formed by I2 nt 31-67 (Fig. 2). The electrophoretic results indicated the formation of a predominantly monomeric structure in different solution conditions. The NMR data revealed that I2s adopted the hairpin structure depicted in Figure 2. The presence of the $A_{32}AAAGAGAA_{40}$:$U_{58}UUUUUUUU_{66}$ double-helical tract at the base of the hairpin was demonstrated by NOE interactions involving exchangeable and non-exchangeable resonances. The two G36:U62 and G38:U60 wobble pairs were clearly established by the characteristic chemical shift and NOE pattern of the G H1 and U H3 iminos (Fig. 2A and Fig. S1A, crosspeaks a and b). Likewise, diagnostic A H2-U H3 interactions allowed identification of eight Watson-Crick A:U pairs in the I2s hairpin (Fig. 2A and Fig. S1B). A35:U63, A37:U61 and A39:U59 were assigned from NOE contacts involving protons of G36:U62 and G38:U60 (Fig. S1A, crosspeaks c-g, i and j, and Fig. S1C, crosspeaks k, l, n and q). The remaining four A:U pairs of the lower stem were identified by cross-strand A H2-U H3 NOEs (indicated by solid horizontal arrows in Fig. S1B), as well as by sequential intra-strand and cross-strand H2-H1' NOEs typical of canonical RNA stems (Fig. S1C).

The NMR data were also consistent with the presence of the two Watson-Crick pairs forming the $A_{45}C_{46}$:$G_{52}U_{53}$ upper stem below the apical loop. The A45:U53 pair was suggested by the remaining A H2-U H3 crosspeak, whose resonances were not connected by NOE interactions with the remaining A:U pairs of the system (Fig. S1B and S1C). Diagnostic crosspeaks between guanine imino and cytosine amino protons that were unrelated to the terminal C67 residue allowed identification of the C46:G52 pair (Fig. S1A, crosspeaks h). There were additional imino resonances in the I2s spectrum (Fig. 2A) whose chemical shifts were consistent with the formation of G:U pairs in the CGUU:GGAA internal loop separating the two canonical stems, but they exchanged too fast with the solvent to be assigned.

The stability of the hairpin formed by intron 2 of *SUS1* was evaluated with UV-thermal denaturation experiments. The results indicated that this stem-loop structure had low thermal

stability: the I2s hairpin melted at 26 °C in the absence of NaCl and at 36 °C with 100 mM NaCl, in ionic conditions approximately similar to those found in a cellular environment. The thermal denaturation curves were reversible in all cases (Fig. 2C and Fig. S2).

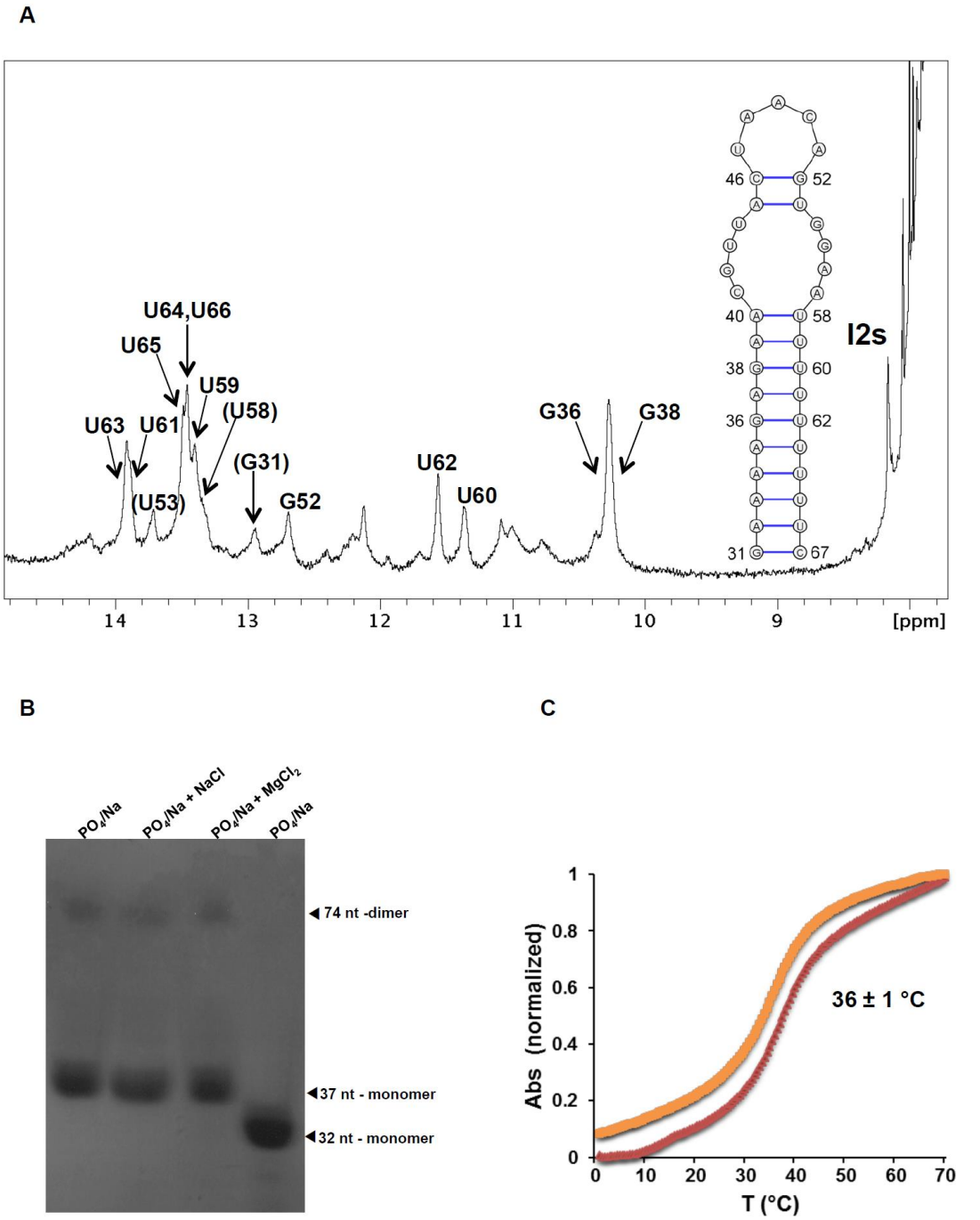


Figure 2
AbuQattam *et al*

FIGURE 2. *In vitro* analysis of intron 2 structure and stability. (A) Imino proton NMR spectrum and NMR-supported secondary structure of the 37-nt I2s oligomer. Conditions: 7 ^{o}C, 2 mM sodium phosphate (pH 6.0) and 100 mM NaCl. The assignments (parentheses indicate tentative assignments) were obtained from NOESY and TOCSY spectra in H_2O and D_2O. (B) Native gel comparing the electrophoretic mobility of 16 μM I2s samples previously annealed in different ionic conditions: (1) 2 mM sodium phosphate (pH 6.0), (2) 2 mM sodium phosphate and 100 mM NaCl, (3) 2 mM sodium phosphate and 2 mM $MgCl_2$. Lane 4 contained a 32-nt RNA hairpin control. (C) UV-monitored thermal denaturation curve of I2s in 2 mM sodium phosphate (pH 6.0) and 100 mM NaCl. The average melting temperature of I2s under these ionic conditions is indicated in the graph.

Partial deletion of the stem-loop structure formed by intron 2 modulates S. cerevisiae SUS1 splicing

The presence of a hairpin structure formed by I2 of *SUS1* in *S. cerevisiae* together with its conservation in several yeast species suggested a possible functional relevance. To address this question we studied an I2 deletion mutant (I2-mut1) in which the double-helical stem of the predicted hairpin was shortened by nine base pairs while the BS was maintained in the apical loop and the 3'SS after the stem (Fig. 3A). This mutant had 18 nt less relative to wild-type I2, but all intronic features relevant for splicing were well within functional range: the total size of the mutant intron was 52 nt (the minimum size has been proposed to be 50 nt), the distance between the 5'SS and the BS was 28 nt (greater than the required 25 nt minimum), and there were 12 nt between the BS and the 3'SS (10 nt are necessary in this case as a minimum) (Meyer et al. 2011; Pérez-Valle and Vilardell 2012). To evaluate the expression efficiency of this mutant *in vivo*, we used the *CUP1* reporter system (Lesser and Guthrie 1993). *cup1*Δ cells were transformed with one of the following reporter constructs: pACT1-CUP1, pSUS1g-CUP1 (containing a *SUS1* gene with wild-type I1 and I2 introns) and pSUS1-I2-mut1-CUP1 (with a

wild-type I1 and a shortened I2 hairpin). Splicing was monitored both by assessing copper tolerance and by semi-quantitative reverse transcription PCR (semi-qRT-PCR). Partial deletion of the stem of the I2 hairpin led to a clear reduction in copper tolerance compared to wild-type I2 (Fig. 3B), suggesting that the I2 hairpin was important for *SUS1* expression. When the different *SUS1* splice forms generated from expressing I2-mut1 in *sus1* cells were measured by semi-qRT-PCR, the accumulation of different pre-mRNAs containing I2, I1 or both introns was evident (Fig. 3C). In agreement with the results of the copper assay, semi-qRT-PCR experiments indicated a reduced amount of fully spliced mRNA relative to the wild-type cultures (Fig. 3D). Since I1-containing isoforms also accumulated with the I2-mut1 mutant (Fig. 3C), we generated an I2-mut1 construct lacking I1 (I2-mut1DI1). The absence of I1 did not change the effect of I2-mut1 on *SUS*1 processing as evidenced by cell growth in copper-containing plates (Fig. S3). This demonstrated that the splicing effects caused by the I2 mutation did not depend on I1.

A fast-migrating *SUS1* transcript was detected in the semi-qRT-PCR gels of I2-mut1 (Fig. 3C, last lane). Sequencing of this band revealed that it corresponded to the *SUS1* E1-E3 transcript, which was already observed (Hossain et al. 2011). Partial deletion of the double-helical stem of the I2 hairpin promoted accumulation of I1- and I2-containing pre-mRNAs (Fig. 3C). The poor splicing efficiency of I2-mut1 could have led to increased exon 2 skipping, which contributes to alternative splicing and the production of the E1-E3 isoform (discussed below).

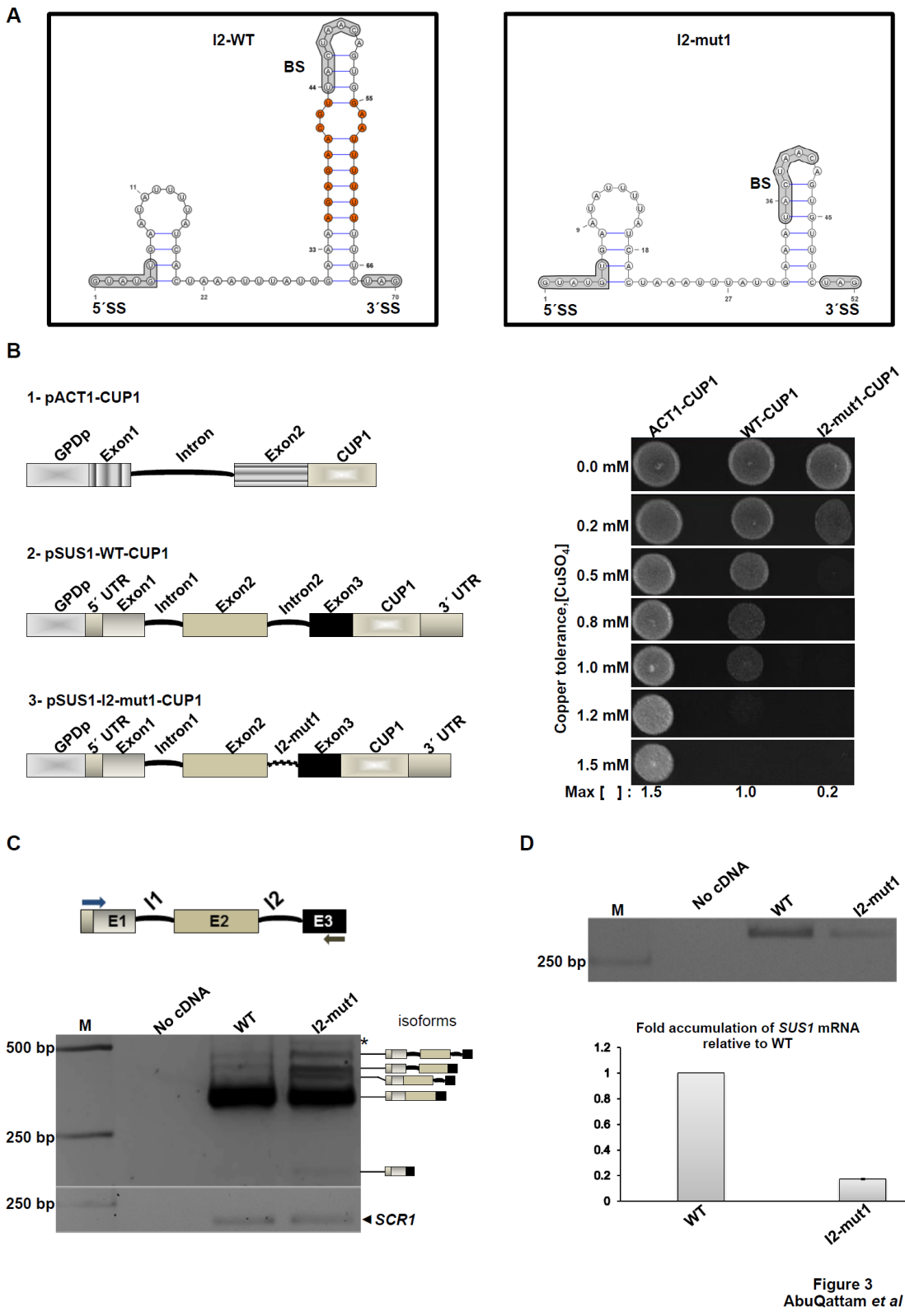


Figure 3
AbuQattam *et al*

FIGURE 3. Partial deletion of the I2 hairpin double-helical stem affects *SUS1* splicing. (A) MFold-predicted secondary structure of wild-type I2. The nucleotides highlighted in orange were removed to construct the intron 2 deletion mutant 1 (I2-mut1). The prediction of I2-mut1 secondary structure is shown on the right. (B) Copper assay of *cup1*Δ cells transformed with plasmids containing SUS1g-CUP1, SUS1-I2-mut1-CUP1 and ACT1-CUP1 as a control. Maximum copper tolerance is indicated. (C) Semi q-RT-PCR (35 cycles) to amplify *SUS1* transcripts resolved in a 3% agarose gel from cells expressing SUS1g-CUP1 (WT) or SUS1-I2-mut1-CUP1 (I2-mut1) in each case. A PCR without cDNA (no cDNA) was included as control. The stick diagrams on the right of the gel show the different isoforms of *SUS1* RNA. (D) Semi q-RT-PCR (25 cycles) to amplify *SUS1* transcripts resolved in a 3% agarose gel from cells expressing SUS1g-CUP1 (WT) or SUS1-I2-mut1-CUP1 (I2-mut1) in each case (upper panel). qPCR showing mRNA accumulation from cells expressing SUS1g-CUP1 (WT) or SUS1-I2-mut1-CUP1 (I2-mut1) normalized to the amount of *SCR1*ncRNA (lower panel). Error bars represent SE for at least three independent experiments (*Spurious PCR product).

Stabilization of the intron 2 hairpin affects SUS1 expression

The functional relevance of the I2 hairpin was further studied by analysing three new structural mutants. Mutant I2-mut2 was designed to disrupt the 10 base-pair stem at the base of the I2 hairpin by replacing the 10-nt $G_{31}AAAAGAGAA_{40}$ purine stretch with $C_{31}UUUUUUUUU_{40}$ (Fig. 4A, left panel). In contrast, I2-mut3 was intended to stabilize this 10 base-pair stem by converting its seven A:U pairs into G:C pairs: thus the As of the $G_{31}AAAAGAGAA_{40}$ purine stretch were replaced with Gs, and the opposing Us in the $U_{58}UUUUUUUUC_{67}$ pyrimidine segment (U_{58}-U_{59}, U_{61}, and U_{63}-U_{66}) with Cs (Fig. 4A, middle panel). On the other hand, I2-mut4 was aimed to disrupt the base pairs affecting BS nt without altering the lower double-helical stem of the hairpin. To achieve this, the $G_{52}UGG_{55}$ segment was replaced with $C_{52}CUU_{55}$ (Fig. 4A, right panel). Note that neither of these I2-mut2, I2-mut3 or I2-mut4 mutants affected the size of the intron, the number of nt separating the BS and the 3'SS, or the PPT following the BS. The effect of these mutants on the structure and stability of the I2 hairpin was assessed with Mfold calculations (Figs. 4A and S4A) as well as with gel electrophoresis and

thermal denaturation analyses of mutant 37-nt I2s oligonucleotides (Figs. S4B and S4C). In the predicted structure of I2-mut3, G:C pairs replaced the seven A:U pairs of the stem at the base of the I2 hairpin. In agreement with this, the I2-mut3s oligomer formed a highly stable monomeric structure with a T_m value of 86 ^oC (Fig. S4B), more than two times higher than the T_m of wild-type I2s, which melted at 36 ^oC in the same ionic conditions (Fig. 2). In the predicted structure of the full-length I2-mut2 intron the 10 base-pair lower stem was completely eliminated and the BS lied in a fully unstructured region (Figs. 4A and S4A). Although the PPT nt formed a small hairpin whose terminal pairs partially encompassed the 3'SS, these terminal U:G and A:U pairs showed significant positional entropy in the ensemble of theoretical structures (data not shown) and are likely unstable. In agreement with these predictions, the I2-mut2s oligonucleotide gave rise to delayed electrophoretic bands and did not exhibit any detectable transition in UV-monitored thermal denaturation experiments (Fig. S4), indicating the formation of a very unstable structure or no structure at all. In the I2-mut4 predicted fold, the BS nt did not form any base pairs in the apical region of the hairpin while the AU-rich lower stem remained unaffected (Figs. 4A and S4A). In accordance with this, I2-mut4s formed a monomeric structure with a slightly reduced T_m value relative to wild-type I2s (32 ^oC *vs.* 36 ^oC in the presence of 100 mM NaCl) (Fig. S4).

We then evaluated the impact of each of these structural mutants on *SUS1* expression as described above for I2-mut1. The I2-mut2, I2-mut3 and I2-mut4 mutants were first cloned in the *CUP1* reporter system and transformed into *cup1*Δ cells. As shown in Fig. 4B left panel, I2-mut3 drastically impaired cell growth relative to wild-type, indicating that stabilization of the I2 stem-loop decreased *SUS1* expression. In contrast, I2-mut2 and I2-mut4 did not impair cell growth when compared to wild-type (Fig. 4B left and right panel respectively). To assess the production of *SUS1* isoforms by these mutants, semi-q-RT-PCR was carried out to amplify *SUS1* transcripts in *sus1*Δ cells bearing mutant constructs. In agreement with the copper assay observations, the I2-mut3 mutant stabilizing the stem-loop structure completely inhibited I2 splicing, leading to a clear enrichment of I2- and I1-I2-containing pre-mRNA *SUS1* transcripts

(Fig. 4C). The I2-mut2 and I2-mut4 mutant sequences did not lead to a detectable increment of I1 or I2 retention in the gels (Fig. 4C), but semi-q-RT-PCR analyses of I2-mut2 transcripts showed a slight reduction of fully-spliced *SUS1* mRNA relative to wild-type (Fig. 4D), while a small increase was observed in the case of I2-mut4 (Fig. 4D).

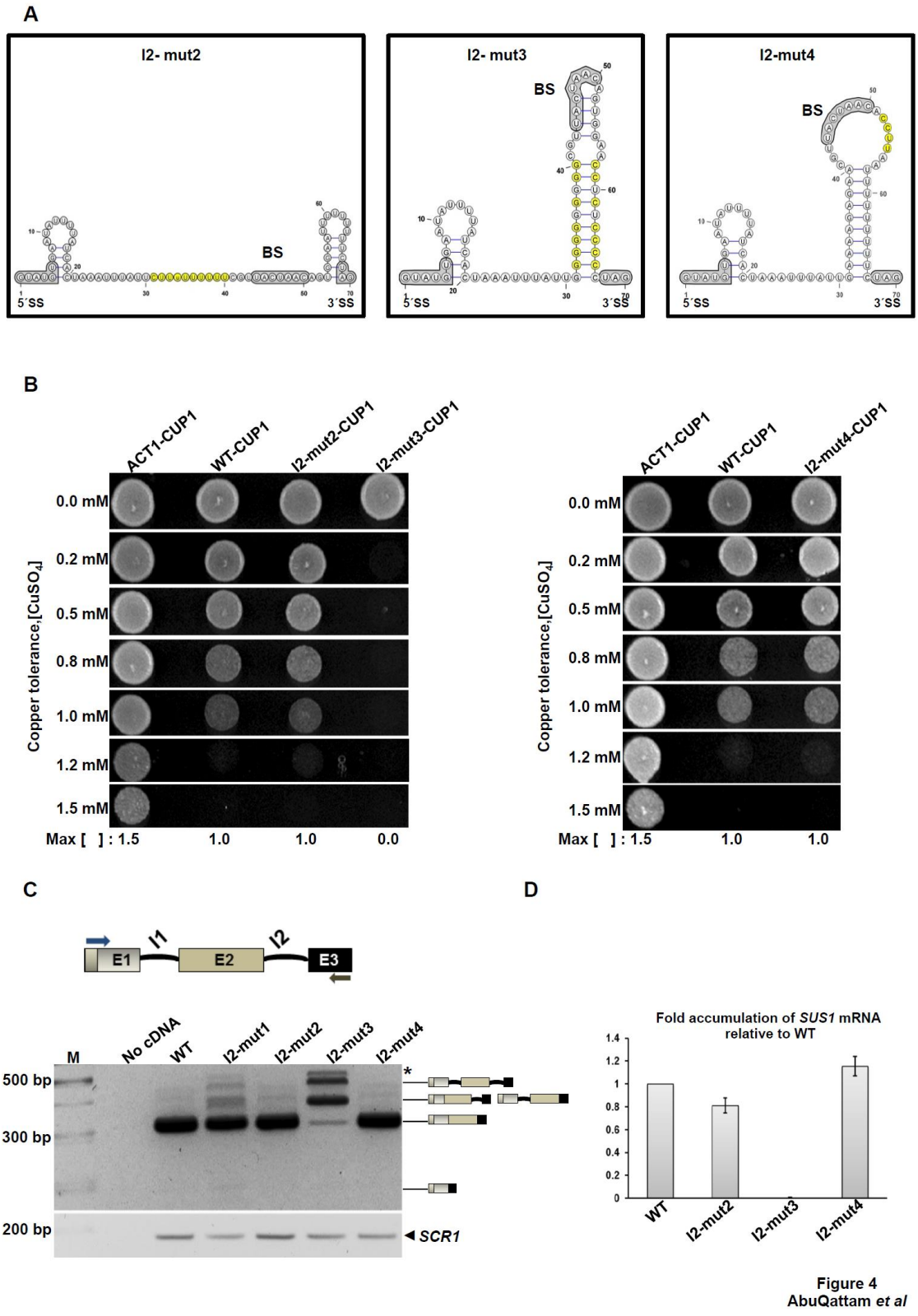


FIGURE 4. Stabilisation of the intron 2 hairpin affects *SUS1* expression. (A) MFold-predicted secondary structure of the I2 hairpin disruption mutant (I2-mut2, left panel), hairpin stabilization mutant (I2-mut3, middle panel) and branch site base paring disruption mutant (I2-mut4, right panel). The nt

highlighted in yellow were substituted to generate each mutant. (B) Copper assay of *cup1*Δ cells transformed with plasmids containing SUS1g-CUP1, SUS1-I2-mut2-CUP1, SUS1-I2-mut3-CUP1, SUS1-I2-mut4-CUP1 and ACT1-CUP1 as a control. Maximum copper tolerance is indicated. (C) Semi q-RT-PCR (35 cycles) to amplify *SUS1* transcripts resolved in a 3% agarose gel from cells expressing SUS1g-CUP1 (WT) or SUS1-I2-mut1-CUP1 (I2-mut1), SUS1-I2-mut2-CUP1 (I2-mut2), SUS1-I2-mut3-CUP1 (I2-mut3) and SUS1-I2-mut4-CUP1 (I2-mut4) in each case. A PCR without cDNA (no cDNA) was included as control. The stick diagrams on the right of the gel show the different isoforms of *SUS1* RNA. (D) qPCR showing mRNA accumulation from cells expressing *SUS1* as in (C) normalized to the amount of *SCR1* ncRNA. Error bars represent SE for at least three independent experiments (*Spurious PCR product).

High-temperature effects on SUS1 splicing upon intron 2 structural mutations

Interesting differences were observed between the structural mutants when the experiments were carried out under stress. Previous analyses with wild-type cells incubated at 42 ºC revealed reduction of *SUS1* expression, accompanied by accumulation of unspliced *SUS1* transcripts and decrease of fully-spliced mRNA (Cuenca-Bono et al. 2011). We tested the effect of the different structural mutants in these conditions by semi-q-RT-PCR (Fig. 5A). The most drastic variation relative to standard conditions (cells growing at 30 ºC) was observed for mutants I2-mut1 and I2-mut3, which compared with the wild-type sequence generated a large accumulation of transcripts containing both introns, together with a drastic reduction of fully spliced mRNA. At 42 ^{o}C, I2-mut2 and I2-mut4 exhibited less intron retention as well as increased levels of fully spliced mRNA relative to wild-type. In addition, I2-mut2 clearly showed less I1 retention relative to wild-type or I2-mut4 (Fig. 5A). This confirmed that I2 splicing can affect an upstream splicing step, supporting the results obtained with mutant I2-mut1.

Sus1 functions in histone deubiquitination and mRNA export are affected in I2 hairpin mutants that inhibit splicing

Sus1 is a conserved factor that is required for mRNA export and histone H2B deubiquitination, and deletion or mutation of *SUS1* has been shown to impair mRNA export and histone H2B deubiquitination (Rodríguez-Navarro et al. 2004; Köhler et al. 2006; Kurshakova et al. 2007; García-Oliver et al. 2011; Kopytova et al. 2010; Galan and Rodríguez-Navarro 2012; Klöckner et al. 2009; Zhao et al. 2008). So far, our analyses have shown that partial deletion (I2-mut1) and stabilization (I2-mut3) of the 37-nt hairpin structure formed by intron 2 led to differential *SUS1* isoform expression. These changes in splicing could lead to a deregulation of Sus1 protein production. To study whether these I2 mutations had functional consequences, cells lacking the *SUS1* gene (*sus1*Δ) were transformed with wild-type and I2-mutant *SUS1* plasmids and their ability to complement *SUS1* deletion phenotypes was evaluated. As shown in Fig. 5B, I2-mut3 was unable to deubiquitinate histone H2B at wild-type levels, in agreement with the low expression of fully spliced *SUS1* isoform in these cells (Fig. 4C). Although I2-mut1 likewise reproducibly reduced the amount of fully spliced *SUS1* (Fig. 3C and 3D), the reduction of ubiquitinated H2B relative to WT was not statistically significant (Fig. 5B). Similar results were obtained using total H2B as a loading control (data not shown). Concomitantly, defects in mRNA export provoked by *SUS1* deletion observed in 90% of the cells (Fig 5C, left panel) were more apparent for I2-mut3 than I2-mut1: 50 and 16% of the cells transformed with these two mutants, respectively, accumulated mRNA, relative to 8% transformed with wild-type *SUS1* (Fig. 5C).

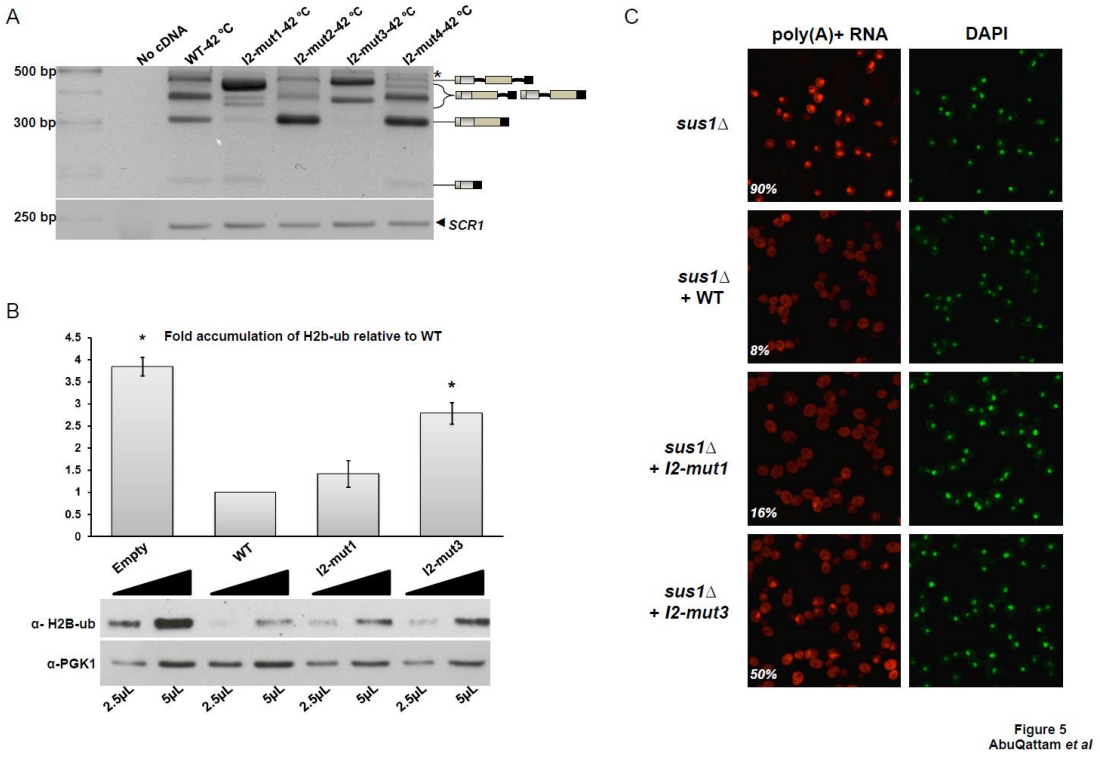


Figure 5. Deletion or stabilization of intron 2 hairpin affects Sus1 cellular roles. (A) Analysis of *SUS1* transcripts by semi-q-RT–PCR from cells expressing SUS1g (WT) or I2 structural mutants (SUS1-I2-mut1-CUP1 (I2-mut1), SUS1-I2-mut2-CUP1 (I2-mut2), SUS1-I2-mut3-CUP1 (I2-mut3) and SUS1-I2-mut4-CUP1 (I2-mut4) from cells incubated at 42 ºC for 20 min. Bands corresponding to the different transcripts are indicated. A PCR without cDNA (no cDNA) was included as control. (B) Quantification and representative blots obtained with increasing concentration of whole cell extracts (WCE) from *sus1*Δ cells transformed with an empty plasmid (Empty), or with plasmids expressing SUS1g (WT), SUS1-I2-mut1 (I2-mut1) or SUS1-I2-mut3 (I2-mut3). Two volumes (2.5 µL or 5 µL) of each WCE were resolved by 15% SDS-polyacrylamide gel, transferred to a nitrocellulose membrane. The membrane was cut and the upper part was incubated with *PGK1* antibody (loading control) and the lower with H2B-ub antibody. Bar graph shows total H2B-ub levels normalized to *PGK1* after quantification of Western-blot signals. Similar results were obtained using total H2B as a loading control (data not shown). Error bars represent SE for at least three independent experiments. (C) Representative images of poly(A)+ RNA localization in *sus1*Δ cells, transformed as in (B) and assayed by *in situ* hybridization using Cy3-labeled oligo(dT) probes. The percentages indicate the number of cells showing mRNA accumulation, and correspond to the averages of at least three independent biological replicates. DAPI staining to mark the nucleus is shown on the right. Cells were incubated at 37 °C in rich media for 2 h.

Discussion

The structure of RNA molecules plays an important role during their metabolism, adding a new layer of regulation to different steps of the gene expression pathway. In this study, we have analysed the impact on splicing of a hairpin structure formed by the second intron of the pre-mRNA of *S. cerevisiae SUS1*. Secondary structure predictions of this intron in this and other yeast species revealed the formation of a stem-loop structure with BS nt located near the apical loop and 3'SS nt close to the 3' terminus of the base-paired stem (Fig. 1). In several cases, 5'SS nt were likewise predicted to form base pairs with downstream nt that might affect splicing efficiency. However, our study focused on the hairpin formed by the BS and 3'SS segment, leaving the 5'SS region unmodified. In all species in which the I2 hairpin was present a pyrimidine-rich sequence was located between the BS and the 3'SS. This polypyrimidine tract (PPT) has been shown to be important for splicing in higher eukaryotes (Coolidge et al. 1997; Liu and Mertz 1996; Roscigno et al. 1993). In contrast, in yeast species the 5'SS, BS and 3'SS sequences are highly conserved, but the PPTs aren't (Rymond and Rosbash 1985; Fouser and Friesen 1987). Interestingly, the PPT of *SUS1* I2 seemed to be partially conserved in yeast, and a uridine at position -8 preceding the 3'SS was present in all cases (Fig. 1). It has been suggested that splicing in fungi may be different from that in vertebrates and may require additional proteins that interact with PPTs upstream of the BS (Kupfer et al. 2004)**.** However, other studies have indicated that PPTs downstream of the BS do provide a positioning function (Matlin et al. 2007). Furthermore, non-conserved sequences can also influence alternative splice-site selection via structural motifs, supporting the idea that pre-mRNA introns can be highly evolved molecules with functional and structural constraints (Deshler and Rossi 1991).

To confirm the theoretical predictions, we studied the structure and stability of *S. cerevisiae* I2, represented by a 37-nt I2s oligonucleotide encompassing the predicted hairpin. The NMR data indicated that I2s formed a hairpin structure in solution containing a 32-nt double-helical stem and a 5-nt UAACA apical loop containing the last four nt of the BS (Fig. 2). The stem comprised an AU-rich tract of 10 base pairs interrupted by a CGUU:GGAA internal loop that

was separated from the apical loop by two Watson-Crick pairs. The UV-thermal denaturation experiments showed that this intron 2 hairpin had surprisingly low thermal stability. It is well established that both the BS and the PPT of the introns need to be single-stranded to allow binding of the BBP (SF1 in mammals) and Mud2 ($U2AF_{65}$) proteins in the first steps of spliceosome assembly (Pérez-Valle and Vilardell 2012; Gahura et al. 2011). It is therefore remarkable that two of the BS nt (A_{45} and C_{46}) and the complete PPT of I2 are contained in the double-stranded stem of the I2 hairpin. This observation, together with the UV thermal denaturation and NMR spectroscopy findings showing that the structure was thermally labile, supported a regulatory role for the I2 hairpin on splicing.

To assess how important I2 hairpin formation was for *SUS1* function and regulation, four structural mutants were analysed: I2-mut1 (containing an I2 hairpin with a shortened double-helical stem), I2-mut2 (with a destabilized stem), I2-mut3 (stabilized stem), and I2-mut4 (with an intact stem and an opened apical BS loop) (Figs. 3 and 4). The phenotypes observed when partially deleting the stem of the I2 hairpin (I2-mut1) suggested that this stem-loop structure was important for *SUS1* expression. However, since part of the PPT was deleted we cannot exclude the possibility that this could have affected *SUS1* splicing. Notably, I1 retention increased with the I2-mut1 mutation (Fig. 3C). Previous studies proposed that the upstream intron of *SUS1* modulated the processing of I2 (Hossain et al. 2011). Our results demonstrate that intrinsic features of I2 can also lead to I1 retention, suggesting that the downstream intron affected splicing of its upstream counterpart.

The most drastic effects were observed with the I2-mut3 mutant. The corresponding I2-mut3s oligomer formed a highly stable structure with a melting temperature more than two times higher than that of the wild-type I2s hairpin. This mutation completely inhibited I2 splicing, likely due to the inability of the spliceosome components to act in this environment. The mutants that completely destabilized the double-helical stem of the I2 hairpin (I2-mut2) or further opened the apical BS loop (I2-mut4) did not provoke a detectable increment of I1 or I2 retention, but behaved differently in terms of the amount of fully-spliced *SUS1* mRNA

generated. The I2-mut4 transcripts showed a slight increase of fully-spliced *SUS1* mRNA relative to wild-type, while a small reduction was observed in the case of I2-mut2. The small positive effect of I2-mut4 may be attributed to the marginally greater accessibility of BS nt in this sequence. This would be consistent with the NMR analyses of the wild-type hairpin, which indicated the presence of only two stable base pairs in this region (Fig. 2), and with the slight reduction in melting temperature (4 ^{o}C) observed for I2-mut4s relative to the wild-type oligomer (Figs. 2 and S4C). Like I2-mut1, I2-mut2 was detrimental for splicing probably because the presence of the lower double-helical stem of the I2 hairpin increases the accessibility of BS and 3'SS nt to splicing factors. However, the slight reduction of fully-spliced mRNA observed for I2-mut2, where the double-helical stem of the hairpin was disrupted, was much smaller than that observed for I2-mut1, containing a shortened stem. It is worth noting in this regard that in the I2-mut2 sequence an artificial PPT upstream from the BS was generated (Fig. 4A), which could promote splicing (Kupfer et al. 2004). Thus, positioning the BS in between two strong PPTs may have compensated the splicing defects caused by the destabilization of the I2 hairpin, explaining the smaller effect observed for I2-mut2 relative to I2-mut1.

The functional relevance of the I2 hairpin was further studied by analysing the ability of the I2-mut1 and I2-mut3 mutants to accomplish Sus1 protein functions. The effect of these mutations on deubiquitination and mRNA export (Fig. 5) suggest that the reduced amount of Sus1 protein likely produced by these mutants is still enough to partially complement Sus1 functions, including *sus1*Δ temperature sensitive growth defect (data not shown). This is interesting, since overexpression of Sus1 has been shown to largely impair its functionality (Hossain et al. 2011; Cuenca-Bono et al. 2011).

In conclusion, we have demonstrated the existence of a weakly-stable, 37-nt hairpin structure formed by the second intron of *S. cerevisiae SUS1* that contains BS nt in its apical loop and 3'SS nt immediately after the 3' terminus of the stem. Our study shows that this structure regulates the expression of *SUS1* isoforms by influencing *SUS1* splicing. The effect of different mutants modulating the structure and stability of the hairpin suggests that this I2 stem-loop

likely increases the accessibility of BS and 3'SS nt to the splicing components by maintaining the BS in a readily accessible and easily opened apical loop environment and the 3'SS nt unstructured immediately after the double-helical stem. In this way, the efficiency of I2 removal would be regulated by the formation of the structure.

Moreover, the effect of mutants I2-mut1 and I2-mut2 on I1 retention indicate that the I2 hairpin also modulates the splicing of the upstream intron. In this regard, it may be possible that the weakly-stable hairpin has evolved as a mechanism to coordinate removal of I1, whose splicing is very inefficient and needs more time to occur. On the other hand, exon 2 skipping was promoted by mutant I2-mut1 under standard conditions. *SUS1* exon 2 skipping was previously detected in wild-type cells when different components of the splicing machinery (*MSL1*, *LEA1*, *MUD2*) were deleted, and was explained by the role of these factors in recognizing the suboptimal BS of I1 (Hossain et al. 2009; Will and Lührmann 2011). Interestingly, our results indicate that the presence of these factors is not sufficient to suppress exon skipping when intron 2 is mutated. This opens the possibility that I2-dependent mechanisms also contribute to regulate exon skipping.

Materials and Methods

Sequences and structure prediction of *SUS1*. The sequences of the second intron of *SUS1* across different species of yeast were obtained from the Yeast Genome Database (SGD) and NCBI (http://www.ncbi.nlm.nih.gov/). The secondary structure predictions were carried out with the MFOLD web server (http://mfold.rna.albany.edu/?q=mfold/rna-folding-form) (Zuker 2003), and the structures were drawn using VARNA (http://varna.lri.fr/) (Darty et al. 2009).

Large-scale RNA transcription. The wild-type I2s and mutant I2-mut2s, I2-mut3s and I2-mut4s RNA oligomers used for NMR spectroscopy, gel electrophoresis and UV thermal experiments were prepared by T7-polymerase *in vitro* transcription using synthetic oligonucleotide DNA templates. The RNA transcripts were purified on denaturing gels

containing 20% acrylamide and 8 M urea. After electroelution from the gel, the RNAs were ethanol-precipitated two times and desalted with Sephadex G-25 cartridges. The I2s NMR samples were microdialyzed in aqueous solutions containing 2 mM sodium phosphate (pH 6.0) and 0.1 mM EDTA with no added salts or additionally containing 100 mM NaCl. The RNA concentration in these samples ranged between 0.10 and 0.15 mM.

NMR spectroscopy. NMR spectra were acquired on 600 MHz (cryoprobe-equipped) and 500 MHz Bruker Avance III spectrometers, and analysed using Topspin 1.3 (Bruker Biospin) and Sparky 3.110 (T. D. Goddard, D. G. Kneller, UCSF USA, 2004). The I2s systems were studied using two-dimensional waterwate-NOESY (with 150 to 400 ms mixing times) and watergate-TOCSY experiments (60 ms mixing time) recorded in 90% H_2O/10% D_2O at several temperatures (between 7 °C and 17 °C). TOCSY and NOESY (250 ms) experiments were also acquired in D_2O at 13 °C and 23 °C. The recycle delay was 2 s in all cases.

UV thermal denaturation. The thermal stability of wild-type I2s and mutant I2-mut2s, I2-mut3s and I2-mut4s oligomers 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 approximately 0 °C to 100 °C at a gradient of 0.5-2.0 °C min^{-1} and subsequently decreased at the same rate to evaluate the reversibility of the process. The experiments were carried out using 0.4-0.5 ODU/ml of RNA (1.0-1.3 μM). The thermal denaturing profiles of the wild-type I2s sequence were examined in the following ionic conditions: 2 mM sodium phosphate (pH 6.0) with no additional salts, or containing either 100 mM NaCl or 2 mM $MgCl_2$. The mutant oligomers were studied in 2 mM sodium phosphate (pH 6.0) and 100 mM NaCl. 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.

Gel electrophoresis. These experiments were used to assess the strand stoichiometry of the structures formed by the I2 oligonucleotides, as well as to evaluate the number of structural

species formed by each sequence in a given condition. Native gels were run at 4 °C for approximately 14 hours under constant voltage (90 V). We used 20% 19:1 acrylamide:bisacrylamide gels and 89 mM Tris-Borate (TB) as running buffer. These experiments involved 13-18 µM wild-type or mutant I2s samples, previously annealed as specified above in the following ionic conditions: 2 mM sodium phosphate (pH 6.0) with no additional salts, or containing either 100 mM NaCl or 2 mM $MgCl_2$. All gels were stained with methylene blue and distained with water.

Generation of *SUS1* constructs. All *SUS1* gene constructs contained the last twenty nt of 5´UTR. The I2 mutants (mut1, mut2, mut3, mut4 and mut1-I1Δ,) were constructed by the fusion PCR method (Yon and Fried 1989; Lesser and Guthrie 1993), 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 Fig. S5). Taq DNA polymerase (Roche®) was used to amplify the WT, I2-mut1, I2-mut2, I2mut3 and I2-mut4 constructs of *SUS1*. Due to the presence of a high GC content in I2-mut3, the Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs®) was used to obtain this construct.

Yeast cultures and microbiological techniques. Copper resistance assays were carried out by growing the transformed *cup1*Δ cells at 30 ºC 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 OD_{600}. Subsequently, 10-fold serial dilutions of an equal number of cells were made and drops spotted onto SC-leu plates containing different concentration of $CuSO_4$ (Lesser and Guthrie 1993). Plates were photographed after 3-5 days of incubation at 30 ºC. Yeast cell transformations were done by the LiAc/SS carrier DNA/PEG method (Gietz and Schiestl 2007).

RNA extraction, Reverse transcription PCRs and semi q-RT-PCRs. Total RNA was harvested from *sus1*Δ cells transformed with the CUP1 plasmids bearing SUS1g, SUS1-I2-mut1, SUS1-I2-mut2, SUS1-I2-mut3 or SUS1-I2-mut4 by the Hot/Acid-phenol method (Schmitt et al. 1990), and quantified using a Nanodrop. 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 OD_{600} 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 pairs of primers located upstream exon1 and exon3 was used to amplify the transcripts of *SUS1* (25 cycles to quantify the mRNA and 35 cycles to amplify low abundant 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 visualised 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 by the ImageJ program (http://rsbweb.nih.gov/ij/)

***In Situ* Hybridization (FISH)**

Fluorescent in situ hybridization (FISH) against poly(A)+RNA was done as described by Cuenca-Bono et al. 2011; yeast cells were grown in 100 ml of SC-Leu medium at 30 °C to an 0.3 OD600. Then, cultures were rapidly shifted to 37 °C incubator for 2 h. After hybridization, slides were mounted using VECTASHIELD® Mounting Medium with DAPI. Detection of Cy3-oligo(dT) was performed using a Leica DM600B fluorescence microscope. mRNA accumulation was represented as the percentage of cells showing a bright signal in the nucleus

that colocalized with DAPI staining (DNA). The percentages correspond to averages of at least 3 biological replicates.

Total Protein extraction for histone modification detection and western blot. For the western blot assays, the transformed *sus1*Δ cells at 30 ºC were grown on 50 ml synthetic selective medium lacking Leucine (SC-Leu) to 0.5 OD_{600}, then the cells were harvested and washed by 20% TCA, after nitrogen freezing the cells were thawed and resuspended in 0.5 ml of 20% TCA. Cells were disrupted by glass beads and the lysate was recovered. After centrifuging the pellet during 10 min (300 rpm) the pellet was resuspended in 200 µl of 1x LB plus 50 µl of 2 M unbuffered Tris and then boiled during 3 min at 95 ºC. The suspension was centrifuged during 5 min at 3000 rpm and the supernatant was recovered. The proteins were separated by 15% SDS-PAGE and electrotransferred to nitrocellulose membranes, as previously described (Thiriet and Albert 1995). Membranes were stained with Ponceau S, as a transferring control, and were incubated with specific antibodies: α-H2B ubiquitinated at position 123 (Cell Signaling) and α-PGK1 (Thermoscientific). Proteins were detected with horseradish peroxidase-conjugated anti-rabbit secondary antibodies and ECL Advanced reagents (GE Healthcare). Specific signals were quantified by the ImageJ program (http://rsbweb.nih.gov/ij/).

ACKNOWLEDGMENTS

This work has been supported by MINECO of Spain (BFU2011-23418 and BFU2014-57636-P to S.R.-N. and BFU-2012-30770 to J.G.), by Generalitat Valenciana of Spain (PROMETEO/2013/061, ACOMP/2014/061 and ACOMP/2015/096 to S.R-N., ACOMP/2014/056 to J.G., and Santiago Grisolía fellowship to A.AQ), and by Universidad Católica de Valencia. We are grateful to the undergraduate students Chiara Boccellato, Alejandro Aparicio and Joan Serrano from ERASMUS, ADEIT-UV and SIE-UPV programs for making contributions to this work.

REFERENCES

Acuña LIG, Kornblihtt AR. 2014. Long range chromatin organization: a new layer in splicing regulation? *transcription* **5**.

Coolidge CJ, Seely RJ, Patton JG. 1997. Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res* **25**: 888–896.

Cuenca-Bono B, García-Molinero V, Pascual-García P, Dopazo H, Llopis A, Vilardell J, Rodríguez-Navarro S. 2011. SUS1 introns are required for efficient mRNA nuclear export in yeast. *Nucleic Acids Res* **39**: 8599–8611.

Darty K, Denise A, Ponty Y. 2009. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**: 1974–1975.

Deshler JO, Rossi JJ. 1991. Unexpected point mutations activate cryptic 3' splice sites by perturbing a natural secondary structure within a yeast intron. *Genes Dev* **5**: 1252–1263.

Dong S, Jacobson A, He F. 2010. Degradation of YRA1 Pre-mRNA in the cytoplasm requires translational repression, multiple modular intronic elements, Edc3p, and Mex67p. *PLoS Biol* **8**: e1000360.

Dong S, Li C, Zenklusen D, Singer RH, Jacobson A, He F. 2007. YRA1 autoregulation requires nuclear export and cytoplasmic Edc3p-mediated degradation of its pre-mRNA. *Mol Cell* **25**: 559–573.

Fica SM, Tuttle N, Novak T, Li N-S, Lu J, Koodathingal P, Dai Q, Staley JP, Piccirilli JA. 2013. RNA catalyses nuclear pre-mRNA splicing. *Nature* 1–18.

Fouser LA, Friesen JD. 1987. Effects on mRNA splicing of mutations in the 3' region of the Saccharomyces cerevisiae actin intron. *Mol Cell Biol* **7**: 225–230.

Gahura O, Hammann C, Valentová A, Půta F, Folk P. 2011. Secondary structure is required for 3' splice site recognition in yeast. *Nucleic Acids Res* **39**: 9759–9767.

Galan A, Rodríguez-Navarro S. 2012. Sus1/ENY2: a multitasking protein in eukaryotic gene expression. *Critical Reviews in Biochemistry and Molecular Biology* **47**: 556–568.

García-Oliver E, García-Molinero V, Rodríguez-Navarro S. 2012. mRNA export and gene expression: The SAGA–TREX-2 connection. *BBA - Gene Regulatory Mechanisms* **1819**: 555–565.

Gietz RD, Schiestl RH. 2007. Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols* **2**: 35–37.

Gonzalez-Aguilera C, Tous C, Gomez-Gonzalez B, Huertas P, Luna R, Aguilera A. 2008. The THP1-SAC3-SUS1-CDC31 complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability. *Mol Biol Cell* **19**: 4310–4318.

Hossain MA, Claggett JM, Nguyen T, Johnson TL. 2009. The cap binding complex influences

H2B ubiquitination by facilitating splicing of the SUS1 pre-mRNA. *RNA (New York, NY)* **15**: 1515–1527.

Hossain MA, Rodriguez CM, Johnson TL. 2011. Key features of the two-intron Saccharomyces cerevisiae gene SUS1 contribute to its alternative splicing. *Nucleic Acids Res* **39**: 8612–8627.

Johnson TL, Vilardell J. 2012. Regulated pre-mRNA splicing: the ghostwriter of the eukaryotic genome. *Biochim Biophys Acta* **1819**: 538–545.

Klöckner C, Schneider M, Lutz S, Jani D, Kressler D, Stewart M, Hurt E, Köhler A. 2009. Mutational uncoupling of the role of Sus1 in nuclear pore complex targeting of an mRNA export complex and histone H2B deubiquitination. *J Biol Chem* **284**:.

Kopytova DV, Krasnov AN, Orlova AV, Gurskiy DY, Nabirochkina EN, Georgieva SG, Shidlovskii YV. 2010. ENY2: couple, triple...more? *Cell Cycle* **9**: 479–481.

Köhler A, Pascual-García P, Llopis A, Zapater M, Posas F, Hurt E, Rodríguez-Navarro S. 2006. The mRNA export factor Sus1 is involved in Spt/Ada/Gcn5 acetyltransferase-mediated H2B deubiquitinylation through its interaction with Ubp8 and Sgf11. *Mol Biol Cell* **17**: 4228–4236.

Kupfer DM, Drabenstot SD, Buchanan KL, Lai H, Zhu H, Dyer DW, Roe BA, Murphy JW. 2004. Introns and splicing elements of five diverse fungi. *Eukaryotic Cell* **3**: 1088–1100.

Kurshakova M, Maksimenko O, Golovnin A, Pulina M, Georgieva S, Georgiev P, Krasnov A. 2007. Evolutionarily conserved E(y)2/Sus1 protein is essential for the barrier activity of Su(Hw)-dependent insulators in Drosophila. *Mol Cell* **27**: 332–338.

Lesser CF, Guthrie C. 1993. Mutational analysis of pre-mRNA splicing in Saccharomyces cerevisiae using a sensitive new reporter gene, CUP1. *Genetics* **133**: 851–863.

Liu X, Mertz JE. 1996. Sequence of the polypyrimidine tract of the 3“-terminal 3” splicing signal can affect intron-dependent pre-mRNA processing in vivo. *Nucleic Acids Res* **24**: 1765–1773.

Matlin AJ, Southby J, Gooding C, Smith CWJ. 2007. Repression of alpha-actinin SM exon splicing by assisted binding of PTB to the polypyrimidine tract. *RNA* **13**: 1214–1223.

Meyer M, Plass M, Pérez-Valle J, Eyras E, Vilardell J. 2011. Deciphering 3'ss selection in the yeast genome reveals an RNA thermosensor that mediates alternative splicing. *Mol Cell* **43**: 1033–1039.

Moehle EA, Braberg H, Krogan NJ, Guthrie C. 2014. Adventures in time and space: Splicing efficiency and RNA polymerase II elongation rate. *RNA Biol* **11**: 313–319.

Pascual-García P, Govind CK, Queralt E, Cuenca-Bono B, Llopis A, Chavez S, Hinnebusch AG, Rodríguez-Navarro S. 2008. Sus1 is recruited to coding regions and functions during transcription elongation in association with SAGA and TREX2. *Genes Dev* **22**: 2811–2822.

Pérez-Valle J, Vilardell J. 2012. Intronic features that determine the selection of the 3′ splice site. *WIREs RNA* **3**: 707–717.

Preker P, Guthrie C. 2006. Autoregulation of the mRNA export factor Yra1p requires inefficient splicing of its pre-mRNA. *RNA* **12**: 994–1006.

Rodríguez-Navarro S, Fischer T, Luo M-J, Antúnez O, Brettschneider S, Lechner J, Pérez-Ortín JE, Reed R, Hurt E. 2004. Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery. *Cell* **116**: 75–86.

Rodríguez-Navarro S, Strässer K, Hurt E. 2002. An intron in the YRA1 gene is required to control Yra1 protein expression and mRNA export in yeast. *EMBO Rep* **3**: 438–442.

Rogic S, Montpetit B, Hoos HH, Mackworth AK, Ouellette BF, Hieter P. 2008. Correlation between the secondary structure of pre-mRNA introns and the efficiency of splicing in Saccharomyces cerevisiae. *BMC Genomics* **9**: 355.

Roscigno RF, Weiner M, Garcia-Blanco MA. 1993. A mutational analysis of the polypyrimidine tract of introns. Effects of sequence differences in pyrimidine tracts on splicing. *J Biol Chem* **268**: 11222–11229.

Rymond BC, Rosbash M. 1985. Cleavage of 5“ splice site and lariat formation are independent of 3” splice site in yeast mRNA splicing. *Nature* **317**: 735–737.

Schmitt ME, Brown TA, Trumpower BL. 1990. A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. *Nucleic Acids Res* **18**: 3091–3092.

Thiriet C, Albert P. 1995. Rapid and effective western blotting of histones from acid-urea-Triton and sodium dodecyl sulfate polyacrylamide gels: two different approaches depending on the subsequent qualitative or quantitative analysis. *Electrophoresis* **16**: 357–361.

Wahl MC, Will CL, Luhrmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**: 701–718.

Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, Ouyang Z, Zhang J, Spitale RC, Snyder MP, Segal E, et al. 2015. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* **505**: 706–709.

Warf MB, Berglund JA. 2010. Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem Sci* **35**: 169–178.

Will CL, Lührmann R. 2011. Spliceosome Structure and Function. *Cold Spring Harb Perspect Biol* **3**: a003707–a003707.

Yon J, Fried M. 1989. Precise gene fusion by PCR. *Nucleic Acids Res* **17**: 4895.

Zhao Y, Lang G, Ito S, Bonnet J, Metzger E, Sawatsubashi S, Suzuki E, le Guezennec X, Stunnenberg HG, Krasnov A, et al. 2008. A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. *Mol Cell* **29**: 92–101.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.

An intronic RNA structure modulates expression of the mRNA biogenesis factor Sus1

Ali AbuQattam, Susana Rodríguez-Navarro and José Gallego

SUPPLEMENTARY MATERIAL

Supplementary Figures

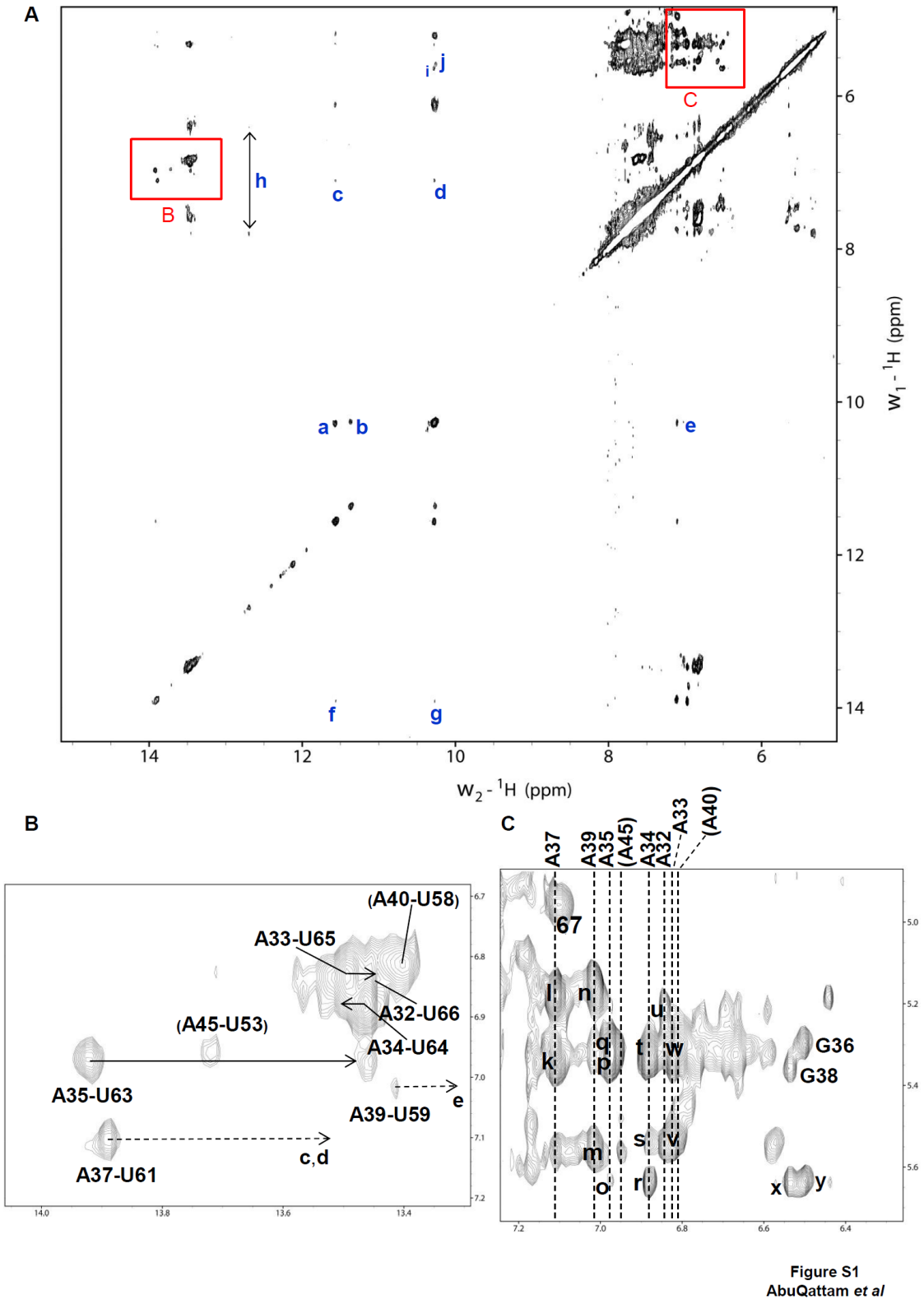


Figure S1
AbuQattam *et al*

Figure S1. **NMR spectroscopy analysis of *SUS1* I2s construct**. (A) Assignment of the imino region of the NOESY spectrum of I2s (H_2O/D_2O, 300 ms mixing time, 7 ^oC, 100 mM NaCl). Selected crosspeaks (a) to (h) are assigned as follows: a, G36 H1-U62 H3; b, G38 H1-U60 H3; c, A37 H2-U62 H3; d, A37

H2-G36 H1; e, G38 H1-A39 H2; f, U63 H3-U62 H3; g, U63 H3-G36 H1; h, C46 H4-G52 H1; i, A37 H1'-G36 H1; j, A39 H1'-G38 H1. (B) Inset of (A) showing NOE interactions between adenine H2 resonances and uracil imino H3 protons. H2-H3 crosspeaks indicating Watson-Crick A-U base pairs are labelled, sequential cross-strand interactions are indicated with solid horizontal arrows, and crosspeaks (c), (d) and (e) are identified in (A). (C) Inset of (A) illustrating the NOE interactions established by A-U pair H2 protons in the aromatic to H1'/H5 spectral region. A-U H2 resonances are indicated with dashed lines, intraresidue H1´-H6/H8 crosspeaks are labelled with residue name and number, and pyrimidine H5-H6 crosspeaks are labelled with residue number. Crosspeaks (k) to (y) are assigned as follows: k, G38 H1'-A37 H2; l, U62 H1'-A37 H2; m, A40 H1'-A39 H2; n, U60 H1'-A39 H2; o, A35 H1'-H2; p, U64 H1'-A35 H2; q, G36 H1'-A35 H2; r, A35 H1'-A34 H2; s, A34 H1'-H2; t, U65 H1'-A34 H2; u, C37 H1'-A32 H2; v, A34 H1'-A33 H2; w, U66 H1'-A33 H2. Crosspeaks (x) and (y) are sequential interactions: x, A37 H1'-G38 H8; y, A35 H1'-G36 H8. In (B) and (C), parentheses indicate tentative assignments.

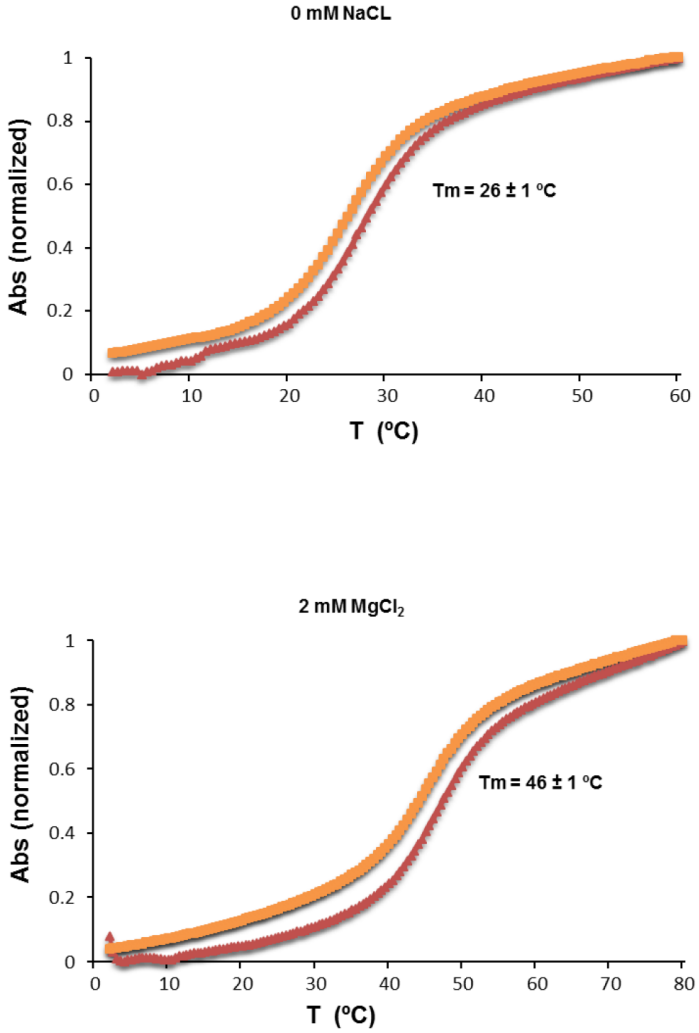


Figure S2
AbuQattam *et al*

Figure S2. UV-monitored thermal denaturation curves of I2s. The curves show I2s melting in aqueous buffers containing 2 mM sodium phosphate (pH 6.0) with no added salts, or additionally containing 2 mM $MgCl_2$. The average melting temperatures measured for I2s under these ionic conditions are indicated in the graphs. The melting curves were monophasic in all cases.

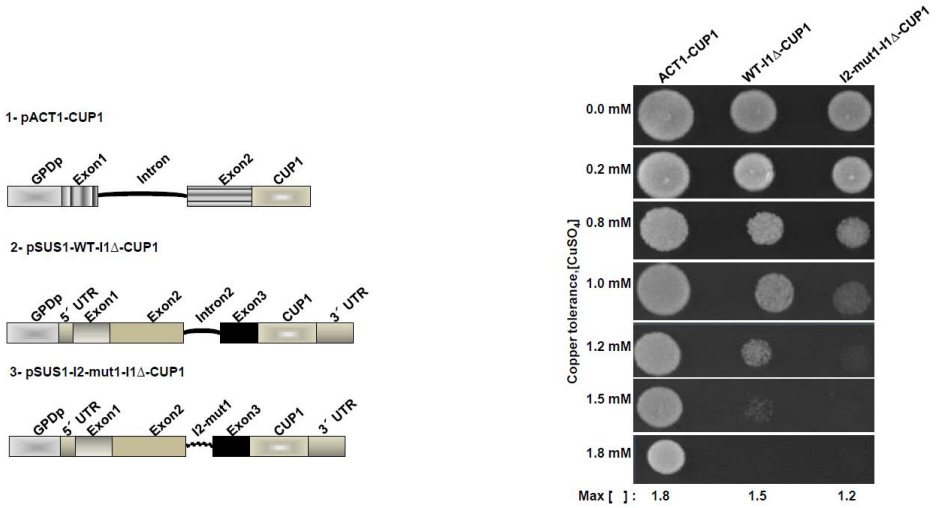


Figure S3
AbuQattam *et al*

Figure S3. The absence of *SUS1* intron 1 did not change the effect of I2-mut1 on *SUS1* processing.

Copper assay of *cup1*Δ cells transformed with plasmids containing SUS1-I1Δ -CUP1, SUS1-I2-mut1-I1Δ -CUP1 and ACT1-CUP1 as a control. Maximum copper tolerance is indicated.

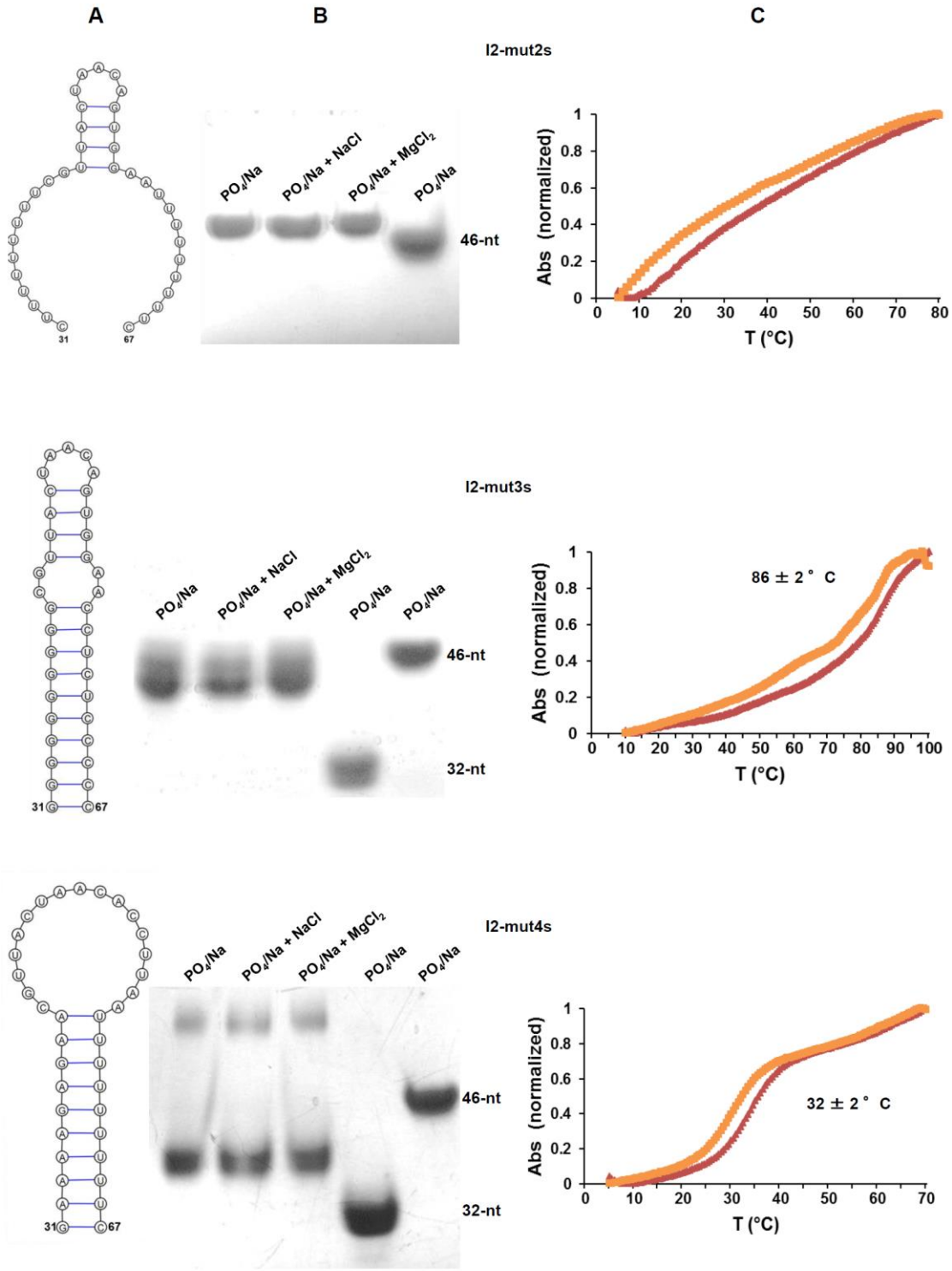


Figure S4
AbuQattam *et al*

Figure S4. *In vitro* analyses of intron 2 mutant oligonucleotides I2-mut2s, I2-mut3s and I2-mut4s. (A) MFold-predicted secondary structures of mutant I2s oligonucleotides. (B) Native gels comparing the electrophoretic mobility of 16-20 μM mutant I2s oligonucleotide samples previously annealed in different ionic conditions: (1) 2 mM sodium phosphate (pH 6.0), (2) 2 mM sodium phosphate and 100 mM NaCl, (3) 2 mM sodium phosphate and 2 mM $MgCl_2$. The right lanes contain 32-nt and/or 46-nt RNA hairpin controls. (C) UV-monitored thermal denaturation curves of mutant I2s oligomers in 2 mM sodium phosphate (pH 6.0) and 100 mM NaCl. The average melting temperatures of the I2-mut2s, I2-mut3s and I2-mut4s constructs under these ionic conditions are indicated in the graphs.

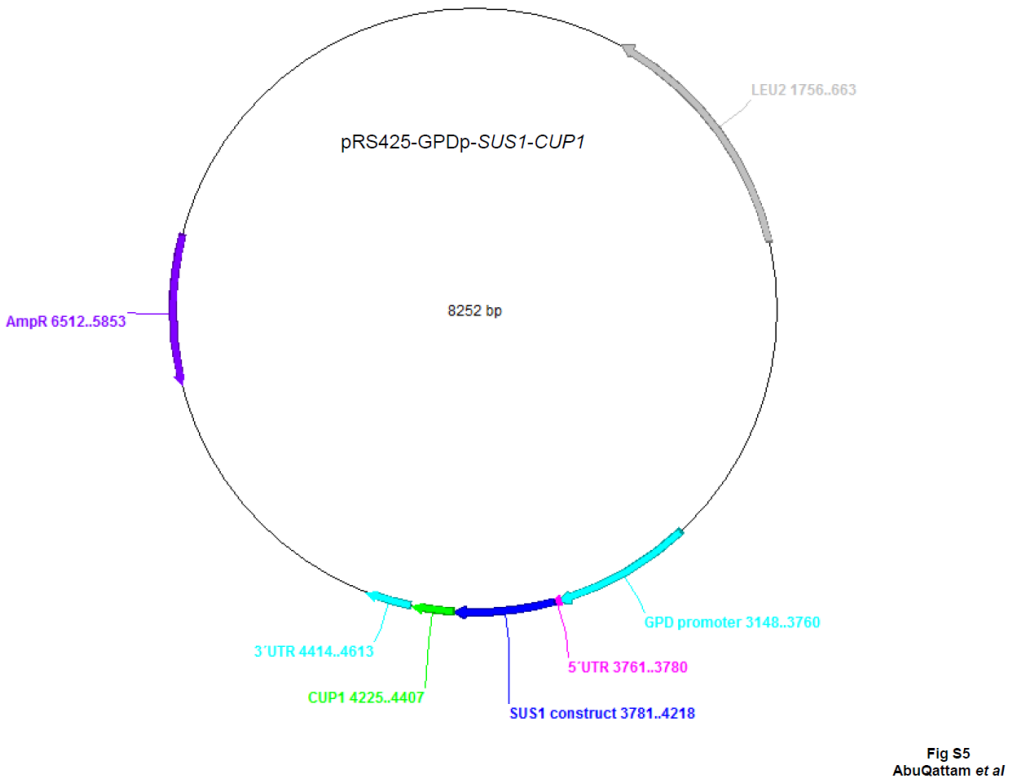


Figure S5. Schematic representation of the modified pRS425 plasmid used for cloning the GPDp-*SUS1-CUP1* constructs. The representation was generated by the ApE plasmid editor**.**

Supplementary Tables

Table S1. List of primers used in this study.

I2-mut1 forward	CCGTAGAACCCAAAGCATTAGGTATGTGAATATTTTATCACTAAATTTATTGAAATACTAACAGTGTTTCTAG
I2-mut1 Reverse	GTATGTGAATATTTTATCACTAAATTTATTGAAATACTAACAGTGTTTCTAGAAATGGTATCGGATTCAACA
I2-mut2 forward	CACTAAATTTATTctttttttttCGTTACTAACAG
I2-mMut2 Reverse	ATCACTAAATTTATTctttttttttCGTTACTAACAGTGG
I2-mutMut3 forward	ATTGggggGgGggCGTTACTAACAGTGGAAccUcUccccCTAGAAATGGTATCGGATTCAACActgcag
I2-mutMut3 reverse	CACTAAATTTATTGggggGgGggCGTTAC
Extreme forward 5´ primer for all constructs	AAAAAAaagcttCAATTCTGGCCTTCACTCCAATGACTATGGATACTGC
Extreme reverse 3´primer for all constructs	AAAAAActgcagTTGTGTATCTACAATCTC
SCR1Forward	AACCGTCTTTCCTCCGTCGTAA
SCR1 Reverse	AGAACTACCTTGCCGCACCA

CHAPTER 2

Impact of exon RNA structure in the expression of *SUS1*, an unusual yeast gene involved in mRNA biogenesis

Ali AbuQattam[a,b], Susana Rodríguez-Navarro[b,c*] and José Gallego[a*]

[a]Facultad de Medicina. Universidad Católica de Valencia, C/Quevedo 2, 46001 Valencia

[b]Gene Expression and RNA Metabolism Laboratory. Centro de Investigación Príncipe Felipe, C/ E. Primo Yúfera 3, 46012 Valencia

[c]Gene Expression and RNA Metabolism Laboratory. Instituto de Biomedicina de Valencia (CSIC), C/ Jaime Roig 11, 46010 Valencia

[*] Correspondence to: srodriguez@cipf.es and jose.gallego@ucv.es

Keywords: exon, NMR, RNA structure, SHAPE, Sus1, splicing, yeast.

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, 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 an unusually structured purine-rich loop. This loop contained two consecutive sheared G:A base pairs and was 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 experiments indicated that all mutants were capable of altering the quantities of unspliced and/or fully-spliced *SUS1* RNA transcripts relative to wild-type. The quantities of mutant RNA species suggested that the structure of E2 may be involved in other processes of *SUS1* RNA biogenesis in addition to splicing. Altogether, the findings described in this report have opened new questions regarding the structure and function of this exon that are under investigation.

Introduction

In eukaryotic cells, introns are eliminated from pre-mRNA molecules by the spliceosome, a dynamic multicomponent machine (Wahl, Will et al. 2009; Cate 2016; Papasaikas and Valcárcel 2016; Sperling 2017). The process of alternative splicing is widespread in higher eukaryotes, where it is considered to be a main source of protein diversity (Nilsen and Graveley 2010; Tress, Abascal et al. 2017). The pervasiveness of this mechanism has been firmly supported by transcriptome sequencing data (Ben-Dov, Hartmann et al. 2008; Stepankiw, Raghavan et al. 2015). Regardless of whether it leads to functional protein isoforms, alternative splicing contributes to modulate gene expression (Braunschweig, Gueroussov et al. 2013; Wong, Au et al. 2016). 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 (Johnson and Vilardell 2012). 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 (Chen 2016; Salzman 2016).

Sus1 (ENY2 in mammals) is a small, evolutionary conserved 11-KDa protein involved in several processes of mRNA biogenesis (Galán and Rodríguez-Navarro 2012). In *Saccharomyces cerevisiae*, Sus1 interacts during transcription elongation with RNA polymerase II and the export factors Yra1 and Mex67 (Pascual-García, Govind et al. 2008), and also accumulates at the nuclear pore, where it is part of the TREX-2 mRNA export complex (Rodríguez-Navarro, Fischer et al. 2004; Jani, Lutz et al. 2009). In addition, it participates in histone H2B deubiquitination as a component of the SAGA complex (Rodríguez-Navarro, Fischer et al. 2004; Köhler, Pascual-García et al. 2006).

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 (Rodríguez-Navarro, Fischer et al. 2004; Pascual-García and Rodríguez-Navarro 2009). 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 (Cuenca-Bono et al. 2011; Hossain et al. 2011). 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 the I2 hairpin structure changed the patterns of Sus1 expression as well as *SUS1* splicing, giving rise to I1 retention and skipping of the second exon (E2) (AbuQattam, Gallego et al. 2016) (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 E2 of *SUS1* influenced the splicing of the flanking introns (I1 and I2) (Cuenca-Bono, García-Molinero et al. 2011; Hossain, Rodriguez et al. 2011). 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 (Fair and Pleiss 2017). 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 (Warf and Berglund 2010; Jin, Yang et al. 2011; McManus and Graveley 2011; García-Robles, Sánchez-Navarro et al. 2012). In yeast, however, studies of the effect of RNA structure have focused on intronic sequences (Pérez-Valle and Vilardell 2012). It was therefore relevant to evaluate whether E2 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) (Wang, Bao et al. 2014) (Figure 1). In this report we explore the role of E2 structure on *SUS1* splicing by combining structural analyses with cellular assays evaluating the impact of mutant E2 sequences.

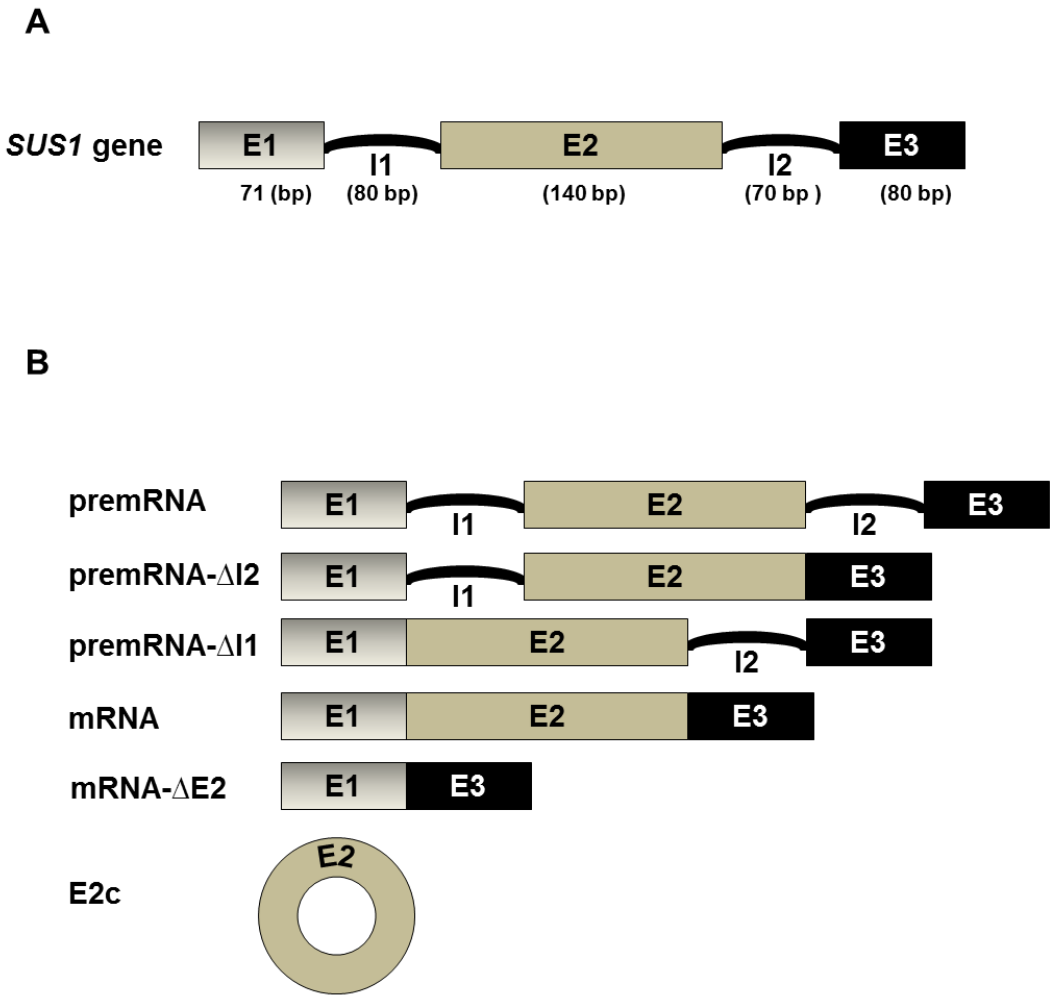


Fig. 1

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

Results

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

The RNA-folding algorithms Mfold (Zuker 2003), RNAfold (Lorenz, Bernhart et al. 2011), RNAstructure (Reuter and Mathews 2010) and MC-Fold (Parisien and Major 2008) 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 (Will, Joshi et al. 2012) and the E2 sequences of the seven yeast species containing two introns in the *SUS1* gene (Cuenca-Bono, García-Molinero et al. 2011) 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* (Cuenca-Bono, García-Molinero et al. 2011; Hossain, Rodriguez et al. 2011), prompted us to experimentally analyse the structure formed by this exon *in vitro*.

We first assessed the secondary structure adopted by the 140-nt E2 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 (Wilkinson, Merino et al. 2006; Deigan, Li et al. 2009; Weeks and Mauger 2011; Hajdin, Bellaousov et al. 2013). Using the average E2 reactivities (Figure 2B) and the RNAStructure software (Reuter and Mathews 2010), 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 (Lescoute and Westhof 2006; de la Pena, Dufour et al. 2009). 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. This latter fact would be consistent with the low SHAPE reactivities exhibited by many junction nt (Figure 2C). 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).

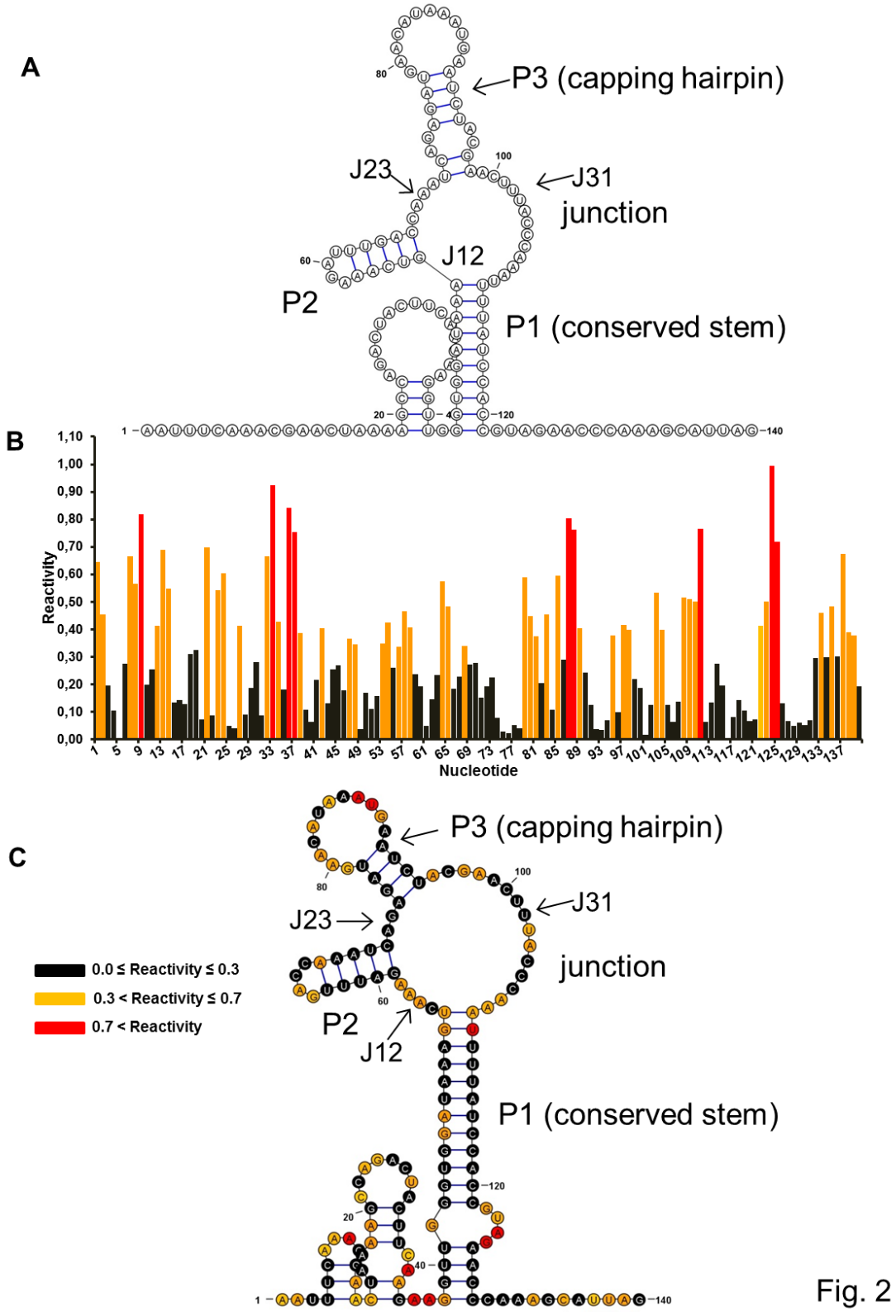


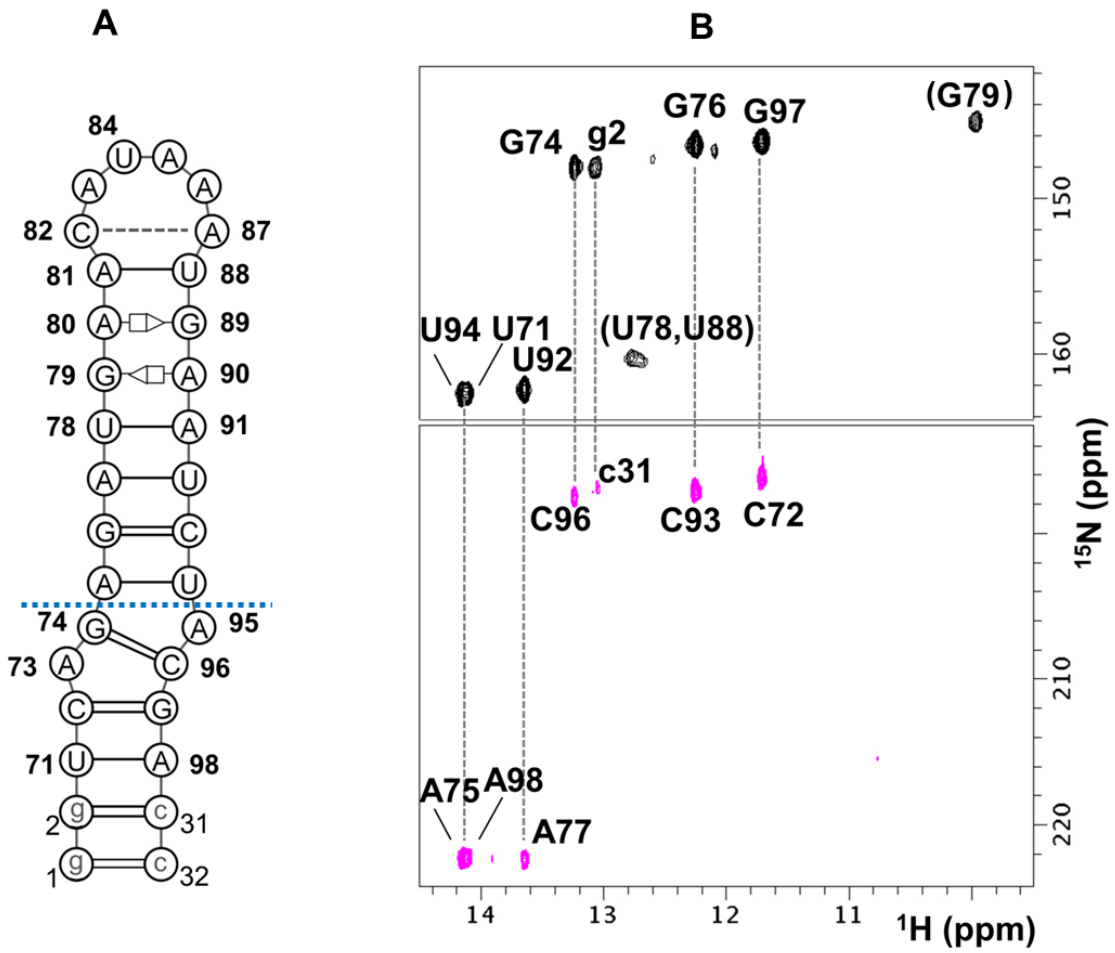
Fig. 2

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 (Lorenz, Bernhart et al. 2011) web server with default parameters. The RNAStructure (Reuter and

Mathews 2010) and MC-Fold (Zuker 2003) 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 (Reuter and Mathews 2010) folding algorithm. In (**A**) and (**B**), 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.

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, G_{79}AACAUAAAUGA_{90}, had an unusually high percentage of purines (75%) and particularly adenine bases (58%), which have been shown to be frequently involved in tertiary interactions (Butcher and Pyle 2011). As part of our efforts to characterize the structure of E2, we next examined by NMR spectroscopy and gel electrophoresis 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 $A_{73}G_{74}$:$A_{95}C_{96}$ 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 S3A), and the NMR data indicated that it adopted the stem-loop structure depicted in Figure 3A.


A
B
84
82
87
81
88
80
89
79
90
78
91
74
95
73
96
71
98
2
31
1
32
G74
g2
G76
G97
(G79)
U94
U71
U92
(U78,U88)
c31
C96
C93
C72
A75
A98
A77
150
160
210
220
15N (ppm)
14
13
12
11
1H (ppm)

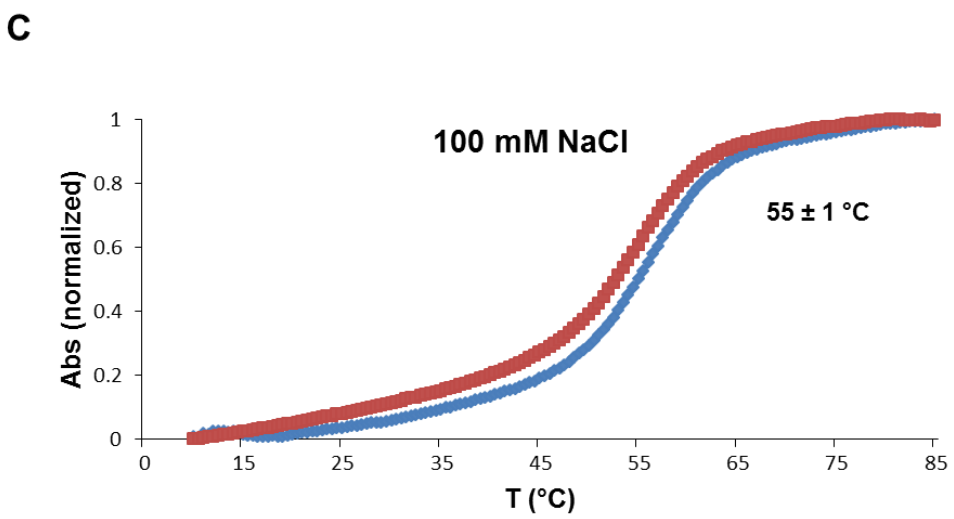

C
100 mM NaCl
55 ± 1 °C
Abs (normalized)
1
0.8
0.6
0.4
0.2
0
0
15
25
35
45
55
65
75
85
T (°C)

Fig. 3

Figure 3. Structure and stability of the E2 capping hairpin, P3. (**A**) NMR-supported structure of the 32-nt 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 (Leontis, Stombaugh et al. 2002). 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**) Assignment of HSQC (top) and HNN-COSY (bottom) spectra of E2s acquired at 18 ^{o}C, based on NOESY and TOCSY analyses in H_2O and D_2O. Parentheses indicate tentative assignments. A broad imino resonance with a chemical shift similar to G79 was also detected and assigned to G89. (**C**) 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.

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 (Figure 3A). 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* Hoogsten-sugar edge) conformation (Figure S4A). These G_{79}:A_{90} and A_{80}:G_{89} sheared pairs were stacked between U_{78}:A_{91}, 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 (Figure 3A). 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 S4B), and by sequential and cross-strand H2-H1' NOEs involving the H2 protons of A80, A81, A90 and A91 (Figure S4B, crosspeaks i, j, k, l, s, t, u and v). The presence of the tandem G:A pairs sandwiched between Watson-Crick U_{78}:A_{91} and A_{81}:U_{88} pairs was likewise supported by the detection at 25 ^{o}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 3B), which are not engaged in hydrogen bonding interactions with the opposite adenines in the sheared pairs (Figure S4A); 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 3B); and finally by the

strong upfield shift of A91 H1' (Figure S4B), typical of a purine residue following a sheared G:A pair (Flinders and Dieckmann 2001). In this last respect, tandem sheared G:A pairs flanked by pyrimidine residues on their 5' ends, as in the case in the E2 capping hairpin (Figure 3A), have been observed to give rise to inter-strand stacking of purine residues (Flinders and Dieckmann 2001).

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 S4B), as well as by a strong sequential A87 H2-U88 H1' NOE (Figure S4, crosspeak r). The remaining A83, U84, A85 and A86 of the E2s apical region closed the terminal loop (Figure 3A). They all exhibited strong TOCSY signals typical of C2'-*endo* sugars and had weaker or non-detectable sequential NOE interactions (Figure S4B), particularly U84, whose base was likely exposed in the solvent.

The $A_{73}G_{74}$:$A_{95}C_{96}$ internal loop in the middle of the E2s stem also adopted an unusual structure. Instead of the expected A_{73}:C_{96} and G_{74}:A_{95} pairs (Figure 2A), G74 established a Watson-Crick pair with C96 flanked by unpaired A73 and A95 bases (Figure 3A). The presence of the G_{74}:C_{96} pair was unambiguously demonstrated by the detection of a hydrogen bond between G74 HN1 and C96 N3 with HNN-COSY experiments (Figure 3B). The unpaired A73 and A95 nt were stacked between G_{74}:C_{96} and the neighbouring canonical pairs of the stems, as revealed by the detection of C72-A73-G74-A75 and U94-A95-C96-G97 sugar-aromatic and aromatic-aromatic NOEs (Figure S4B), by sequential and cross-strand H2-H1' NOEs involving the H2 hydrogens of A73 and A95 (Figure S4B, crosspeaks d, w and x), and by the C3'-*endo* character of the A73, A95 and C96 sugars at 25 ^{o}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 3B and S4B).

We also assessed the stability of the E2s stem-loop with UV thermal denaturation experiments. The E2s hairpin melted reversibly at 55 ^{o}C in an aqueous solution containing 100 mM NaCl, and at 66 ^{o}C in the presence of 2 mM $MgCl_2$ (Figure 3C and Figure S3B). 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 (Figure 3A), which likely contributed to increase the E2s melting temperature.

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 *SUS1* in *S. cerevisiae* 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). As expected, the predicted structure of E2-muts changed significantly relative to that formed by the wild-type sequence (Figure 4A).

To evaluate the expression efficiency of this mutant *in vivo*, we used the *CUP1* reporter system (Lesser and Guthrie 1993). *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. 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 also detected a clear reduction in the amount of fully spliced mRNA relative to the wild-type cultures (Figures 4C and 4D).

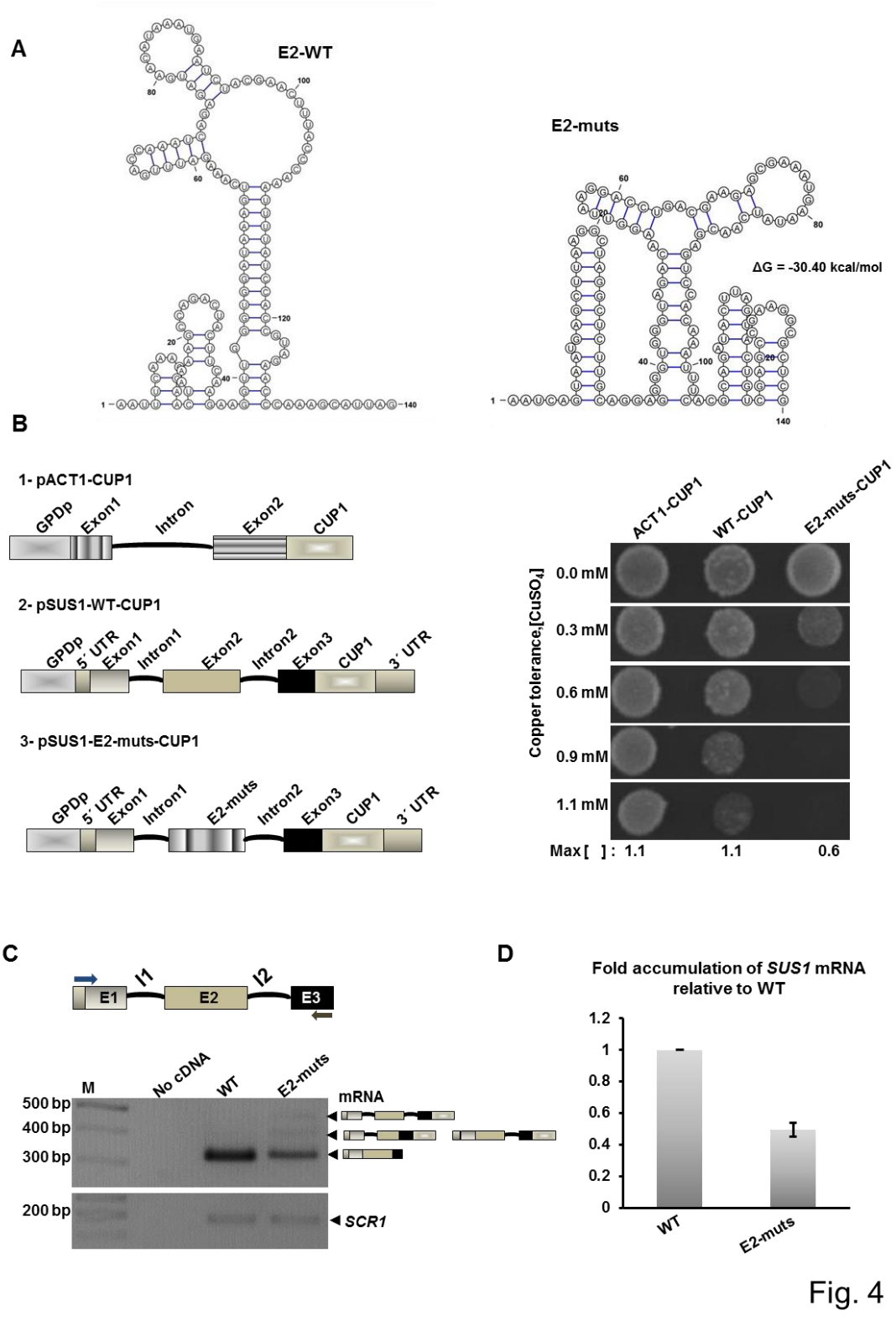


Fig. 4

Figure 4. Synonymous mutation of E2 impacts *SUS1* splicing 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 (Zuker 2003) is shown on the right. For comparison, we

obtained ΔG = -19.80 Kcal mol^{-1} for the wild.-type sequence with the same folding program. (**B**) Copper assay of *cup1*Δ cells transformed with plasmids containing SUS1g-CUP1, SUS1-E2-muts-CUP1 and ACT1-CUP1 as a control. Maximum copper tolerance (mM) is indicated (**C**) Semi q-RT-PCR (25 cycles) to amplify *SUS1* transcripts from cells expressing SUS1g-CUP1 (WT) and SUS1-E2-muts-CUP1 (E2-muts), resolved in a 3% agarose gel. A PCR without cDNA (no cDNA) was included as control. The stick diagram on the right of the gel identifies the *SUS1* RNA species (fully spliced mRNA). (**D**) qPCR showing mRNA accumulation from cells expressing SUS1g-CUP1 (WT), SUS1-E2-muts-CUP1 (E2-muts) and SUS1-E2-mut1-CUP1 (E2-mut1) normalized to the amount of *SCR1* (lower panel). Error bars represent SE for at least three independent experiments.

4. The three-way junction structure formed by E2 modulates SUS1 splicing and Sus1 expression.

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 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). Mutant E2-mut2 was designed to disrupt the conserved stem (P1) at the base of the E2 junction. Thus, the $G_{43}GUGG_{47}$:$C_{117}CACC_{121}$ base pairs were replaced with $\underline{C}_{43}\underline{C}U\underline{C}\underline{C}_{47}$:$C_{117}C\underline{U}CC_{121}$ (mutated nt are underlined) (Figure 5B, left panel). Mutant E2-mut2r was intended to restore this stem by additionally replacing $C_{117}C_{118}$ with GG, $C_{120}C_{121}$ with GG and U_{45} with A while keeping the E2-mut2 mutations, so that equivalent $\underline{C}_{43}\underline{CACC}_{47}$:$\underline{G}_{117}\underline{GUGG}_{121}$ base pairs would replace the wild-type pairings (Figure 5B, right panel). Mutants E2-mut2 and E2-mut2r introduced 4 and 5 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).

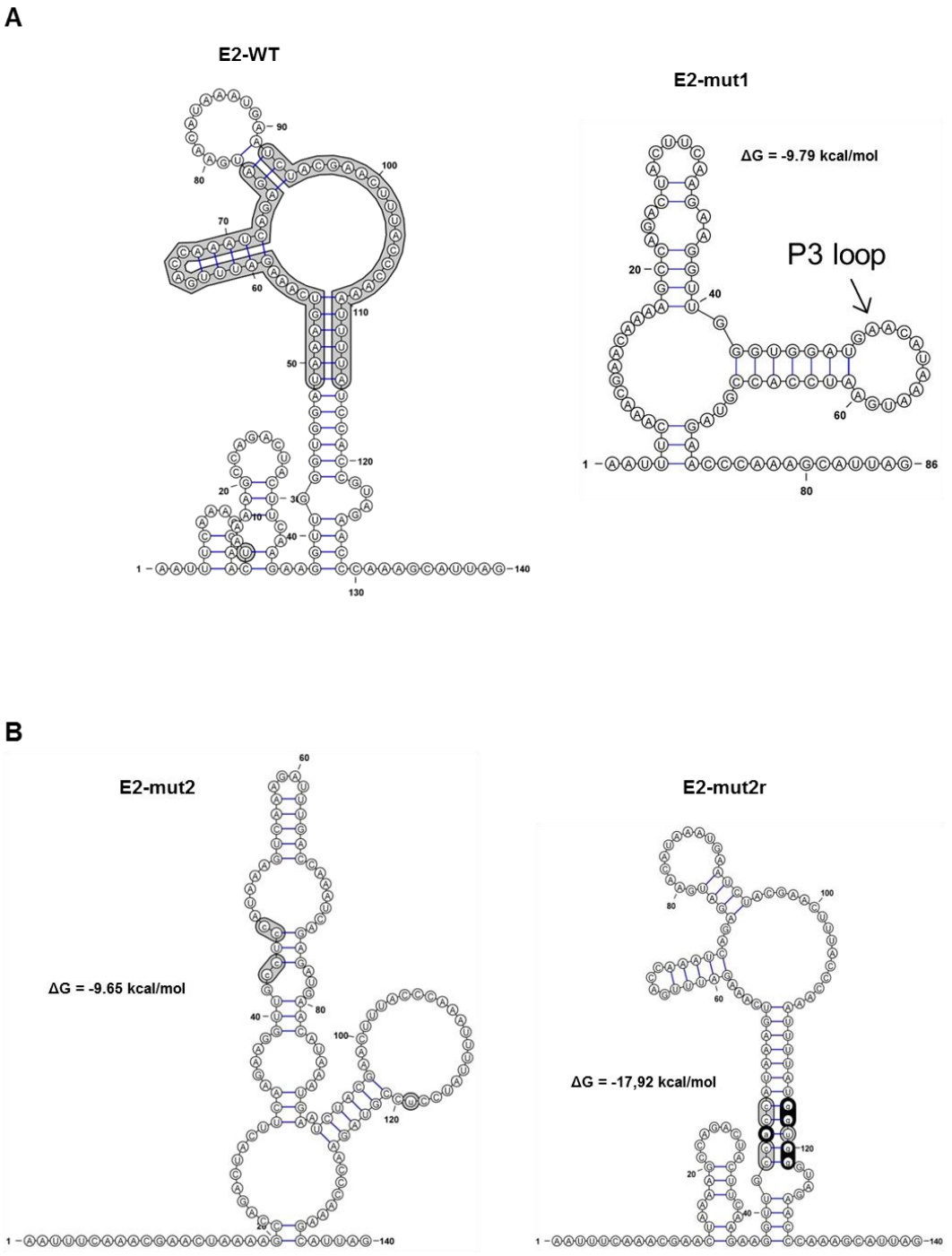


Fig. 5

Figure 5. Predicted secondary structure of *SUS1* mutants affecting the three-way junction or 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 (Zuker 2003), is shown on the right. In this mutant, the three-way junction was eliminated while keeping the P3 purine-rich loop (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 nt were mutated to generate each mutant. For comparison, ΔG = -19.80 Kcal mol^{-1} was obtained for wild-type E2 with the same folding program.

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 *cup1*Δ 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), indicating that deletion of the E2 junction decreased *SUS1* expression (Figure 6A). Likewise, the E2-mut2 mutant disrupting the P1 stem also had a deleterious effect (copper tolerance up to 0.9 mM) that was not observed in the E2-mut2r mutant, which restored cupper resistance to wild-type levels (copper tolerance up to 1.1 mM) (Figure 6B). To assess the production of *SUS1* isoforms by these mutants, semi-q-RT-PCR was carried out to amplify *SUS1* transcripts in *sus1*Δ cells bearing wild-type and mutant constructs. In agreement with the copper assay observations, a slight decrement was detected in the levels of E2-mut1 fully spliced mRNA relative to wild-type, indicative of reduced splicing efficiency (Figures 7A and 7B). Although we observed a reduction of copper tolerance for the E2-mut2 mutant destabilizing the P1 stem of the junction, semi-q-RT-PCR did not reveal significant differences of fully-spliced *SUS1* mRNA relative to wild-type, and a similar result was obtained with the E2-mut2r mutant restoring the P1 stem (Figure 7C). *SUS1* expression is reduced when cells are incubated at higher temperatures (for instance 20 min at 42 ºC), and this reduction of *SUS1*

expression is accompanied by accumulation of unspliced *SUS1* transcripts and a concomitant decrease of fully spliced mRNA (Cuenca-Bono, García-Molinero et al. 2011). Recently, we also showed that *SUS1* I2 structural mutants affected the ratio of the different species in these conditions (AbuQattam, Gallego et al. 2016). We thus tested the effect of the destabilization of the P1 stem (E2-mut2) and its restoration by E2-mut2r at 42 ºC. As shown in Figure 7D, accumulation of unspliced forms of SUS1 transcripts became evident in E2-mut2. Notably, restoring the structure of the P1 stem abolished this accumulation.

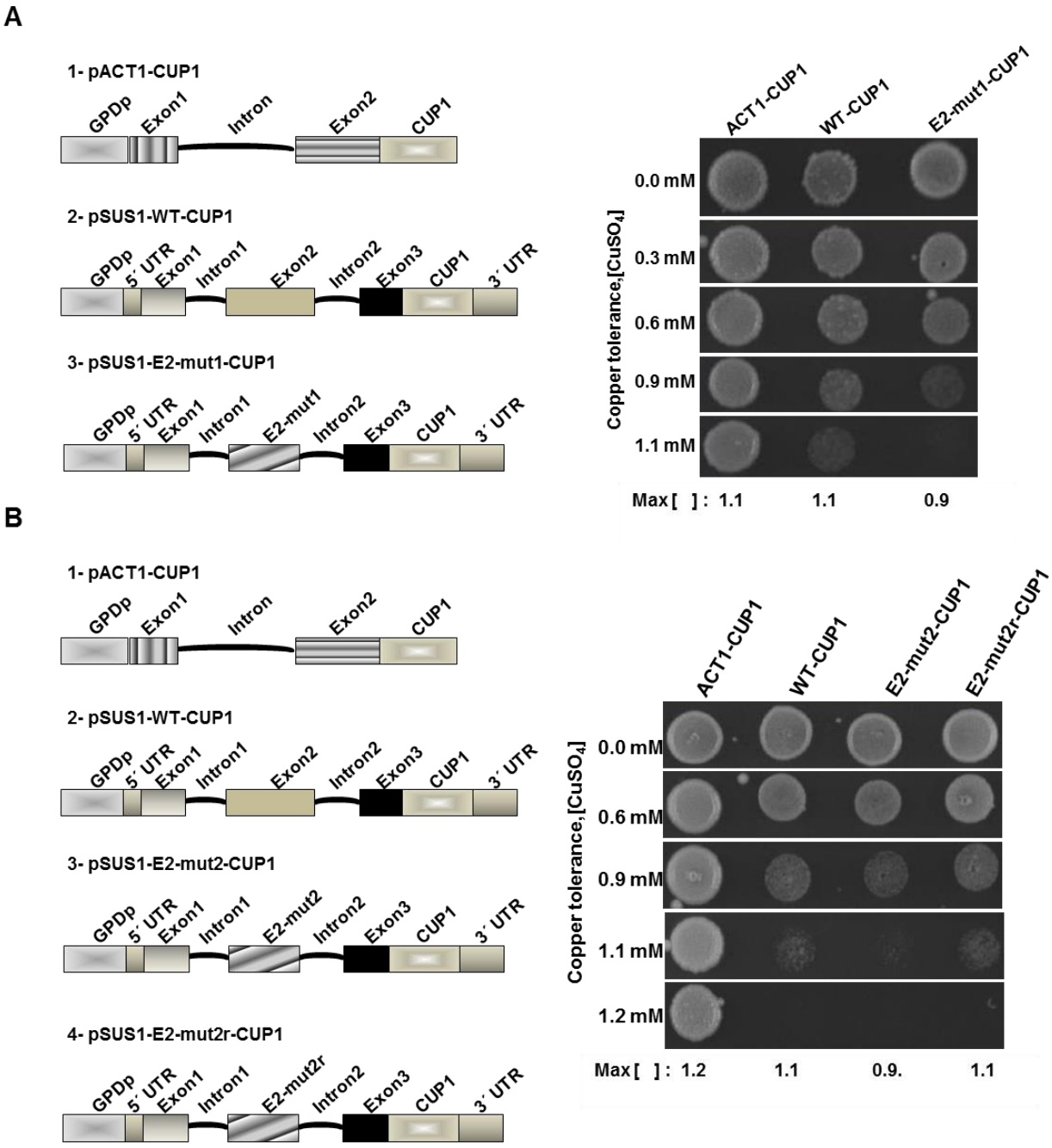

A
1- pACT1-CUP1
GPDp
Exon1
Intron
Exon2
CUP1
2- pSUS1-WT-CUP1
GPDp
5´ UTR
Exon1
Intron1
Exon2
Intron2
Exon3
CUP1
3´ UTR
3- pSUS1-E2-mut1-CUP1
E2-mut1
ACT1-CUP1
WT-CUP1
E2-mut1-CUP1
0.0 mM
0.3 mM
0.6 mM
0.9 mM
1.1 mM
Copper tolerance,[CuSO4]
Max [ ] : 1.1 1.1 0.9
B
1- pACT1-CUP1
2- pSUS1-WT-CUP1
3- pSUS1-E2-mut2-CUP1
E2-mut2
4- pSUS1-E2-mut2r-CUP1
E2-mut2r
ACT1-CUP1
WT-CUP1
E2-mut2-CUP1
E2-mut2r-CUP1
0.0 mM
0.6 mM
0.9 mM
1.1 mM
1.2 mM
Copper tolerance,[CuSO4]
Max [ ] : 1.2 1.1 0.9. 1.1

Fig. 6

Figure 6. The three-way junction and the P1 stem of E2 affect Sus1 gene expression. (**A**) Copper assay of *cup1*Δ cells transformed with plasmids containing SUS1g-CUP1 (WT), SUS1-E2-mut1-CUP1 and ACT1-CUP1 as a control. Maximum copper tolerance (mM) is indicated (**B**) Copper assay of *cup1*Δ cells transformed with plasmids containing SUS1g-CUP1, SUS1-E2-mut2-CUP1, SUS1-E2-mut2r-CUP1 and ACT1-CUP1 as a control. Maximum copper tolerance (mM) is indicated.

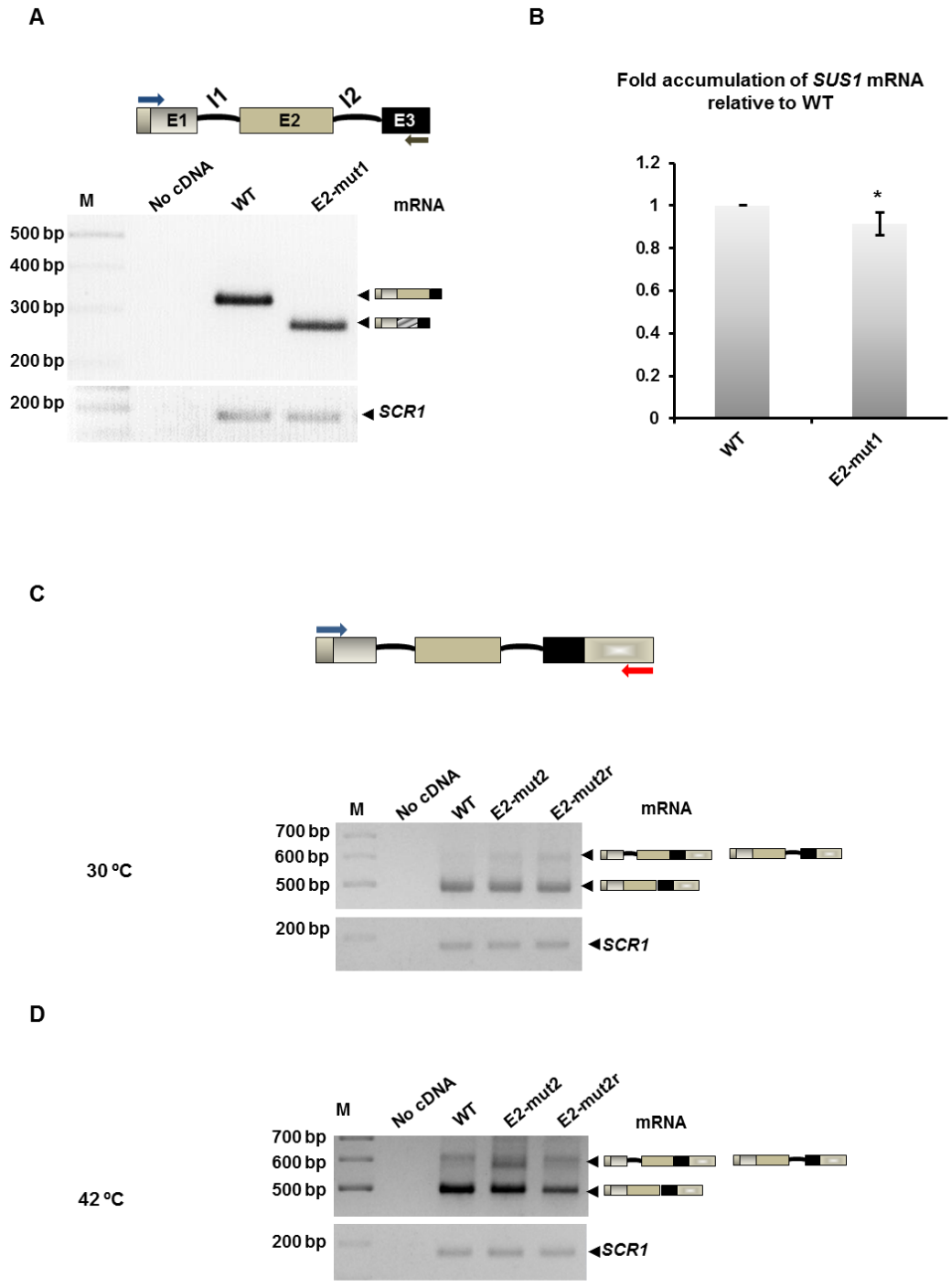


Fig. 7

Figure 7. *SUS1* splicing efficiency is affected by the three-way junction and the P1 stem of E2. (**A**) Semi q-RT-PCR (25 cycles) to amplify *SUS1* transcripts resolved in a 3% agarose gel from cells expressing SUS1g-CUP1 (WT) and SUS1-E2-mut1-CUP1 (E2-mut1). A PCR without cDNA (no cDNA) was included as control. The stick diagrams on the right of the gel identify the *SUS1* RNA transcripts. (**B**) 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. (**C**) Semi q-RT-PCR (25 cycles) to amplify *SUS1* transcripts resolved in a 3% agarose gel from cells expressing SUS1g-CUP1 (WT) or SUS1-E2-mut2-CUP1 (E2-mut2) and SUS1-E2-mut2r-CUP1 (E2-mut2r) at 30 ºC in each case. A PCR without cDNA (no cDNA) was included as control. The stick diagram on the right of the gel identifies the *SUS1* RNA species (fully spliced mRNA). (**D**) as in (C) after 20 minutes of incubation the cultures at 42 ºC.

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 (Wan, Qu et al. 2014). One highly regulated step of mRNA biogenesis is splicing, the process by which introns are removed from pre-mRNAs transcripts. While the RNA structure of intron sequences has been shown to play an important role in splicing, little is known about how exon RNA folding influences splicing in *S. cerevisiae*. 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 (Cuenca-Bono, García-Molinero et al. 2011; Hossain, Rodriguez et al. 2011).

We first analysed the secondary structure adopted by the full-length E2 RNA with SHAPE, a technique based on detecting differences in ribose 2'-hydroxy group reactivity modulated by local backbone flexibility (Weeks and Mauger 2011). The results clearly indicated that the exon formed a 16-base pair discontinuous stem (P1; Figure 2C). 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 (Lescoute and Westhof 2006) connected stem P1 with hairpins P2 and P3 in the SHAPE-driven E2 structure (Figure 2C). 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 (Lescoute and Westhof 2006). 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 are widely distributed in all kingdoms of life and have been observed to play essential functional and architectural roles in riboswitches, ribozymes and ribonucleoprotein systems (de la Pena, Dufour et al. 2009). 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 adopt a complex three-dimensional fold. Compared to other type-C intersections, the J31 strand of the E2 junction is unusually long (15 nt; Figure 2C) but not unprecedented: a J31 strand of 14 nt has been observed in 23S rRNA (Lescoute and Westhof 2006). Likewise, P3 nt U84 and A86, which were exposed in the terminal loop of the isolated P3 hairpin studied by NMR (Figure 3A; see below), had low reactivities in the context of the full-length sequence studied by SHAPE (Figure 2C). This would also agree with the hypothesis that the P3 loop may establish tertiary contacts in the context of the full-length E2 RNA.

The apical loop of the P3 hairpin comprised 12 nt and had an unusually high percentage of purines (75%) and particularly adenines (58%), which have been shown to be frequently involved in RNA-RNA tertiary interactions (Butcher and Pyle 2011). The complex structure of the E2 junction and the unusual base composition of the P3 loop led us to undertake a more detailed study of the structure of the P3 hairpin using NMR spectroscopy. The results indicated that both the apical and the internal loops of the hairpin were unusually structured (Figure 3A). The NMR data indicated that the apical loop comprised two consecutive sheared G:A pairs flanked by Watson-Crick U_{78}:A_{91} and A_{81}:U_{88} pairs (Figure 3A). The spectral data were also consistent with an additional C:A pair stacked on A_{81}:U_{88}, and indicated that the terminal P3 loop was closed by the remaining A83, U84, A85 and A86 residues (Figure 3A). On the other hand, the $A_{73}G_{74}$:$A_{95}C_{96}$ internal loop predicted to be formed in the middle of the P3 stem by most folding algorithms also adopted an unusual structure: the non-opposing G74 and C96 bases established a Watson-Crick pair flanked by unpaired but stacked adenine bases (Figure

3A). Not surprisingly, UV-monitored denaturation experiments indicated significant thermal stability for the P3 hairpin (Figure S3), which agreed with the detection by NMR spectroscopy of non-canonical base pairs and stacked residues in the apical and internal loops.

Although it will be necessary to obtain a high-resolution structure of the P3 hairpin to confirm the structural characteristics depicted above, 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 (Dagenais, Girard et al. 2017). 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 (Flinders and Dieckmann 2001). 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 (Figure 3A). Interestingly, the active site of the VS ribozyme, like that of the hairpin ribozyme, is formed by the tertiary association of the substrate loop containing the scissile phosphate with another loop (Lilley 2011; Dagenais, Girard et al. 2017). In the case of the hairpin ribozyme, the loops belong to helices emerging from the same four-way junction (Ferré-D'Amaré and Scott 2010; Lilley 2011). Based on the detected analogies, a thought-provoking hypothesis would be that E2 might contain a potentially scissile phosphate in loop P3 that could be cleaved as a result of the tertiary association of that loop with domain P1 or some other domain located within the *SUS1* RNA sequence, or even in a different RNA molecule. Ribozymes have been unexpectedly found in many genomes (including humans), but up to now they have only been identified in introns (de la Peña and García-Robles 2010; García-Robles, Sánchez-Navarro et al. 2012). Alternatively, the unusual structure of loop P3 might serve as a binding platform for a trans-acting protein factor. Further structural and biochemical analyses will be necessary to test these hypotheses and to determine the three-dimensional structure of the E2 RNA.

The presence of a novel E2 RNA structure and its partial conservation hinted for a functional role in Sus1 expression. We engineered and studied with functional assays four structural E2 mutants with the aim of altering the structure formed by E2. In mutant E2-muts, all of the E2 RNA sequence was modified with synonymous mutations that did not alter the length of the exon or the sequence of the protein. As a consequence, the predicted structure of the E2 RNA changed significantly relative to that formed by the wild type sequence (Figure 4A). In mutant E2-mut1, the three-way junction was removed while keeping unchanged the P3 purine-rich loop (Figure 5A). The remaining two structural mutants were designed to affect the conserved P1 stem. Mutant E2-mut2 disrupted this stem, while mutant E2-mut2r restored it (Figure 5B). Like E2-muts, mutants E2-mut2 and E2-mut2r did not affect the size of the exon.

The most drastic effects were observed with the synonymous mutant (E2-muts) (Figure 4). The maximum copper tolerance for this construct was 0.6 mM $CuSO_4$, which is consistent with the reduced *SUS1* copper tolerance observed in our previous results substituting E2 with a 140-nt sequence derived from *TAF14* (0.5 mM $CuSO_4$) (Cuenca-Bono, García-Molinero et al. 2011). Cells expressing E2-muts showed a clear reduction in the abundance of fully spliced *SUS1* transcript that was accompanied by a smaller increase of unspliced forms (see Figure 4C). This might indicate that the sequence and the structure of E2 are important not only for splicing, but may also be involved in the regulation of transcript stability and/or degradation (Johansson, He et al. 2007; Cuenca-Bono, García-Molinero et al. 2011).

Elimination of the three-way junction while keeping the purine-rich loop (E2-mut1) led to small reduction of *SUS1* copper tolerance relative to wild-type (0.9 mM $CuSO_4$ *vs* 1.1 mM $CuSO_4$ for wild-type), and a similar effect was observed for the E2-mut2 mutant disrupting the P1 stem. As expected, the E2-mut2r compensatory mutant canceled the effect observed with E2-mut2, restoring cupper tolerance to wild-type levels (Figure 6). Interestingly, E2-mut1 and E2-mut2 behaved similarly in copper resistance compared to wild-type, but affected splicing efficiency in different ways (compare Figures 7A and 7C). While E2-mut1 reduced the amount of fully spliced *SUS1* transcript without increasing the amount of unspliced transcripts, disruption of the

P1 stem slightly augmented intron retention. Again, this might indicate that the E2 RNA structure is involved in several aspects of transcript metabolism, which could include transcription, export, stability and degradation. The P1 stem, conserved across the seven yeast species containing two introns in the *SUS1* gene, might be particularly important for splicing events affecting both introns. A possible explanation for this may reside on the fact that the long (16-base pair) P1 stem decreases the effective nt length of the E2 exon located between the two introns. This might facilitate processes such as exon definition (Fair and Pleiss 2017), exon skipping and circRNA generation. The last two activities have been observed for *SUS1* E2 (Figure 1). Supporting this model, under thermal stress E2-mut2 clearly accumulated intron-containing species, and the E2-mut2r restoring the P1 stem canceled this effect (see Figure 7D). It is worth noting in this regard that this phenotype was observed under stress conditions, a situation in which the ratio of the different *SUS1* transcripts is more regulated (Cuenca-Bono, García-Molinero et al. 2011; AbuQattam, Gallego et al. 2016). These results demonstrate that similarly to intron 2 RNA structure (AbuQattam, Gallego et al. 2016), intrinsic features of E2 RNA structure can lead to intron retention.

In this regard, circRNAs are generated by back-splicing, and widespread generation of circRNAs also happens in yeasts (Wang, Bao et al. 2014). Splicing of *SUS1* has been shown to generate E2 in circular form (E2c), which can be amplified by PCR under normal growth conditions ((Wang, Bao et al. 2014) 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 several exciting new open questions including the role of RNA structure in E2c generation and E2c function, and the possibility that E2c plays a regulatory role *in trans* in Sus1 expression.

In conclusion, in this work we have shown that E2 adopts a three-way junction structure that includes a conserved double-helical stem and a novel structural motif contained in a purine-rich loop. The functional data indicated that exon 2 RNA structure exerted a role in *SUS1* transcript

biogenesis that was not restricted to splicing. The structural results and the functional data have opened several new possibilities that are now under investigation in our laboratories.

Materials and Methods

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 (Notredame, Higgins et al. 2000) web server.

Secondary structure predictions. Secondary structure predictions were carried out with the Mfold (Zuker 2003), RNAfold (Lorenz, Bernhart et al. 2011), RNAStructure (Reuter and Mathews 2010) and MC-Fold (Parisien and Major 2008) web servers using default parameters. The structures were drawn using VARNA (http://varna.lri.fr/) (Darty, Denise et al. 2009). We used the locARNA (Will, Joshi et al. 2012) 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 (Cuenca-Bono, García-Molinero et al. 2011).

Preparation of RNA samples for experiments *in vitro*. The DNA template used for transcribing the full-length *S. cerevisiae SUS1* E2 RNA sequence 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 (Wilkinson, Merino et al. 2006). 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 (Wilkinson, Merino et al. 2006). 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 unlabelled E2s samples, we also generated $^{13}C/^{15}N$-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.

SHAPE analysis. SHAPE was performed with N-methylisatoic anhydride (NMIA) essentially as described (Wilkinson, Merino et al. 2006). 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 $MgCl_2$ 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 (Karabiber, McGinnis

et al. 2013), which also normalized the reactivity values. Manual normalization of the reactivity data (Deigan, Li et al. 2009) 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 (Reuter and Mathews 2010) and MC-Fold (Parisien and Major 2008). We used m= 1.8 and b= -0.6 Kcal mol^{-1} in the SHAPE pseudo-free energy change term of RNAStructure (Hajdin, Bellaousov et al. 2013). 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 coloured red.

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 $MgCl_2$. 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.

Native gel electrophoresis. We used this technique to assess the number of structural species formed by E2s as well as their strand stoichiometry. The experiments were run at 4 °C for approximately 14 hours under constant voltage (90 V), and utilized 20% 19:1 acrylamide:bisacrylamide gels and 89 mM Tris-Borate (TB) as running buffer. They involved

8.9 µM E2s 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. Gels were stained with methylene blue and distained with water.

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 analysed using Topspin 3.5 (Bruker Biospin) and Sparky 3.113 (T. D. Goddard, D. G. Kneller, UCSF USA, 2004). The unlabelled E2s samples were studied with two-dimensional watergate-NOESY experiments (250 ms mixing time) recorded in 90% H_2O/10% D_2O 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 D_2O 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 $^{13}C/^{15}N$-labelled E2s transcript was analysed at 25 ^{o}C in D_2O with two-dimensional ^{1}H-^{13}C HSQC and HCCH-TOCSY and three-dimensional ^{13}C-edited NOESY-HMQC (100 and 200 ms mixing times), HCCH-e-COSY and HCP experiments. Two-dimensional ^{1}H-^{15}N HSQC experiments were also recorded in 90% H_2O/10% D_2O at 18 ^{o}C, as well as HNN COSY experiments allowing detection of hydrogen bonds between bases *via* two bond N-N couplings (Dingley and Grzesiek 1998). For ^{1}H-^{15}N 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 $^{2}J_{NN}$ coupling was set to 15 ms, and we typically collected 2048 and 128 complex points in the t_2 and t_1 dimensions, respectively, with 300 scans for each t_1 increment. The recycle delays ranged between 1.0 and 1.3 s for all experiments with labelled E2s samples.

Assignments and secondary structure determination. The assignment of E2s involved exchangeable imino and amino signals as well as non-exchangeable aromatic and H1', H2' and

H3' protons of all residues, and was based on standard analyses of NOESY, TOCSY and COSY spectra acquired in H_2O and D_2O at different temperatures, supported by the study of two-dimensional ^{1}H-^{13}C HCCH TOCSY, ^{1}H-^{13}C HSQC, ^{1}H-^{15}N HSQC and ^{1}H-^{15}N HNN-COSY data obtained from ^{15}N,^{13}C-labeled E2s samples dissolved in D_2O or H_2O. These analyses allowed us to identify canonical and non-canonical base pairs as well as stacking interactions present in the E2s hairpin structure.

Generation of *SUS1* constructs. All *SUS1* gene constructs contained the last twenty nt of the 5´UTR. The E2 mutants (muts, mut1, mut2 and mut2r) were constructed by the fusion PCR method (Yon and Fried 1989; Lesser and Guthrie 1993), 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*.

Yeast cultures and microbiological techniques. Copper resistance assays were carried out by growing the transformed *cup1*Δ cells at 30 ºC 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 OD_{600}. Subsequently, 10-fold serial dilutions of an equal number of cells were made and drops spotted onto SC-leu plates containing different concentration of $CuSO_4$ (Lesser and Guthrie 1993). Plates were photographed after 3-5 days of incubation at 30 ºC. Yeast cell transformations were done by the LiAc/SS carrier DNA/PEG method (Gietz and Schiestl 2007).

RNA extraction, Reverse transcription PCRs and semi q-RT-PCRs. Total RNA was harvested from *sus1*Δ 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 (Schmitt et al. 1990), and quantified using a Nanodrop. 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 OD_{600} 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 pairs of primers located upstream exon1 and exon3 was used to amplify the transcripts of *SUS1* (the number of cycles was adjusted depending on the abundancy 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 visualised 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/)

Acknowledgements

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., and Santiago Grisolía fellowship to A.AQ), and by Universidad Católica de Valencia. The authors wish to thank Amparo López-Carrasco and Ricardo Flores for help with the SHAPE experiments.

REFERENCES

AbuQattam, A., J. Gallego, et al. (2016). "An intronic RNA structure modulates expression of the mRNA biogenesis factor Sus1." RNA **22**(1): 75-86.

Ben-Dov, C., B. Hartmann, et al. (2008). "Genome-wide analysis of alternative pre-mRNA splicing." J Biol Chem **283**(3): 1229-1233.

Braunschweig, U., S. Gueroussov, et al. (2013). "Dynamic integration of splicing within gene regulatory pathways." Cell **152**(6): 1252-1269.

Butcher, S. E. and A. M. Pyle (2011). "The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks." Acc Chem Res **44**(12): 1302-1311.

Cate, J. H. (2016). "STRUCTURE. A Big Bang in spliceosome structural biology." Science **351**(6280): 1390-1392.

Cuenca-Bono, B., V. Garcia-Molinero, et al. (2011). "SUS1 introns are required for efficient mRNA nuclear export in yeast." Nucleic Acids Res **39**(19): 8599-8611.

Chen, L. L. (2016). "The biogenesis and emerging roles of circular RNAs." Nat Rev Mol Cell Biol **17**(4): 205-211.

Dagenais, P., N. Girard, et al. (2017). "Insights into RNA structure and dynamics from recent NMR and X-ray studies of the Neurospora Varkud satellite ribozyme." Wiley Interdiscip Rev RNA.

Darty, K., A. Denise, et al. (2009). "VARNA: Interactive drawing and editing of the RNA secondary structure." Bioinformatics **25**(15): 1974-1975.

de la Pena, M., D. Dufour, et al. (2009). "Three-way RNA junctions with remote tertiary contacts: a recurrent and highly versatile fold." Rna **15**(11): 1949-1964.

de la Peña, M. and I. García-Robles (2010). "Ubiquitous presence of the hammerhead ribozyme motif along the tree of life." RNA **16**(10): 1943-1950.

Deigan, K. E., T. W. Li, et al. (2009). "Accurate SHAPE-directed RNA structure determination." Proc Natl Acad Sci U S A **106**(1): 97-102.

Dingley, A. J. and S. Grzesiek (1998). "Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide (2)J(NN) couplings." Journal of the American Chemical Society **120**(33): 8293-8297.

Fair, B. J. and J. A. Pleiss (2017). "The power of fission: yeast as a tool for understanding complex splicing." Curr Genet **63**(3): 375-380.

Faza, M. B., S. Kemmler, et al. (2009). "Sem1 is a functional component of the nuclear pore complex-associated messenger RNA export machinery." J Cell Biol **184**(6): 833-846.

Ferré-D'Amaré, A. R. and W. G. Scott (2010). "Small self-cleaving ribozymes." Cold Spring Harb Perspect Biol **2**(10): a003574.

Flinders, J. and T. Dieckmann (2001). "A pH controlled conformational switch in the cleavage site of the VS ribozyme substrate RNA." J Mol Biol **308**(4): 665-679.

Galán, A. and S. Rodríguez-Navarro (2012). "Sus1/ENY2: a multitasking protein in eukaryotic gene expression." Crit Rev Biochem Mol Biol **47**(6): 556-568.

Garcia-Oliver, E., V. Garcia-Molinero, et al. (2012). "mRNA export and gene expression: The SAGA-TREX-2 connection." Biochim Biophys Acta **1819**(6): 555-565.

García-Robles, I., J. Sánchez-Navarro, et al. (2012). "Intronic hammerhead ribozymes in mRNA biogenesis." Biol Chem **393**(11): 1317-1326.

Hajdin, C. E., S. Bellaousov, et al. (2013). "Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots." Proc Natl Acad Sci U S A **110**(14): 5498-5503.

Hossain, M. A., C. M. Rodriguez, et al. (2011). "Key features of the two-intron Saccharomyces cerevisiae gene SUS1 contribute to its alternative splicing." Nucleic Acids Res **39**(19): 8612-8627.

Jani, D., S. Lutz, et al. (2012). "Functional and structural characterization of the mammalian TREX-2 complex that links transcription with nuclear messenger RNA export." Nucleic Acids Res **40**(10): 4562-4573.

Jani, D., S. Lutz, et al. (2009). "Sus1, Cdc31, and the Sac3 CID region form a conserved interaction platform that promotes nuclear pore association and mRNA export." Mol Cell **33**(6): 727-737.

Jin, Y., Y. Yang, et al. (2011). "New insights into RNA secondary structure in the alternative splicing of pre-mRNAs." RNA Biol **8**(3): 450-457.

Johansson, M. J., F. He, et al. (2007). "Association of yeast Upf1p with direct substrates of the NMD pathway." Proc Natl Acad Sci U S A **104**(52): 20872-20877.

Johnson, T. L. and J. Vilardell (2012). "Regulated pre-mRNA splicing: the ghostwriter of the eukaryotic genome." Biochim Biophys Acta **1819**(6): 538-545.

Karabiber, F., J. L. McGinnis, et al. (2013). "QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis." RNA **19**(1): 63-73.

Köhler, A., P. Pascual-García, et al. (2006). "The mRNA export factor Sus1 is involved in Spt/Ada/Gcn5 acetyltransferase-mediated H2B deubiquitinylation through its interaction with Ubp8 and Sgf11." Mol Biol Cell **17**(10): 4228-4236.

Kopytova, D. V., A. N. Krasnov, et al. (2010). "ENY2: couple, triple...more?" Cell Cycle **9**(3): 479-481.

Lang, G., J. Bonnet, et al. (2011). "The tightly controlled deubiquitination activity of the human SAGA complex differentially modifies distinct gene regulatory elements." Mol Cell Biol **31**(18): 3734-3744.

Leontis, N. B., J. Stombaugh, et al. (2002). "The non-Watson-Crick base pairs and their associated isostericity matrices." Nucleic Acids Res **30**(16): 3497-3531.

Lescoute, A. and E. Westhof (2006). "Topology of three-way junctions in folded RNAs." RNA **12**(1): 83-93.

Lilley, D. M. (2011). "Catalysis by the nucleolytic ribozymes." Biochem Soc Trans **39**(2): 641-646.

Lorenz, R., S. H. Bernhart, et al. (2011). "ViennaRNA Package 2.0." Algorithms Mol Biol **6**: 26.

McManus, C. J. and B. R. Graveley (2011). "RNA structure and the mechanisms of alternative splicing." Curr Opin Genet Dev **21**(4): 373-379.

Nilsen, T. W. and B. R. Graveley (2010). "Expansion of the eukaryotic proteome by alternative splicing." Nature **463**(7280): 457-463.

Notredame, C., D. G. Higgins, et al. (2000). "T-Coffee: A novel method for fast and accurate multiple sequence alignment." J Mol Biol **302**(1): 205-217.

Papasaikas, P. and J. Valcárcel (2016). "The Spliceosome: The Ultimate RNA Chaperone and Sculptor." Trends Biochem Sci **41**(1): 33-45.

Parisien, M. and F. Major (2008). "The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data." Nature **452**(7183): 51-55.

Pascual-García, P., C. K. Govind, et al. (2008). "Sus1 is recruited to coding regions and functions during transcription elongation in association with SAGA and TREX2." Genes Dev **22**(20): 2811-2822.

Pascual-García, P. and S. Rodríguez-Navarro (2009). "A tale of coupling, Sus1 function in transcription and mRNA export." RNA Biol **6**(2): 141-144.

Pérez-Valle, J. and J. Vilardell (2012). "Intronic features that determine the selection of the 3' splice site." Wiley Interdiscip Rev RNA **3**(5): 707-717.

Reuter, J. S. and D. H. Mathews (2010). "RNAstructure: software for RNA secondary structure prediction and analysis." BMC Bioinformatics **11**: 129.

Rodriguez-Navarro, S., T. Fischer, et al. (2004). "Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery." Cell **116**(1): 75-86.

Salzman, J. (2016). "Circular RNA Expression: Its Potential Regulation and Function." Trends Genet **32**(5): 309-316.

Sperling, R. (2017). "The nuts and bolts of the endogenous spliceosome." Wiley Interdiscip Rev RNA **8**(1).

Stepankiw, N., M. Raghavan, et al. (2015). "Widespread alternative and aberrant splicing revealed by lariat sequencing." Nucleic Acids Res **43**(17): 8488-8501.

Tress, M. L., F. Abascal, et al. (2017). "Alternative Splicing May Not Be the Key to Proteome Complexity." Trends Biochem Sci **42**(2): 98-110.

Wahl, M. C., C. L. Will, et al. (2009). "The spliceosome: design principles of a dynamic RNP machine." Cell **136**(4): 701-718.

Wan, Y., K. Qu, et al. (2014). "Landscape and variation of RNA secondary structure across the human transcriptome." Nature **505**(7485): 706-709.

Wang, P. L., Y. Bao, et al. (2014). "Circular RNA is expressed across the eukaryotic tree of life." PLoS One **9**(6): e90859.

Warf, M. B. and J. A. Berglund (2010). "Role of RNA structure in regulating pre-mRNA splicing." Trends Biochem Sci **35**(3): 169-178.

Weeks, K. M. and D. M. Mauger (2011). "Exploring RNA structural codes with SHAPE chemistry." Acc Chem Res **44**(12): 1280-1291.

Wilkinson, K. A., E. J. Merino, et al. (2006). "Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution." Nat Protoc **1**(3): 1610-1616.

Will, S., T. Joshi, et al. (2012). "LocARNA-P: accurate boundary prediction and improved detection of structural RNAs." RNA **18**(5): 900-914.

Wong, J. J., A. Y. Au, et al. (2016). "Intron retention in mRNA: No longer nonsense: Known and putative roles of intron retention in normal and disease biology." Bioessays **38**(1): 41-49.

Zhao, Y., G. Lang, et al. (2008). "A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing." Mol Cell **29**(1): 92-101.

Zuker, M. (2003). "Mfold web server for nucleic acid folding and hybridization prediction." Nucleic Acids Res **31**(13): 3406-3415.

Impact of exon RNA structure in the expression of *SUS1*, an unusual yeast gene involved in mRNA biogenesis

Ali AbuQattam, Susana Rodríguez-Navarro and José Gallego

SUPPLEMENTARY MATERIAL

Supplementary Figures

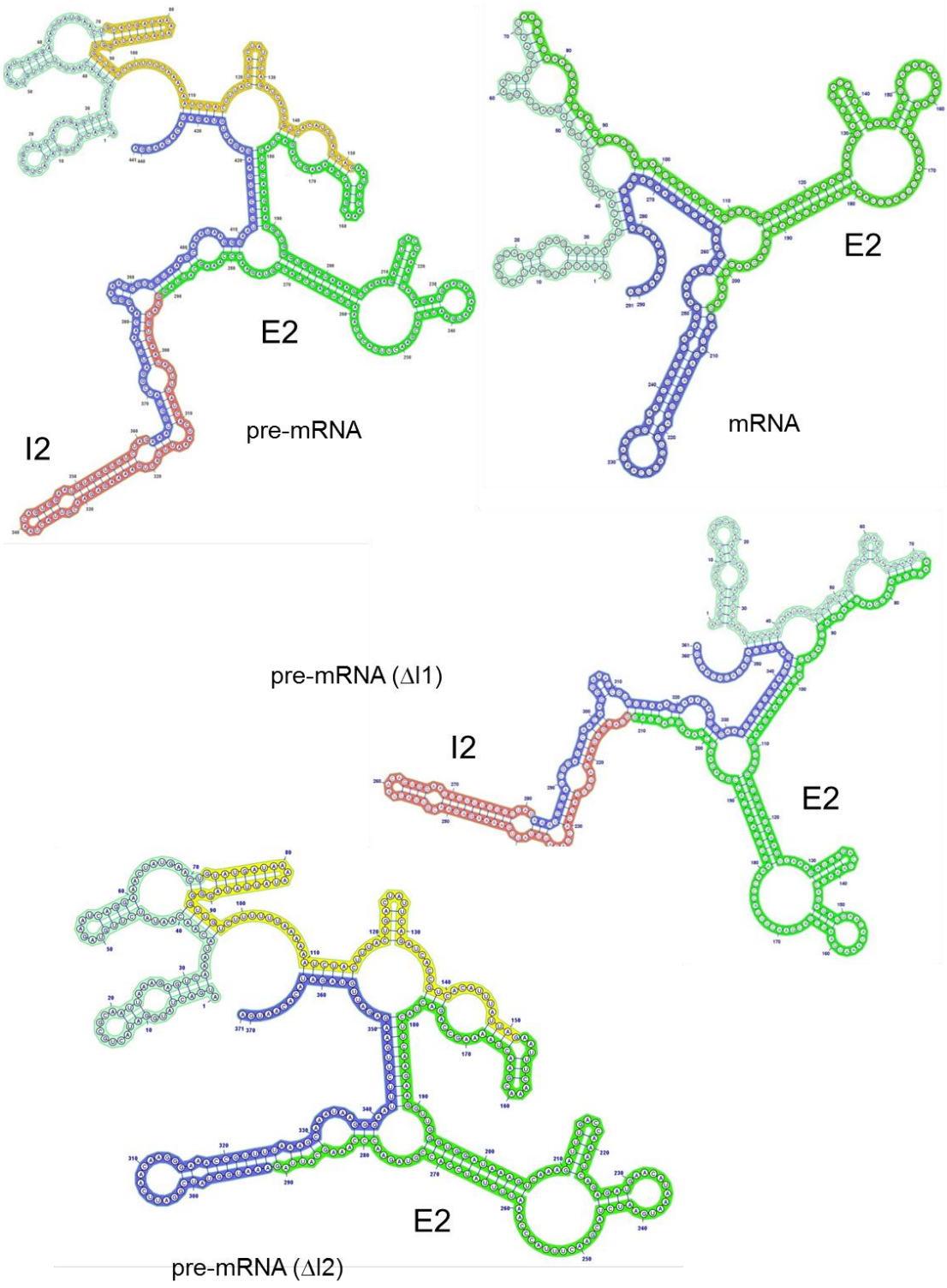


Fig. S1

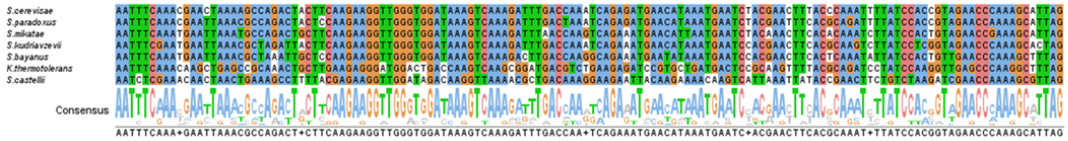
Figure S1. Predicted secondary structure of different *SUS1* RNA species of *S. cerevisae*: unspliced pre-mRNA, partly-spliced pre-meRNA lacking I1, partly-spliced pre-mRNA lacking I2, and fully-spliced mRNA. E2 (green-shaded nt) and the previously studied I2 (red-shaded nt) (AbuQattam, Gallego et al. 2016) form the same structures regardless of the sequence context. The other components of the gene E1,

E3 and I1 are represented in cyan, blue and gold, respectively. The predictions were carried out with Mfold (Zuker 2003) with default parameters, and the minimum free-energy structure is shown in all cases.

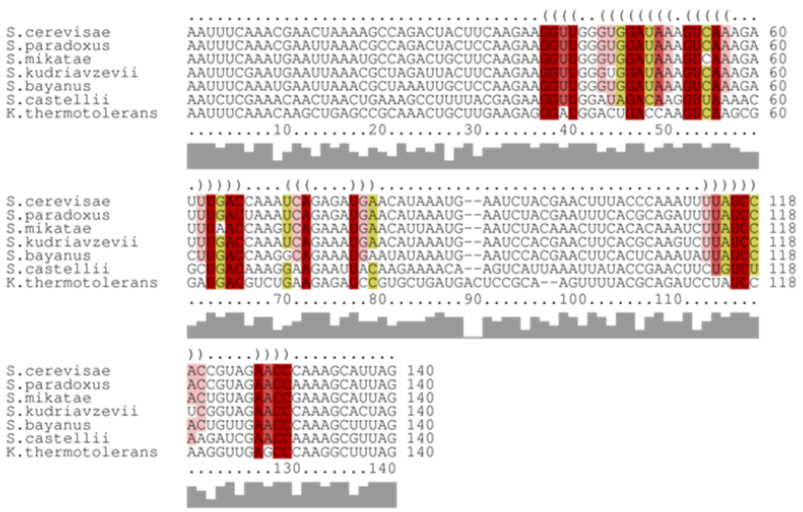
A

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%

B



C



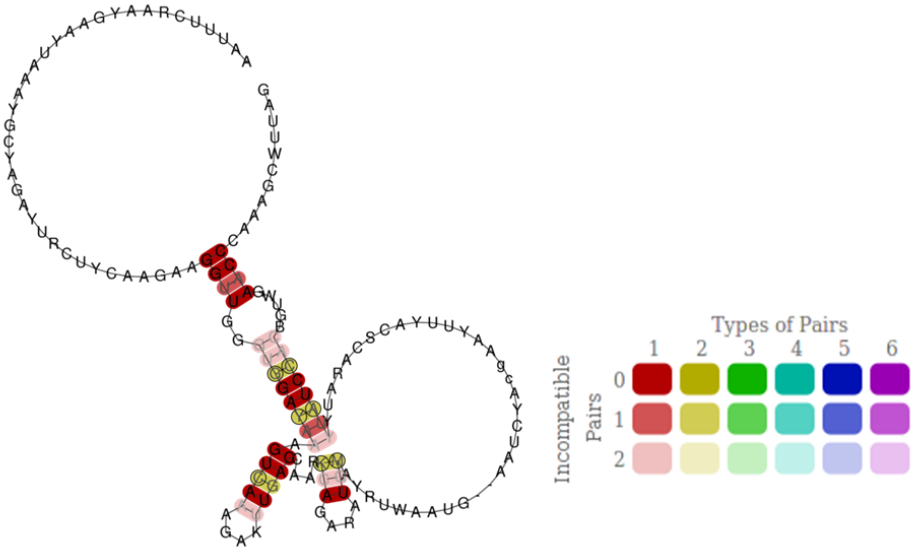


Fig. S2

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 (Cuenca-Bono, García-Molinero et al. 2011): *S. cerevisae, S. paradoxus, S. mikatae, S. kudriavzevii, S. bayanus, S. Castellii, and K. thermotolerans.* The sequence of E2 is more conserved than that of E1 or E3. (**B**) Sequence alignment of *SUS1* E2 across the same species. (**C**) Secondary structure conservation of E2 RNA across the same species, carried out with locARNA (Will, Joshi et al. 2012).

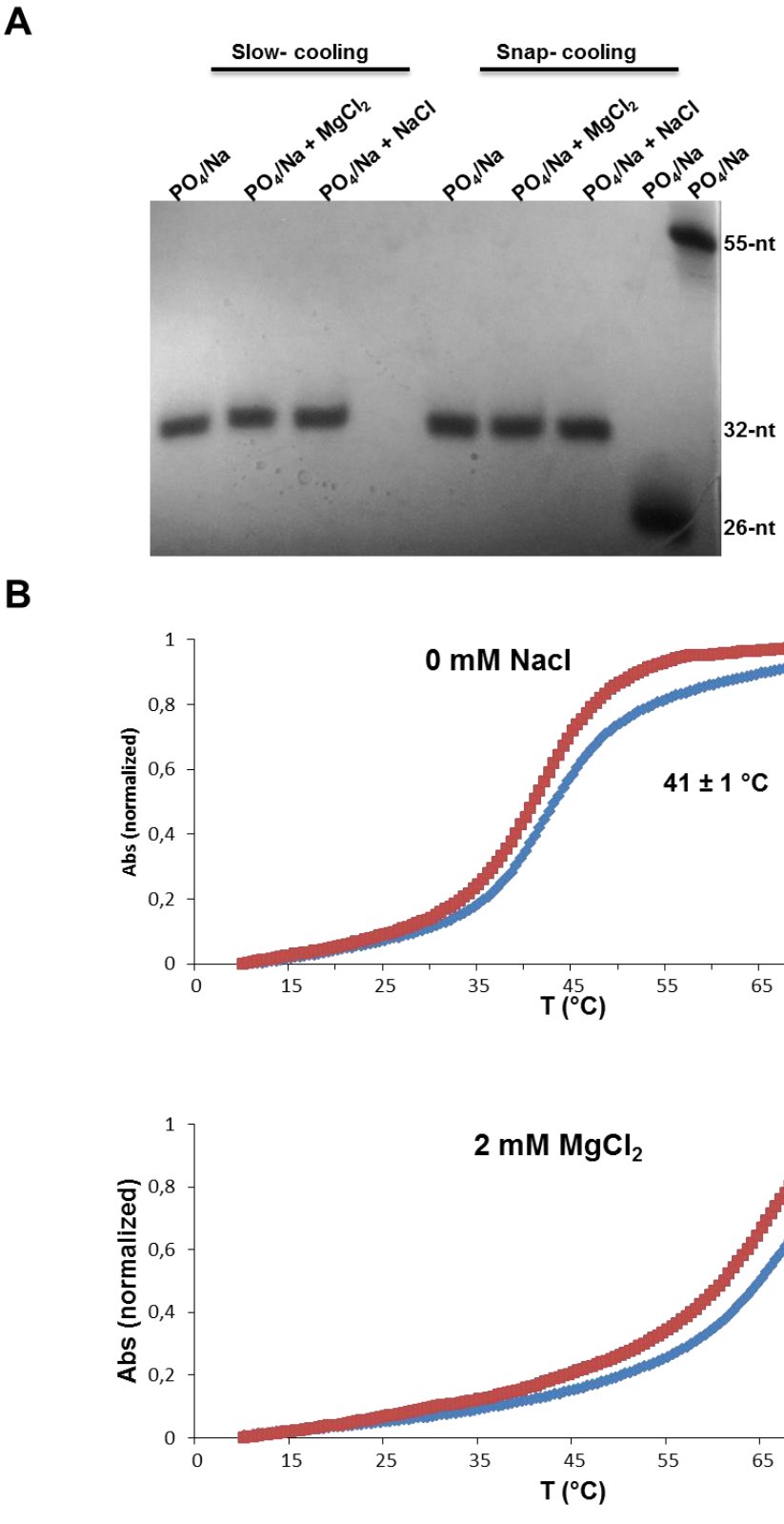

A
Slow- cooling
Snap- cooling
PO4/Na
PO4/Na + MgCl2
PO4/Na + NaCl
PO4/Na
PO4/Na + MgCl2
PO4/Na + NaCl
PO4/Na
PO4/Na
55-nt
32-nt
26-nt
B
0 mM NaCl
41 ± 1 °C
Abs (normalized)
T (°C)
2 mM MgCl2
66 ± 1 °C
Abs (normalized)
T (°C)

Fig. S3

Figure S3. *In vitro* analysis of the structure and stability of the purine-rich P3 RNA hairpin formed by the second exon of the *SUS1* gene of *S. cerevisiae*, represented by the E2s sequence. (**A**) Native gel comparing the electrophoretic mobility of 8.9 μM 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 $MgCl_2$. The last two lanes contained 26-nt and 55-nt RNA hairpin controls dissolved in 10 mM sodium phosphate (pH 6.0) and 0.1 mM EDTA. The E2s samples were previously heated to 95 ^{o}C and either cooled-down slowly (annealing) or rapidly (snap-cooling). (**B**) 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 $MgCl_2$. The average melting temperatures measured for E2s under each ionic condition are indicated in the graphs.

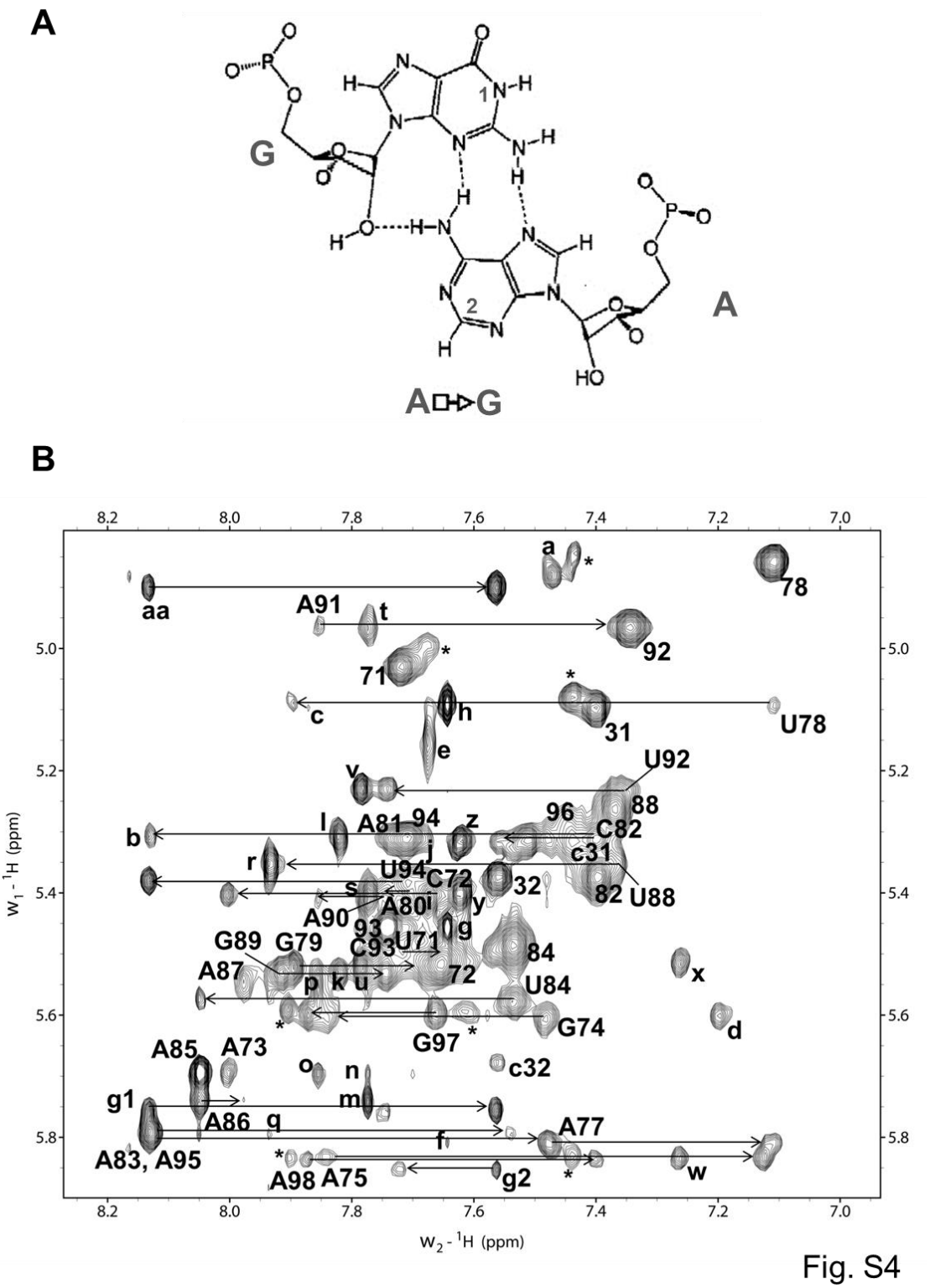


Figure S4. NMR spectroscopy analysis of the purine-rich P3 RNA hairpin formed by the second exon of the *SUS1* gene of *S. cerevisiae*, represented by the E2s sequence. (**A**) Structure of a sheared (*trans* Hoogsten-sugar edge) G:A pair. (**B**) Assignment of the H2/H6/H8-H1'/H5 region of the E2s NOESY spectrum (D_2O, 250 ms mixing time, 25 ^oC). Intraresidue H1´-H6/H8 cross-peaks are labeled with residue name and number, intraresidue H5-H6 crosspeaks are labeled 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 $A_{73}G_{74}$:$A_{95}C_{96}$ internal loop nucleotides likely caused broadening of some G74 and C96 resonances as well as the doubling of several resonances of the lower g_1-C_{72}:G_{97}-c_{32} stem (extra resonances are labeled with asterisks), and G76 H1' was also broadened. All of the assignments were supported by analyses of NOESY and TOCSY spectra at 36 ^{o}C as well as by studies of two-dimensional HCCH TOCSY and ^{1}H-^{13}C HSQC data obtained from an ^{15}N,^{13}C-labeled E2s sample in the same temperature and solution conditions.

A

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

B

E2-WT	100%
E2-muts	100%
E2-mut1	21.42%
E2-mut2	91.3%
E2-mutr	89.13%

Fig. S5

Figure S5. Impact of *SUS1* E2 mutations on RNA and protein sequence. (**A**) Alignment (top) and sequence identity (bottom) of wild-type and mutant E2 RNA sequences. The sequence identities are expressed relative to wild-type. (**B**) Alignment (left) and sequence identity (right) of the amino acid sequences coded by wild-type and mutant E2 RNAs. The sequence identities are expressed relative to wild-type. None of the E2 mutants introduced stop codons in the RNA sequence, and no significant

variations of codon usage were detected with the Genscript Rare Codon Analysis Tool (https://www.genscript.com/tools/rare-codon-analysis) (data not shown).

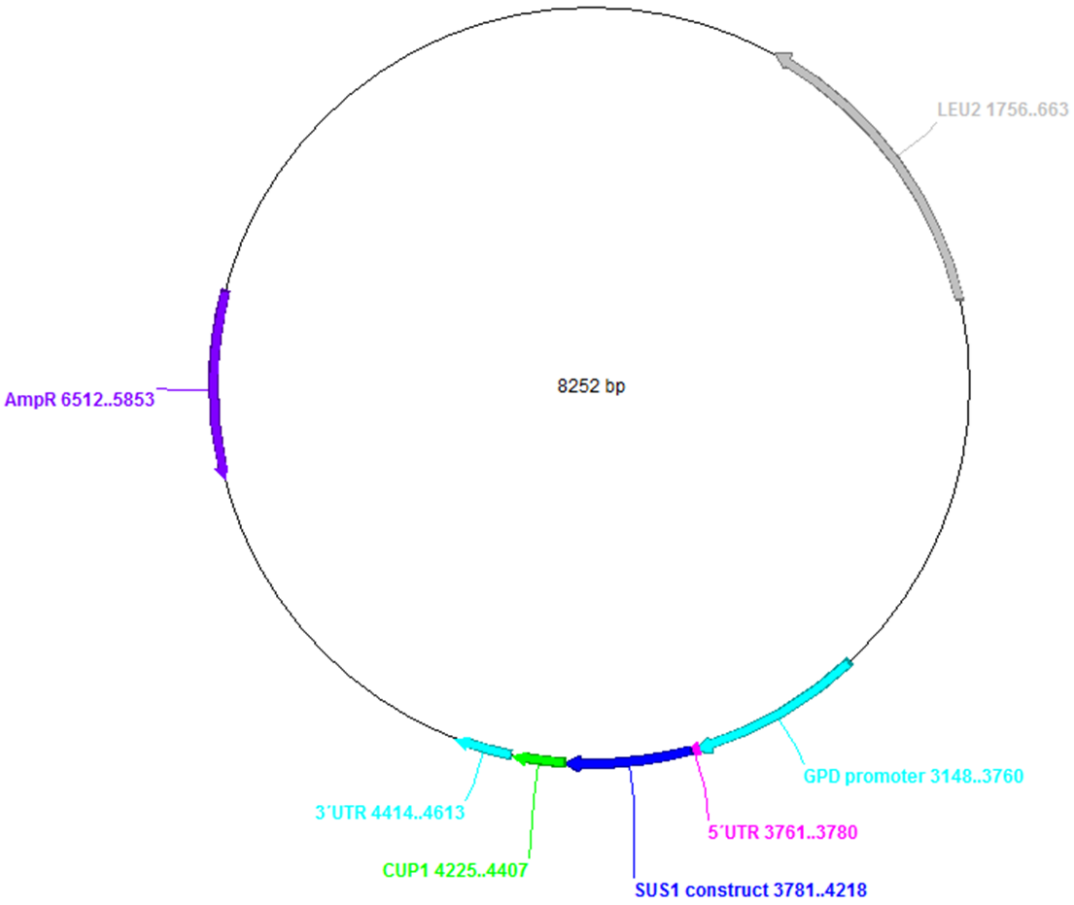


Fig. S6

Figure S6. 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.

E2-mut1 Forward	GAAGGTTGGGTGGATGAACATAAATGAATCCACCGTAGAACCCAAAGC ATTAGGTATGT
E2-mut1 Reverse	TCATCCACCCAACCTTCTTGAAGTAGTCTGGCTTTTGTTCGTTTGAAATT CTAATA
E2-mut2 Forward	CCAAATCAGAGATGAACATAAATGAATCTACGAACTTTACCCAAATTTT ATCCtCCGTAGAACCC
E2-mut2 Reverse	GTTCATCTCTGATTTGGTCAAATCTTTGACTTTATggAggCAACCTTCTTGA AGTAG
E2-mut2r Forward	GATTTGACCAAATCAGAGATGAACATAAATGAATCTACGAACTTTACCC AAATTTTATggtggGTAGAACCCAAAGC
E2-mut2r Reverse	CATCTCTGATTTGGTCAAATCTTTGACTTTATggtggCAACCTTCTTGAAG
Extreme Forward 5´ primer for all constructs and semi-qPCR	AAAAAAaagcttCAATTCTGGCCTTCACTCCAATGACTATGGATACTGC
Extreme Reverse 3´primer for all constructs and semi-qPCR	AAAAAActgcagTTGTGTATCTACAATCTC
Semi-qPCR of E2-mut2 and E2-mut2r	AAAAAActgcagTTCTCCGAATTAATTAACTTCC

SCR1 Forward	AACCGTCTTTCCTCCGTCGTAA
SCR1 Reverse	AGAACTACCTTGCCGCACCA
SHAPE cassette Forward	TAATACGACTCACTATAggccttcgggccaaAATTTCAAACGAACTAAAAGCC
SHAPE cassette Reverse	gaaccggaccgaagcccgatttggatccggcgaaccggatcgaCTAATGCTTTGGGTTCTACGG
VIC-labelled SHAPE	GAACCGGACCGAAGCCCG
NED-labelled SHAPE	GAACCGGACCGAAGCCCG
E2c Forward	AAAGGATCCATTTCAAACGAACTAAAAGC
E2c Reverse	AAAGGATCCTAATGCTTTGGGTTCTACGG

CONCLUSIONS

Overall, the results gathered in this thesis project indicate that the RNA structures formed by the second intron and the middle exon of the *SUS1* gene of *Saccharomyces cerevisiae* are relevant for splicing and also influence other Sus1 expression processes. Specific conclusions can be summarized as follows:

1. The RNA of the downstream intron (I2) of *SUS1* forms a weakly-stable hairpin structure in solution containing branch site nucleotides in its apical loop, and 3' splice site nucleotides after the stem terminus.
2. All four mutants containing altered I2 RNA structures accumulate unspliced *SUS1* pre-mRNA and/or induce distorted levels of fully spliced mRNA relative to wild-type.
3. The I2 mutants that inhibit splicing have significantly impaired *SUS1* expression.
4. Sus1 functions in histone deubiquitination and mRNA export are affected in the I2 mutants that inhibit splicing.
5. The RNA of the middle exon (E2) of *SUS1* forms a three-way junction structure that belongs to topological family C of this type of intersections.
6. The P1 double helical stem of this intersection is conserved across the seven yeast species containing two introns in the *SUS1* gene
7. The P3 hairpin emerging from the junction exhibits significant thermal stability and is closed by an unusually structured purine-rich loop.
8. This purine-rich loop contains two consecutive sheared G:A base pairs and is structurally related to the substrate loop of the *Neurospora* VS ribozyme.
9. Synonymous mutation of E2 impacts *SUS1* splicing and affects Sus1 expression.
10. A mutant removing the three-way junction of E2 while conserving the purine-rich P3 loop had decreased levels of fully-spliced *SUS1* mRNA and poorer Sus1 expression.
11. Disruption of the conserved P1 stem decreases Sus1 expression, while compensatory mutations restoring this stem recover expression to wild-type levels.
12. Mutations affecting the conserved P1 stem increase the levels of partially-spliced *SUS1* RNA transcripts.

13. Quantification of spliced and unspliced RNA species in the different E2 mutants suggests that the structure of E2 may be involved in other processes of *SUS1* RNA biogenesis in addition to splicing.
14. The findings described in this thesis have opened new questions regarding the structure and function of the middle exon of the *SUS1* gene that are currently under investigation.