A consensus motif in *ASH1* and further transcripts unifies several RNA motifs required for interaction with the She2p/She3p transport machinery and mRNA localization in yeast

Markus Seiler[1#*], Annika Niedner[2#], Simone Heber[2,3], Michael Feldbrügge[4], Ralf-Peter Jansen[5], Dierk Niessing[2,3*], Kathi Zarnack[1*]

[1]Buchmann Institute for Molecular Life Sciences (BMLS), Goethe University Frankfurt, Max-von-Laue-Str. 15, 60438 Frankfurt a. M., Germany

[2]Institute of Structural Biology, Helmholtz Zentrum München, Ingolstädter Landstrasse. 1, 85764 Neuherberg, Germany

[3]Institute of Pharmaceutical Biotechnology, Ulm University, James Franck Ring N27, 89081 Ulm, Germany

[4]Institute of Microbiology, Heinrich-Heine University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany

[5]Interfaculty Institute of Biochemistry, Tübingen University, Auf der Morgenstelle 34, 72076 Tübingen, Germany

[#] Equal contribution

[*] Corresponding authors

Keywords: She2p, She3p, localization machinery, RNA localization, RNA zip-code, modeling, consensus motif, surface plasmon resonance, computational screening

Abstract

Asymmetric localization of the *ASH1* transcript is a central step in the regulation of mating type switching in *Saccharomyces cerevisae* and a paradigm for localized mRNAs specifically recognized by the She2p/3p transport machinery. Four RNA elements in *ASH1* are known to mediate bud localization, but it remained unclear what a consensus motif of all four She2p/3p recognition sites (SRS) might look like. In this study, we derive the SRS consensus motif, which is characterized by an L-shaped, double-stranded RNA conformation with two cytosine-containing sequence motifs at conserved positions and a central asymmetric bulge. It is noteworthy that the spatial arrangement of these features can be retained, despite their altered order within the nucleotide sequence. We termed these variations "configurations" and confirmed in biochemical studies that SRS function in *ASH1* elements is preserved when changing configurations. We tested other known and predicted SRS elements in target transcripts of She2p/3p by biochemical and biophysical studies and confirmed the conserved SRS function even in switched configuration. Consistently, we present the *SRS instance collection*, which includes further predictions of SRS instances of transcript targets of the She2p/3p translocation pathway.

42 **Introduction**

43 The active intracellular transport of effector molecules contributes to the high degree of
44 organization in life and asymmetric mRNA distribution is one way to ensure correct spatiotemporal
45 expression of encoded proteins[1]. In recent years, it has been studied in various polarized cells, in *D.*
46 *melanogaster* and *Xenopus laevis* oocytes and embryos[2–4], and also in mammalian fibroblasts[5] and neurons[6].
47 mRNA transport is involved in embryonic patterning, cellular polarity, and asymmetric cell division.
48 Failure of mRNAs to be properly transferred has been shown to cause cellular dysfunction and disease[7]. In
49 these pathways, mRNAs are recognized by the cellular translocation machinery via *cis*-acting elements in
50 their sequences, called zip-codes or localization elements that may be determined by short RNA sequence
51 motifs, conformational folds[8–10] and spatial restraints between these features[11]. These zip-codes are
52 recognized by RNA-binding proteins that target the mRNAs to their final destination[12]. Within the mRNAs,
53 the zip-codes often reside in the 3' untranslated regions (3' UTRs) of mRNAs, but also in 5' UTRs and
54 coding regions[13–18].

55 Budding yeast (*S. cerevisiae*) has evolved as a model organism for mechanistic analyses on specific
56 targeting of mRNAs to the bud of emerging daughter cells[19,20]. The current model of how the RNA transport
57 machinery is built suggests a step-wise assembly of its core compounds: She2p interacts specifically with
58 its target mRNA and recruits the adapter protein She3p together with the type V myosin Myo4p[21–24]. This
59 mRNA ribonucleoprotein (mRNP) complex is then transported along the actin cytoskeleton. The paradigm,
60 the yeast *ASH1* mRNA, is localized to the bud tip enabling the asymmetric distribution of its encoded
61 transcription factor Ash1p into the daughter cell nucleus where it represses mating-type switching[25–28]. To
62 date, over twenty transcripts have been reported to be targeted via the She2p/3p machinery. Besides the
63 well-characterized She2p/3p target mRNAs *ASH1*, *EAR1*, and *WSC2*, further candidates were identified
64 using combined RNA immunoprecipitation and microarray techniques[24,29]. In some target mRNAs, the
65 regions containing putative zip-codes had been narrowed down by yeast three-hybrid and *in vivo*
66 localization studies[30]. However, besides of *ASH1*, precise positions of the She2p/3p recognition sites (SRS)
67 remained unknown in most cases. Within the *ASH1* mRNA, four localization elements with intrinsic SRS

68 function were mapped in accurate fashion[13,14,19]. Three of these (E1, E2A and E2B) were shown to reside
69 within the coding region, whereas E3 overlapped the stop codon and extended into the 3' UTR
70 (**Supplementary Fig. 1A, Tab. 1**). Each *ASH1* SRS element was sufficient for triggering the assembly of
71 the She2p/3p complex and for localizing a reporter RNA to the bud and was predicted to exhibit RNA folds
72 with extended stem-loop structures. An *in-vivo* screen with mutated *ASH1* fragments supported the notion
73 that a three-dimensional fold was involved in zip-code function. The fold was described as a stem with
74 conserved nucleotides on either side: a singular cytosine and a CGA triplet, with a distance of six
75 nucleotides between the cytosines. Searching for similar RNA folds identified functional SRS elements in
76 *IST2* and *EAR1*[10]. However, these folds lacked an apparent consistency with some of the four *ASH1* SRS
77 elements and a coherent model for specific SRS recognition by the She2p/3p transport machinery has
78 remained elusive. Recently, the interaction interface between E3 and She2p/3p was resolved using X-ray
79 crystallography: a symmetric tetramer of She2p exhibited two interfaces interacting with E3 on opposite
80 sides of the complex[8]. As suggested before[31], the affinity was increased by She3p in a synergistic manner.
81 Intriguingly, the ternary complex formation was accompanied by an induced fit in the SRS element: For
82 the unbound E3 RNA, SHAPE experiments suggested an elongated conformation[8]. When interacting with
83 She2p/3p, E3 displayed a sharp turn in one of its bulge regions, resulting in a kinked global conformation
84 that was also confirmed by NMR structures.

85 In this study, we define a coherent SRS consensus motif that unifies the features of E1, E2A, E2B
86 and E3 in *ASH1* and that also classifies SRS elements of further bud-localizing transcripts (*WSC2*, *EAR1,*
87 *IST2*). It exhibits a kinked stem-asymmetric bulge-stem-bulge structure, with the recurrent singular cytosine
88 in the asymmetric bulge and the CGA motif in a second bulge. One characteristic of the consensus motif is
89 its arrangement of features, conserved in the overall spatial arrangement but present in two variations
90 ("configurations") in linear sequences. Electrophoretic mobility shift assays and surface plasmon resonance
91 studies showed that E3 and other SRS elements could be switched between configurations without
92 impairing She2/3p binding. Using computational screening, we identified potential SRS instances in further
93 twelve transcripts known to interact with She2/3p. Finally, we introduce the *SRS instance collection*,

94 describing in detail predicted SRS elements and their positions in further RNA targets of the She2p/3p
95 translocation pathway.
96
97 **Results**
98 **Bioinformatics predictions reveal a consensus capturing all She2/3p recognition sites of *ASH1***
99 We wondered if a general concept may exist to describe all four localization elements of the *ASH1*
100 mRNA that act as She2p/3p recognition sites (SRS). For this study, we could rely on a recently published
101 crystal structure that showed in detail how the *ASH1* localization element E3 interacts in a kinked stem
102 structure with She2p/3p[8]. Of note, this interaction involved a molecularly induced fit upon binding that
103 resulted in an asymmetric bulge in the RNA stem with one of its functional cytosines deeply buried inside
104 the protein interface. The asymmetric bulge of E3 displayed a large stretch ("Lbulge") and a small one
105 ("Sbulge"), with a length ratio of 6 nt : 3 nt (**Fig. 1A, right**) or 4 nt : 1 nt, respectively, if considering
106 atypical base pairing inside the bulge as further base pairs. We hypothesized that the ability to form such
107 an asymmetric bulge could be a pre-requisite of the E3 conformation, whose composition would require
108 different lengths to be kinked. In a next step, we performed a systematic survey of such arrangements in
109 further SRS elements of *ASH1*. However, common RNA conformation predictors do not consider precise
110 structural rearrangements due to RNA-protein interaction very well. In a step-wise strategy
111 (**Supplementary Fig. 2, Material and Methods),** we first evaluated which of the commonly used
112 prediction tools would predict the best approximation of the kinked E3 conformation in complex with
113 She2p/She3p[8] just from the single E3 nucleotide sequence. Of note, we found *Kinefold*[32] to predict the
114 asymmetric bulge in close approximation (5 nt : 2 nt) that could even be closer approximated when
115 considering atypical base pairing as found in the asymmetric bulge of the crystal structure, resulting in (6
116 nt : 3 nt) (**Fig.1**, **Supplementary Fig. 3A**).
117 For E2A, we found the predicted conformation with an asymmetric bulge with exactly the same
118 length ratio as found in E3 (5 nt : 2 nt) using *Kinefold* (**Fig. 1B left)**. Of note, after 180° rotation, the
119 predicted E2A conformation harmonized perfectly with the E3 conformation (**Fig. 1A right, 1B right**).

120 Furthermore, this 180° rotation caused the previously diverging sequence motifs of E3 to fit now to the
121 single cytosine and CGA motifs of E2A. Surprisingly, sequence comparison revealed 89% sequence
122 identity between E3 and E2A in the asymmetric bulge and flanking regions (**Supplementary Fig. 1B**). For
123 E1, the 180° rotation of its predicted conformation resulted in an arrangement comparable with E3 (**Fig.**
124 **1D**), but with different length proportions of the asymmetric bulge and a lower degree of sequence identity.
125 These findings suggested one blueprint, with a common overall spatial arrangement of its
126 (conformational and sequence) features in E3, E2A and E1 that we referred to as "SRS consensus motif"
127 (**Fig. 1E,F**). Of note, despite of this conserved spatial arrangement of features, the linear order of features
128 was interchangeable in the nucleotide sequences (**Fig. 1F**). This was due to the fact that the conformations
129 of E3, E2A and E1 are harmonized by a 180 degree rotational step of E3. Subsequently, the 5' start of E3
130 was at the opposite site of the SRS consensus motif than for E2A and E1, resulting in different feature
131 orders in the nucleotide sequence (**Fig. 1E**). We termed these variations "configurations", with "config-I"
132 observed in E3 and "config-II" in E2A and E1. Further, we could harmonize the annotation of the sequence
133 motifs in E3 by labeling the single cytosine motif in the asymmetric bulge as C_1 and the cytosine in the
134 CGA motif as C_2 (**Fig. 1A, right, 1E**), analogously to E2A and E1. For E2B, at a first glance, the
135 conformational prediction of E2B displayed the hypothesized "Lbulge" even shorter than the "Sbulge" (**Fig.**
136 **1C, left**). Remarkably, Kinefold predicted an alternative conformation with an asymmetric bulge in
137 agreement with the SRS consensus motif (6 nt : 3 nt), but with a higher content of energy (calculated free
138 energy = -10.3 kcal/mol versus -7.4 kcal/mol), which could be rearranged by sliding between sequence
139 repeats in one inverted repeat (**Fig. 1C, right**). Altogether, we had derived a unified model of the SRS
140 consensus motif (see **Info-Box**). In a subsequent step, we could confirm that the model was sufficient to
141 summarize previous functional studies with mutationally-derived decreases in RNA localization in all four
142 *ASH1* SRS elements[10,13,14,25] (**Material and Methods)**. For an overview, the distinct modeling steps
143 (initially characterized elements, Kinefold prediction, rotation, and mutual feature alignments) are
144 displayed in (**Fig. 2**).
145

146 **The two configurations are interchangeable in E3 without loss-of-function *in-vitro***

147 One unexpected characteristic of the SRS consensus motif was the spatial conservation of its
148 features in its overall conformation while the order of features in the nucleotide sequences was found to be
149 swapped for the different configurations. To validate if we had grasped the essence of the consensus, we
150 investigated if an artificial switch between the different sequence orders (and the configurations) would
151 conserve SRS function. For this purpose, we used the well-characterized construct $E3_{51nt}$wt[8] and switched
152 its feature order in the nucleotide sequence (config-I) to design a config-II construct ($E3_{51nt}$switch, **Fig. 3A**).
153 Based on the observation that the E3 crystal structure exhibited an interaction of She2p/3p with C_2G but
154 not with C_2GA, and for comparability reasons with previous studies, we limited the construct design to
155 C_2G. Using an electrophoretic mobility shift assay (EMSA), we examined $E3_{51nt}$switch for interaction with
156 She2p and She3p in a titration approach. We used the fusion protein She2p-She3p, which comprised both,
157 She2p and a She3p fragment and was demonstrated to be sufficient to mediate the synergistic binding of
158 both proteins[8]. For the wildtype construct $E3_{51nt}$wt, we observed an interaction strength at the expected
159 nanomolar protein concentration (**Fig. 3A**)[8]. As a control, we tested the construct $E3_{51nt}$M6 characterized
160 by reduced interaction[8] due to a C_{1813}G substitution of C_2. In agreement of our model, $E3_{51nt}$wt and
161 $E3_{51nt}$switch, interacted with She2p-She3p at comparable nanomolar protein concentration (**Fig. 3A**). These
162 similar binding affinities contrasted with $E3_{51nt}$M6, which displayed severely impaired interaction.

163 Since the $E3_{51nt}$M6 mutant showed impaired but still considerable binding to She2p-She3p-fusion
164 protein, we used the full-length proteins to assess the binding features of E3 elements. Incubation with
165 constant She3p concentration of 0.56 μM and titration of She2p (0.02 - 5 μM) confirmed that both
166 constructs, $E3_{51nt}$wt and $E3_{51nt}$switch, supported binding even at the lowest She2p concentration used in this
167 experiment (0.02 μM). In contrast, binding of the full-length proteins to $E3_{51nt}$M6 was abolished at
168 concentrations up to 5 μM (**Supplementary Fig. 4B**). The same was true for a reciprocal approach, in
169 which increasing concentrations of She3p (0.02 - 5 μM) were tested in the presence of high amounts of
170 She2p (5 μM) (**Supplementary Fig. 4C**). Moreover, both experiments showed that She3p efficiently binds
171 E3 in the absence of She2p. Surprisingly, this binding was sensitive to mutation of C_2. In conclusion, these

172 studies successfully confirmed that E3 can be switched from config-I to config-II without compromising
173 its functionality to interact with She2p/3p.

174 When we analogously examined E2A and E1, we reversed the direction of switching (now from
175 config-II to config-I), resulting in the constructs $E1_{27nt}$switch and $E2A_{32nt}$switch (**Supplementary Fig. 5**).
176 Again, in EMSA validations, we could demonstrate that both constructs largely exhibited the binding
177 properties to She2p-She3p-fusion. In summary, the "config-switch" approach supported our model, which
178 suggested that SRS features could be rearranged in the RNA sequence while retaining function. We could
179 demonstrate that E1, E2A, and E3 recognized She2p/3p regardless of configuration, demonstrating the
180 essential characteristics of RNA-She2p/3p recognition were reflected in the SRS consensus motif. As a
181 conclusion, these findings indicate that SRS function requires very few conserved bases in defined positions
182 of a stem-loop, which must accommodate a bulge that is flexible enough to open up upon binding by She2p-
183 She3p to adopt a kinked L-shaped conformation in the co-complex.

184 In a fine-grained approach, we used the analytical power of surface plasmon resonance (SPR), to
185 investigate the influence of the functional cytosines C_1 and C_2 on SRS function. Previous studies had
186 determined a preferred distance of six nucleotides between these cytosines[10]. Using the functionally
187 minimal[8] construct $E3_{29nt}$, we introduced varying distances between C_1 and C_2 and performed titration series
188 in a surface plasmon resonance experiment with a She2p-She3p-fusion construct (125 nM) (**Fig. 3B**). As a
189 result, for the wildtype construct $E3_{29nt}$wt, we observed a binding affinity of approx. K_D = 0.89 nM (**Fig.
190 3B, Supplementary Fig. 6**). In contrast, constructs with C_1-C_2 distances of five and seven nucleotides led
191 to a reduction in binding of at least one order of magnitude (**Fig. 3B**, **Supplementary Tab. 1**), confirming
192 the findings of 6 nt as optimal C_1-C_2 distance.

193

194 **A computational algorithm to screen for the consensus in known She2p/3p substrates**

195 We performed computational screening (details, see **Material and Methods**) to identify further
196 instances of the SRS consensus motif in further target sequences of the She2p/3p transport machinery. For
197 this purpose, we screened a list of 24 transcripts reported to show *in-vitro* binding by She2p/3p or bud

198 localization *in-vivo*[24,29,30]. The identified sequence regions represented potential SRS instances and were
199 evaluated for their potential to form the correct conformation of one of the configurations using *Kinefold*.
200 We intended to extend this conformational evaluation by integrating data describing the degree of unpaired
201 state of nucleotides (obtained in a remarkable genome-wide study using parallel analysis of RNA structure
202 (PARS)[33]). This strategy was promising, but for our purposes the data were not detailed enough, due to
203 competing conformations that could not be differentiated in the averaged values per nucleotide position,
204 even for some of the strong *ASH1* SRS elements (**Material and Methods, Supplemental Fig. 7**).

205 For the three transcripts *EAR1, IST2* and *WSC2*, their SRS elements had been identified already,
206 with their sequence regions narrowed down to fragments[10,30]. Now, with knowledge of the SRS consensus
207 motif, these SRS elements could be classified as a config II arrangements (**Fig. 4C, Supplementary Fig.**
208 **8**). Remarkably, for *EAR1* and *IST2*, the asymmetric bulge was even more eccentric than observed in *ASH1*
209 (*EAR1:* 7 nt : 0 nt; *IST2*: 9 nt : 1 nt). In addition, we identified an additional upstream SRS instance in each
210 transcript, *EAR1* and *IST2* (**Tab. 2**). In contrast, for *WCS2*-C, the initially identified SRS fragment turned
211 out to comprise possibly two SRS elements, a canonical one in the center of this fragment, and a potential
212 SRS instance in the flanking regions, however, with a C_1-C_2 distance of 7 nt (**Supplemental Fig. 8D**).
213 However, we could not classify the second published instance in *WSC2*, *WSC2*-N, in agreement with our
214 model (see **Discussion**). As a first outcome of the screening we had identified seven SRS instances (**Fig.**
215 **3C**). When we studied the seven canonical SRS elements in detail, we noticed a recurring di-adenosine in
216 the L-bulge ("AA_L") as a common feature in all elements (**Fig. 3C**), possibly indicating relevance in SRS
217 function. This was surprising due to the fact that the crystal structure of E3 showed the interaction between
218 She2p/3p and the AA_L nucleotides occurs exclusively via the backbone atoms of the nucleotides, and thus
219 unspecifically[8], with their nucleobases having intramolecular contact within the RNA structure. This
220 observation makes relevance in the context of SRS recognition less likely.

221 As mentioned before, for most known SRS instances (in 23 of 30 instances in our set), functionality
222 had been exclusively studied for large RNA fragments or complete transcripts. For 13 of these transcripts,
223 we could identify now the positions of SRS instances. For 12 transcripts, we found one or two SRS instances

224 in each sequence, and we found four instances in *TPO1* (**Fig. 4A**). Altogether, in 17 of 24 sequences (*ASH1*,
225 *WSC2*, *EAR1*, *IST2* included), we predicted 30 SRS instances, with nine instances (30%) in config-I and 21
226 instances (70%) in config-II, respectively (**Fig. 4A**). Of notice, more than two thirds of the SRS instances
227 contained the AA_L di-nucleotide in the Lbulge (**Fig. 4B)**.
228
229 **Validation of SRS elements in *EAR1*, *IST2*, *KSS1* and *TCB3***
230 As described before, in *IST2* and *EAR1*, sequence fragments with SRS function had been identified
231 already, and these sequences were also identified in our computational screening (termed *EAR1*-SRS2,
232 *IST2*-SRS2; nomenclature see **Tab. 2**). We could classify their functional features, in agreement with the
233 SRS consensus motif, provided that the consensus was valid for these instances. In such a view, the SRS
234 elements would resemble the very first characterization (Olivier et al., 2005), however in the larger context
235 of the SRS consensus motif. To address this aspect, we studied specifically designed SRS constructs,
236 analogous to our previous studies. In accordance with the consensus, the newly designed constructs
237 $EAR1_{26nt}$switch and $IST2_{29nt}$switch were strictly reduced to those regions assumed to be essential for SRS
238 function and also switched in configuration (each to config-I) (**Fig. 5A**). As we could show in EMSA
239 studies, $EAR1_{26nt}$switch was bound with affinities comparable to the unmodified wild-type element
240 $EAR1_{50nt}$wt (**Fig. 5A**). Although for $IST2_{29nt}$switch binding was clearly retained, it was considerably weaker
241 when compared to $IST2_{62nt}$wt. In addition, we tested the predicted *KSS1*-SRS1 (config-II, **Tab. 2**) and
242 *TCB3*-SRS1 (config-I) for which no detailed position in their transcripts had previously been determined.
243 As a result, we observed strong binding *in-vitro* for both, $KSS1_{72nt}$wt and for the minimal construct
244 $KSS1_{29nt}$switch (switched to config-I), confirming *KSS1*-SRS1 as a potential SRS element (**Fig. 5B**). In an
245 analogous manner, we observed for the minimal construct $TCB3_{38nt}$wt (config-I) strong interaction with the
246 fusion protein suggesting *TCB3*-SRS1 as further SRS element (**Fig. 5B**). In summary, we were able to
247 demonstrate the identification and *in-vitro* confirmation of new SRS instances. Furthermore, for some of
248 these instances we were able to prove that the newly designed, size-reduced constructs retain the SRS
249 function even in switched configuration. In conclusion, the fact that our model, derived exclusively from

250 *ASH1* SRS elements, enabled the identification, size reduction, and config-switch of designed SRS
251 constructs in previously unexplored sequences while preserving SRS function confirmed its applicability
252 as an SRS consensus motif.
253
254 **Validation of further features in the SRS of *EAR1***
255 In *EAR1* (*YMR171C*), it had been shown for its SRS element (now termed *EAR1*-SRS2) that
256 substitutions in the sequence motifs decreased function[10]. The substitution of both, C_1 and C_2 (updated
257 nomenclatures) in the constructs "M1" and "M2"[10] had shown severe effects *in-vivo*. To examine the SRS
258 consensus motif and *EAR1*-SRS2 in more detail, we chose a fine-grained approach using the $EAR1_{31nt}$switch
259 construct (switch to config-II) as the base construct for further changes (**Fig. 6A**) and determined its binding
260 affinity using SPR analysis as described before. As result, we obtained an affinity of the original construct
261 $EAR1_{31nt}$switch of K_D = 11 nM (**Fig. 6D, Supplementary Fig. 9**). Remarkably, compared to the E3 minimal
262 element, the affinity with the She2p/3p fusion protein was decreased by one order of magnitude ($E3_{29nt}$wt:
263 K_D = 0.89 nM). When we modified $EAR1_{31nt}$switch in the inverted repeat IR1, by inversions or additionally
264 introduced base pairs, the modified constructs showed affinities in the range of the original construct
265 suggesting no direct impact on SRS function (**Fig. 6D**). Also, when we analyzed the relevance of the di-
266 nucleotide AA_L in the Lbulge by deletions of different size (**Fig. 6C**), we observed binding affinities in the
267 order of magnitude of the basic construct (**Fig. 6D**). The analysis of the functional cytosines C_1 and C_2 was
268 complex since two cytosines (C_{1601} and C_{1602}) precluded an unambiguous assignment of the C_1 position
269 (**Fig. 6B**). The substitution of either cytosine (constructs $C_{1\text{-}1}$U, $C_{1\text{-}2}$U) resulted in a considerable binding
270 decrease suggesting that both positions contribute to She2p/3p binding (**Fig. 6D**). Unexpectedly, the
271 deletion of one of the C_1 positions (ΔC_1) or a simultaneous substitution of both cytosines ($C_{1\text{-}1}$ U & $C_{1\text{-}2}$ U)
272 restored She2p/3p binding (see **Discussion**). In contrast, substituting C_{1590} in the three-nucleotide C_{1590}GA
273 motif (C_2U) returned unequivocal results manifesting in a dramatic loss of binding. Altogether, these results
274 confirmed that neither sequence composition of IR1 nor the AA_L motif in the Lbulge were relevant for
275 She2p/3p binding *in vitro*. Alterations of the C_1-C_2 distance were tolerated in some SRS contexts but not

276 others, with a strong bias for a C_1-C_2 distance of six nucleotides. We therefore kept a C_1-C_2 distance
277 constraint of six nucleotides for subsequent computational SRS screening.
278
279 **A collection of predicted *SRS* instances in known target transcripts of She2p/3p**
280 Computational screening for the SRS consensus motif had identified SRS instances in 17 of 24
281 sequences (**Fig. 4A**). After successful *in-vitro* validation of several of the new SRS instances (in *EAR1*,
282 *IST2*, *KSS1*, *TCB3*), we finally compiled a collection of these screening results to present to the community
283 (*SRS instance collection*, **Tab. 2**). This collection contains information on 30 SRS instances in 17 transcripts
284 that have been reported to show She2p/3p binding *in-vitro* or bud localization *in-vivo* [24,29,30]. In addition, a
285 standardized nomenclature for SRS elements is introduced and further information is provided on positional
286 data, classification and other details of SRS instances, several of them are promising new candidates for
287 future experimental investigations.

288 **Discussion**

289 In this work, we have derived the unified model of a three-dimensional RNA motif that functions
290 as She2p/3p recognition motif (SRS) with the She2p/She3p protein transport machinery, a prerequisite for
291 active transport of transcripts to the yeast bud. We used a computer-aided and conceptual modeling
292 approach that based on the hypothesis that an asymmetric bulge in the RNA conformation might be a
293 prerequisite of SRS function. This approach led to the identification of a common three-dimensional
294 arrangement of SRS functional features (conformational features and sequence motifs) in the *ASH1*
295 localization elements E3, E2A and E1 (and with limitations also for E2B) in a blueprint, thus demonstrating
296 the long-missing consensus motif of all four SRS elements. Unexpectedly, although this arrangement of
297 features was conserved in its overall three-dimensional fold, , there were two variations with respect to the
298 linear arrangement in the nucleotide sequences that we termed "configurations" ("config-I" in E3, "config-
299 II" in E2A, E1 and E2B). The consensus model proved to be correct based on our consistent results from
300 electric mobility shift assays, in which we analyzed the SRS function of RNA-constructs with re-designed
301 feature arrangement. According to our model, this rearrangement of the features affected their linear order
302 in the nucleotide sequences, but allowed the preservation of their overall three-dimensional arrangement,
303 hence representing a switch between their original configuration and the respective counterpart
304 configuration ("config-switch"). We were able to consistently demonstrate that the config-switch retained
305 SRS function in all experiments we performed indicating that the consensus motif comprises the main
306 requirements of SRS function. The fact that also size-reduced constructs proved to keep SRS function was
307 a further hint that we had grasped the essential features of SRS function. It can be described in all *ASH1*
308 SRS elements as follows: an asymmetric bulge, flanked by inverted repeats, forms a kinked, L-shaped RNA
309 conformation. The larger side of the bulge contains the single cytosine motif C_1 inside of the first three
310 positions. The C_2GA motif resides on the opposite strand of C_1, separated by an inverted repeat in
311 downstream direction on this strand. The two configurations are marked by the two different 5' entry sites
312 into this kinked conformation. These findings were further confirmed, when we compared previous
313 thorough mutagenesis studies of the *ASH1* SRS elements and could assign these results well according to

314 their relevance in the SRS consensus model. Further, by studying the binding kinetics of $E3_{29nt}$wt in surface
315 plasmon resonance analysis, we could confirm the optimal distance between C_1 and C_2 with six nucleotides
316 (ca. 28 Å) as it was suggested before[10]. It is noteworthy that the existence of two configurations with double-
317 stranded folds that can be converted into each other by swapping their features in the nucleotide sequence
318 finds an analogy in studies on the beta-actin RNA zip code, where the order of its single-stranded 5' and 3'
319 elements is interchangeable within the target transcripts bound to ZBP1[11].

320 Due to its universal character to describe all four *ASH1* SRS elements, the exploitation of the SRS
321 consensus motif went beyond previously computational screens. Hence, we created a computational screen
322 that searched for both configurations, followed by a conformational check of the overall RNA fold
323 performed by *Kinefold*. For some of the further transcripts that had been found to localize at the bud tip of
324 yeast cells, sequence fragments had already been narrowed down with retained SRS function. In this set of
325 sequence fragments, we could classify SRS elements of *EAR1*, *IST2* and *WSC2* as further instances of the
326 SRS consensus motif. *EAR1* was first identified as localized in the bud in the study of [29] with its SRS
327 element characterized in further studies[10] and classified as a config-II SRS element in this work (*EAR1*-
328 SRS2, **Tab. 2**). For the minimal construct $EAR1_{31nt}$switch, using SRP analysis, we could confirm previous
329 observations that the nucleotide sequences of the inverted repeats did not contribute to SRS function[8,13]. In
330 further consistence with previous *in-vivo* studies[10], the loss of C_2 showed and individual substitutions of
331 each of the two potential C_1 positions (in C_1-C_2-distance of five or six nucleotides) affected SRS function.
332 Unexpectedly, a single deletion of C_{1601} or a simultaneous substitution of C_{1601} and C_{1602}, both restored SRS
333 function. When studying the potential di-nucleotide AA_L sequence motif by deletion experiments we could
334 not detect any impact on SRS function, even if the Lbulge region was almost completely deleted. So far,
335 we can neither explain the apparent discrepancy in the C_1 mutation effects in the *EAR1* constructs nor the
336 retained SRS function with deleted asymmetric bulge. It may reflect both complex sequence composition
337 and inherent redundancy in this particular SRS instance.

338 So far, we had observed only the occurrence of one config-I (E3) whereas we had identified six
339 instances of config-II (**Fig. 4C**). This dominance of config-II SRS elements continued when we applied the

340 computational screen to our complete sequence set and predicted 30 SRS instances in 17 transcripts, with
341 22 of them (73%) exhibiting config-II. So far, the reason for the unequal frequency of the two configurations
342 remains unclear. From this outcome, we used EMSA approaches to analyze two instances of each subset:
343 SRS instances with relatively well-known positions in their transcripts or with unknown SRS positions. For
344 the first subset, we could confirm SRS function in both SRS elements *EAR1*-SRS2 and *IST2*-SRS2[10]. For
345 the second subset, we could verify *KSS1*-SRS1 and *TCB3*-SRS1 *in-vitro.* Remarkably, when we switched
346 configuration in *EAR1*-SRS2, *IST2*-SRS2 and *KSS1*-SRS1 (config-II to config-I), we could demonstrate
347 again conserved SRS function for <u>all</u> switches, strongly supporting the SRS consensus model. Of note, as
348 a first successful validation *in-vivo*, we had been able to identify and confirm the nucleotide positions of a
349 SRS element in the *CLB2* transcript[34]: In the transcript, only a few but functionally crucial positions of the
350 SRS instance *CLB2*-SRS1 (config-II, 1111 nt - 1145 nt) were altered in non-canonical mutations to reveal
351 a significant loss-of-function in bud-localization. It is also interesting to note that the regulatory effect of
352 well-defined asymmetric bulges seems not to be limited to SRS function. As described in a study[35], the
353 structure of the pre-mRNA autoregulatory complex of the L30 ribosomal protein in yeast displays an
354 asymmetric bulge (length-ratio 5 nt : 2 nt) with a cytosine positioned in an extension of the Lbulge, and
355 separated from another cytosine (in a further bulge) by an IR in a distance of 6 nt. Remarkably, it is this
356 asymmetric bulge region that is part of a mutual induced fit, suggesting analogies with the observed induced
357 fit mechanisms in the crystal structure of E3 and She2p/3p.

358 Further aspects of SRS characteristics remain to be investigated. For example, a further extension
359 of the description of the SRS consensus motif is thinkable: It is reasonable to ask if the SRS function would
360 be conserved if Lbulge and Sbulge were joined together without IR2 as intervening feature. Although
361 experimental analysis of such a setting is beyond the scope of this study, we would like to put to discussion
362 such a theoretical construct, to possibly take a step forward to complete the picture (**Supplementary Fig.**
363 **10A**). The SRS element *WSC2*-N[36] that could be not classified, might be such an example (**Supplementary**
364 **Figure 10B, left**). Such an extended perspective might have further implications also for the understanding
365 of some of the SRP results of *EAR1* described before. Both elements, *EAR1*-SRS2 and *IST2*-SRS2 displayed

366 two CGA motifs with the following consequences: Whereas we considered *EAR1*-SRS2 as config-II with
367 C_2GA residing in the terminal bulge, this terminal bulge also may be considered as a joint of Lbulge and
368 Sbulge, hence offering a new perspective: the cytosine motif in this terminal bulge may be considered as
369 C_1 while its counterpart inside of the asymmetric bulge may be extended and labelled to C_2GA
370 (**Supplementary Figure 10B, right**). In such a view, the functionality of *EAR1*-SRS2 would exhibit
371 redundant features of two superimposed and oppositely rotated SRS elements in the bigger context of an
372 extended SRS consensus motif. This perspective may also explain partially the ambiguous results in the
373 SPR experiments of our study. Secondly, a further extension of the description of the SRS consensus might
374 be formed by considering a range in the restraints of the C_1-C_2 distance. In consistency with previous
375 studies[10] and with our SPR results we focused on the optimal parameters of six nucleotides in this study to
376 minimize false positive matches. However, we could not exclude a minimal functionality for instances with
377 (C_1-C_2 distance $\neq$ 6). An example might be a second, potential SRS instance that we identified in the
378 flanking regions of *WSC2*-SRS2 (*WSC2*-C) but with a C_1-C_2 distance of seven nucleotides (**Supplementary**
379 **Fig 7C**). In such a case, *WSC2*-SRS2 may compensate the suboptimal C_1-C_2 distance of its neighboring
380 SRS element by inducing SRS conformation in a cooperative manner when both SRS instances interact
381 with the binding interfaces on the opposite sites of the She2p tetramer[37].

382 As a second aspect, localization elements that mediate bud localization require the interaction with
383 the She2p/3p protein transport machinery, but further motifs may be required as binding platforms of
384 auxiliary factors. For example, for E2B, we speculated that its energetically optimal conformation had to
385 be converted to an alternative conformation to represent the SRS consensus motif. Even strong SRS
386 elements such as E2A[12] can be subject to major reorganizations, as shown by the PARS data which indicate
387 competing conformations for E2A[33] (**Supplementary Fig. 7**). Therefore, the She2p/3p interaction can be
388 seen as a central step in the zip-code function, but it might be framed by further rearrangements in the
389 chronological order. In such a scenario, transient binding of auxiliary cofactors could stabilize the
390 asymmetric bulge and increase SRS function. AA_L might be such a co-factor binding motif. Despite the
391 fact that AA_L was present in each case of our seven well-characterized, bud-localizing SRS elements,

392 neither the crystal structure[8] nor the results of SPR studies suggest involvement in She2p/3p binding
393 specificity. Future research will be required to clarify the possible function of AA_L *in-vivo*, also with regard
394 to transient interactions. In addition, for E3 and E2A, it was tempting to speculate if their extraordinary
395 sequence identity was a product of convergent evolution, due to further requirements in affecting zip-code
396 function. Maybe we have to face a trade-off between canonical SRS consensus motif and stability of the
397 individual RNA conformation, with the weakening of one aspect may be compensated by strengthening of
398 the other. This would complicate the identification of SRS instances, especially if auxiliary co-factors
399 would alter the probabilities of formation of SRS conformations *in-vivo*. Summarizing, all these aspects
400 stress two aspects: (i) The new SRS consensus motif delivers a conceptual framework of how to classify
401 further aspects of new SRS instances, and (ii) It remains necessary to combine experimental screenings
402 (identification of bud-localizing transcripts, final validation of instances) and computational screenings
403 (prediction of positional information in transcripts to narrow down experimental analysis and suggest
404 crucial positions in mutations studies). Consequently, the *SRS instance collection* contains only such SRS
405 instances where bud localization had been confirmed already, at least at the transcript level.

406 In conclusion, the SRS consensus motif enabled the description of the four SRS elements of *ASH1*
407 in one model, hence providing the long-missing consensus. It further unified the described functional
408 features of the known SRS elements in *IST2*, *EAR1* and *WSC2* and enabled the identification of new SRS
409 elements in further bud-localizing transcripts. The consensus model proved to be consistent with previous
410 mutagenesis studies of these elements and we could demonstrate that a switch between the configurations
411 retained function in SRS constructs of several transcripts.

412 **Materials and Methods**

413 **Overview on the theoretical data flow to derive the SRS consensus model**

414 The theoretical part of deriving the SRS consensus model comprised the following steps
415 (**Supplementary Fig. 2)**: (1) Hypothesis (asymmetric bulge as hallmark of SRS) by analysis of crystal
416 structure of *ASH1*-E3 (2) Selection of an appropriate conformational prediction tool (3) Conformational
417 prediction of instances and inspection to derive a consensus (**Fig. 2**) (4) Cross-check of the consensus with
418 published functional data (5) Creation of a computational screening pipeline (6) Check for further data
419 sources to filter results (**Supplementary Fig. 7**) (7) Screening for instances in selected sequence sets (8)
420 Inspection of instances for further patterns

421

422 **Identification of an appropriate conformational prediction tool**

423 The inspection of the crystal structure of *ASH1*-E3 suggested an asymmetric bulge as a pre-requisite
424 of kinked RNA conformation. The occurrence of such asymmetric bulges may be favorized in mRNPs by
425 the energy release due to molecular interaction. However, its occurrence is less likely in an individual RNA
426 and is less likely to be predicted as plausible conformation by computational prediction tools that commonly
427 deal with individual RNA sequences. In a survey, when investigating which of common tools[32,38–40] was
428 most likely to predict the asymmetric bulge of E3, we found Kinefold[32] as the one that best tolerated
429 asymmetric bulges (**Supplementary Fig. 3**). It predicted the asymmetric bulge consisting of a 5 nt Lbulge
430 and a 2 nt Sbulge, which closely approximates the length ratio of the asymmetric bulge observed in the E3-
431 She2p/3p crystal structure[8] that displayed a length ratio of 6 nt : 3 nt (considering only Watson-Crick base
432 pairs) or 4 nt : 1 nt (considering also atypical base pairings as observed in the crystal structure), respectively
433 (**Fig. 1A**).

434

435 **Cross-check of the consensus with published functional data**

436 We used published data of functional studies that involved both, mutagenesis and deletion of RNA
437 regions to cross-check for consistency with the SRS consensus model[10,13,25,30]. We characterized such

438 mutations in *ASH1* SRS elements that could clearly be assigned to destroy or to keep SRS function
439 according to the consensus model. All these constructs were in accordance with the consensus model (data
440 not shown).
441
442 **Development of a computational screening approach**
443 For the purpose of identifying further instances of the SRS consensus motif, we transformed the
444 consensus features into "patterns" to screen for "matches" in RNA sequences. We created a computational
445 screening pipeline consisting of in-house computational scripts and an external conformational prediction
446 algorithm. For the screening, three aspects were required: (1) Screening the sequence for conformational
447 aspects. For each RNA, the sequence was screened for two nested pairs of inverted repeats (each with a
448 minimal length of four nucleotides) that framed an asymmetric bulge region. We had to consider both
449 configurations in different screening patterns. The obtained constituted potential SRS instances. (2)
450 Screening the sequence for functional motifs. For each SRS instance, the Lbulge region was screened for
451 C_1, the opposite strand for C_2GA (positioned beyond IR1), with a C_1-C_2 distance of 6 nt between. The
452 screening strategy based on the approach of modular feature screening of mosaic-like patterns[41]. This
453 modular approach allowed to introduce further features to be screened (e.g., AA_L). It also tolerated length
454 variations between the distinct (sequence and conformational) features (e.g., one version tolerated C_1
455 positioned in the last nucleotide of IR1). As a consequence, permutated sets of varying patterns could be
456 summarized into two screening patterns (one for each configuration). (3) Validation by conformational
457 prediction. For each SRS instance, it was calculated if it would express a stable conformational fold that
458 fitted to the conformational arrangement of the SRS consensus. This analysis was performed using *Kinefold*
459 as energy-minimization approach that had proven best for large asymmetric bulges. On each site , the SRS
460 stretches were extended by 10 nt of the flanking sequences of their RNA sequences, to tolerate some degree
461 of freedom during calculation. In a series of tests using known SRS elements, a flanking region extension
462 by 10 nt had proven to be best in a range of 0 nt to 50 nt (data not shown). Of note, natural limitations of
463 common RNA conformational predictions occur due to the fact that such calculations are limited to predict

464 the conformation of distinct RNA stretches but less reliable regarding statements of the complete transcript
465 and second, they would ignore any energetic benefits by binding of molecular interactors (e.g., She2p/3p
466 proteins). As a consequence, we analyzed the 25th best conformational predictions of *Kinefold* (maximum
467 value = 25), to consider slight structural dynamics in each RNA molecule.
468
469 **Check for further data sources to filter for the consensus**
470 For the purpose to possibly extend the validation step of new SRS instances by conformational
471 predictions. we intended to cross-check the predicted conformations with experimental data to get deeper
472 insights on actual RNA conformations *in-vivo*. For our purposes, we tested the remarkable genome-wide
473 study based on PARS (Parallel Analysis of RNA Structure)[33]. For each nucleotide in transcripts, PARS-
474 scores described the percentage in occurring in single-strand ("S1-reads" of RNase S1) or double-strand
475 ("V1-reads" of RNase V1) regions of the RNA conformations. We extracted the S1- and V1-data and tested
476 the strong localization element E3[13] as positive control. As a result the PARS data was well suited to
477 describe the SRS conformation as observed in the E3 crystal structure[8] (**Supplemental Fig. 7**). However,
478 for further controls, the data comparison of paired/ unpaired occurrence of nucleotides between PARS data
479 and SRS consensus model failed (**Supplemental Fig. 7**). Commonly, RNA molecules display not one
480 distinct conformation in-vivo but are subject to conformational dynamics resulting in collections of
481 competing conformations. PARS data per nucleotide averaged the binding state (paired/unpaired) over all
482 competing conformations. It seemed that even the two SRS elements in *ASH1*-E1 and *ASH1*-E2A (proven
483 to be efficient localization elements[13]) were not sufficiently stable in the consensus conformations
484 according to their PARS data to distinguish them from alternative conformations
485
486 **Further bioinformatics resources**
487 For the investigation of molecular structures, we used Pymol[42] as 3D- viewer. Initial protein
488 characterization of She2p and She3p was performed using the ELM server[43]. RNA sequence alignments
489 were performed using Jalview[44]. RNA sequences were extracted of the Saccharomyces Genome Database[45].

490 For the actual screening, the mRNA sequences of *ASH1*, *IST2*, *EAR1* (aka *YMR171C*), *WSC2* and 20 further
491 mRNA sequences known to She2p/3p interact[10,13,25,29,30,36] were chosen. As sequence data, we used the
492 CDRs, extended by 200 nt downstream by default, due to the fact that for several mRNAs, the actual 3'
493 UTR lengths were still unknown. Functional annotation of Supplementary Table S1 followed [46]. Graphical
494 visualization of RNA conformations was performed using VARNA[47].
495
496 **Protein expression**
497 Expression in *E. coli* and purification of She2p, She3p, their fusion protein, and variants thereof
498 were performed as described in [8,23]. Purity and solubility were confirmed by size-exclusion chromatography
499 and SDS page.
500
501 **RNA preparation and isotope labeling**
502 For small scale *in vitro* transcription the MEGAshortscript T7 kit (life technologies) was used
503 according to the manufacturer's instructions. Synthetic DNA oligonucleotides containing a T7 promoter
504 and the respective sequence were used as a template. After *in vitro* transcription, the template was digested
505 by DNAse and the RNA was extracted using phenol:chloroform:isoamyl alcohol (25:24:1). Homogeneity
506 of RNA was assessed by native and denaturing TBE agarose gel electrophoresis.
507
508 **Electrophoretic mobility shift assay**
509 In a total volume of 20 µL 5 nM radioactively labeled RNA, 100 µg/mL yeast tRNA competitor,
510 and protein at a given concentration were mixed in buffer (20 mM Hepes, pH 7.8, 200 mM NaCl, 2 mM
511 $MgCl_2$, 2 mM DTT) with 4 % (v/v) glycerol. After 25 min at room temperature, the samples were resolved
512 by native TBE PAGE (6 % polyacrylamide) and gels incubated for 15 min in fixing solution (10 % (v/v)
513 acetic acid, 30 % (v/v) methanol) and subsequently vacuum dried. Gels were analyzed with radiograph
514 films or by a using Phosphorimager (FujifilmFLA5100).
515

516 **Surface plasmon resonance analysis (SPR)**

517 SPR experiments were performed with a BIACORE 3000 system (GE Healthcare). For protein-
518 RNA interactions, biotinylated RNA was streptavidin-captured on an SA-Chip (GE Healthcare) surface
519 according to manufacturers instructions. Measurements were performed at a flow rate of 30 μL/min in
520 standard buffer. Data were analyzed using the BIAevaluation software (GE Healthcare). Sample
521 sensograms were referenced against the signal in a ligand-free reference channel, and averaged values at
522 the equilibrium calculated. Steady state binding curves were obtained by plotting the response units at
523 equilibrium against protein concentration with the Origin Pro 9.0 (OriginLab) software.

524

525 **Computational quality check of the folds of the experimental constructs**

526 Part of the design of experimental constructs (minimal, config-switch) for the *in-vitro* and *in-vivo*
527 approaches was a quality check with *Kinefold*: The conformations of their sequences were predicted to
528 ensure that their expected conformation would be in agreement with the SRS consensus model

529

530 **Acknowledgements**

531 We would like to thank Herve Isambert for helpful discussions regarding Kinefold and Toby
532 Gibson for further support. This work was supported by the Deutsche Forschungsgemeinschaft (DFG,
533 German Research Foundation) to R.-P.J. (DFG-FOR2333, Project Number 270067186), to M.F.
534 (CEPLAS EXC 1028, CRC1208, DFG-FOR2333, DFG Fe448/8, DFG Fe448/9), to D.N. (DFG-FOR2333,
535 Project Number 270067186), and to Z.K. (DFG-FOR2333, Project Number 270067186).

536

537 **Declaration of Interest**

538 The authors declare no potential conflicts of interest.

539 **References**

540 (1)Eliscovich, C.; Singer, R. H. RNP Transport in Cell Biology: The Long and Winding Road. *Curr.*
541 *Opin. Cell Biol.* **2017**, *45*, 38–46. https://doi.org/10.1016/j.ceb.2017.02.008.

542 (2)Bashirullah, A.; Cooperstock, R. L.; Lipshitz, H. D. RNA Localization in Development. *Annu. Rev.*
543 *Biochem.* **1998**, *67*, 335–394. https://doi.org/10.1146/annurev.biochem.67.1.335.

544 (3)Johnstone, O.; Lasko, P. Translational Regulation and RNA Localization in Drosophila Oocytes and
545 Embryos. *Annu. Rev. Genet.* **2001**, *35*, 365–406.
546 https://doi.org/10.1146/annurev.genet.35.102401.090756.

547 (4)Kloc, M.; Bilinski, S.; Chan, A. P.; Allen, L. H.; Zearfoss, N. R.; Etkin, L. D. RNA Localization and
548 Germ Cell Determination in Xenopus. *Int. Rev. Cytol.* **2001**, *203*, 63–91.
549 https://doi.org/10.1016/s0074-7696(01)03004-2.

550 (5)Palacios, I. M.; St Johnston, D. Getting the Message across: The Intracellular Localization of mRNAs
551 in Higher Eukaryotes. *Annu. Rev. Cell Dev. Biol.* **2001**, *17*, 569–614.
552 https://doi.org/10.1146/annurev.cellbio.17.1.569.

553 (6)Dahm, R.; Kiebler, M.; Macchi, P. RNA Localisation in the Nervous System. *Semin. Cell Dev. Biol.*
554 **2007**, *18* (2), 216–223. https://doi.org/10.1016/j.semcdb.2007.01.009.

555 (7)Abouward, R.; Schiavo, G. Walking the Line: Mechanisms Underlying Directional mRNA Transport
556 and Localisation in Neurons and Beyond. *Cell. Mol. Life Sci. CMLS* **2021**, *78* (6), 2665–2681.
557 https://doi.org/10.1007/s00018-020-03724-3.

558 (8)Edelmann, F. T.; Schlundt, A.; Heym, R. G.; Jenner, A.; Niedner-Boblenz, A.; Syed, M. I.; Paillart, J.-
559 C.; Stehle, R.; Janowski, R.; Sattler, M.; Jansen, R.-P.; Niessing, D. Molecular Architecture and
560 Dynamics of ASH1 mRNA Recognition by Its mRNA-Transport Complex. *Nat. Publ. Group* **2017**,
561 *24* (2), 152–161.

562 (9)Niedner, A.; Edelmann, F. T.; Niessing, D. Of Social Molecules: The Interactive Assembly of ASH1
563 mRNA-Transport Complexes in Yeast. *RNA Biol.* **2014**, *11* (8), 998–1009.
564 https://doi.org/10.4161/rna.29946.

565 (10) Olivier, C.; Poirier, G.; Gendron, P.; Boisgontier, A.; Major, F.; Chartrand, P. Identification of a
566 Conserved RNA Motif Essential for She2p Recognition and mRNA Localization to the Yeast Bud.
567 *Mol. Cell. Biol.* **2005**, *25* (11), 4752–4766. https://doi.org/10.1128/MCB.25.11.4752-4766.2005.
568 (11) Patel, V. L.; Mitra, S.; Harris, R.; Buxbaum, A. R.; Lionnet, T.; Brenowitz, M.; Girvin, M.; Levy,
569 M.; Almo, S. C.; Singer, R. H.; Chao, J. A. Spatial Arrangement of an RNA Zipcode Identifies
570 mRNAs under Post-Transcriptional Control. *Genes Dev.* **2012**, *26* (1), 43–53.
571 https://doi.org/10.1101/gad.177428.111.
572 (12) Chartrand, P.; Singer, R. H.; Long, R. M. RNP Localization and Transport in Yeast. *Annu. Rev. Cell*
573 *Dev. Biol.* **2001**, *17*, 297–310. https://doi.org/10.1146/annurev.cellbio.17.1.297.
574 (13) Chartrand, P.; Meng, X. H.; Singer, R. H.; Long, R. M. Structural Elements Required for the
575 Localization of ASH1 mRNA and of a Green Fluorescent Protein Reporter Particle in Vivo. *Curr.*
576 *Biol. CB* **1999**, *9* (6), 333–336. https://doi.org/10.1016/s0960-9822(99)80144-4.
577 (14) Gonzalez, I.; Buonomo, S. B. C.; Nasmyth, K.; von Ahsen, U. ASH1 mRNA Localization in Yeast
578 Involves Multiple Secondary Structural Elementsand Ash1 Protein Translation. *Curr. Biol. CB*
579 **1999**, *9* (6), 337–340.
580 (15) Kanke, M.; Jambor, H.; Reich, J.; Marches, B.; Gstir, R.; Ryu, Y. H.; Ephrussi, A.; Macdonald, P.
581 M. Oskar RNA Plays Multiple Noncoding Roles to Support Oogenesis and Maintain Integrity of the
582 Germline/Soma Distinction. *RNA N. Y. N* **2015**, *21* (6), 1096–1109.
583 https://doi.org/10.1261/rna.048298.114.
584 (16) Mori, Y.; Imaizumi, K.; Katayama, T.; Yoneda, T.; Tohyama, M. Two Cis-Acting Elements in the
585 3' Untranslated Region of Alpha-CaMKII Regulate Its Dendritic Targeting. *Nat. Neurosci.* **2000**, *3*
586 (11), 1079–1084. https://doi.org/10.1038/80591.
587 (17) Serano, J.; Rubin, G. M. The Drosophila Synaptotagmin-like Protein Bitesize Is Required for
588 Growth and Has mRNA Localization Sequences within Its Open Reading Frame. *Proc. Natl. Acad.*
589 *Sci. U. S. A.* **2003**, *100* (23), 13368–13373. https://doi.org/10.1073/pnas.1835727100.

590 (18) Thio, G. L.; Ray, R. P.; Barcelo, G.; Schüpbach, T. Localization of Gurken RNA in Drosophila
591 Oogenesis Requires Elements in the 5' and 3' Regions of the Transcript. *Dev. Biol.* **2000**, *221* (2),
592 435–446. https://doi.org/10.1006/dbio.2000.9690.
593 (19) Long, R. M.; Gu, W.; Lorimer, E.; Singer, R. H.; Chartrand, P. She2p Is a Novel RNA-Binding
594 Protein That Recruits the Myo4p-She3p Complex to ASH1 mRNA. *EMBO J.* **2000**, *19* (23), 6592–
595 6601. https://doi.org/10.1093/emboj/19.23.6592.
596 (20) Münchow, S.; Sauter, C.; Jansen, R. P. Association of the Class V Myosin Myo4p with a Localised
597 Messenger RNA in Budding Yeast Depends on She Proteins. *J. Cell Sci.* **1999**, *112 (Pt 10)*, 1511–
598 1518.
599 (21) Bertrand, E.; Chartrand, P.; Schaefer, M.; Shenoy, S. M.; Singer, R. H.; Long, R. M. Localization of
600 ASH1 mRNA Particles in Living Yeast. *Mol. Cell* **1998**, *2* (4), 437–445.
601 https://doi.org/10.1016/s1097-2765(00)80143-4.
602 (22) Böhl, F.; Kruse, C.; Frank, A.; Ferring, D.; Jansen, R. P. She2p, a Novel RNA-Binding Protein
603 Tethers ASH1 mRNA to the Myo4p Myosin Motor via She3p. *EMBO J.* **2000**, *19* (20), 5514–5524.
604 https://doi.org/10.1093/emboj/19.20.5514.
605 (23) Heym, R. G.; Zimmermann, D.; Edelmann, F. T.; Israel, L.; Ökten, Z.; Kovar, D. R.; Niessing, D. In
606 Vitro Reconstitution of an mRNA-Transport Complex Reveals Mechanisms of Assembly and Motor
607 Activation. *J. Cell Biol.* **2013**, *203* (6), 971–984. https://doi.org/10.1083/jcb.201302095.
608 (24) Takizawa, P. A.; Vale, R. D. The Myosin Motor, Myo4p, Binds Ash1 mRNA via the Adapter
609 Protein, She3p. *Proc. Natl. Acad. Sci.* **2000**, *97* (10), 5273–5278.
610 (25) Chartrand, P.; Meng, X. H.; Huttelmaier, S.; Donato, D.; Singer, R. H. Asymmetric Sorting of
611 Ash1p in Yeast Results from Inhibition of Translation by Localization Elements in the mRNA. *Mol.*
612 *Cell* **2002**, *10* (6), 1319–1330. https://doi.org/10.1016/s1097-2765(02)00694-9.
613 (26) Long, R. M.; Singer, R. H.; Meng, X.; Gonzalez, I.; Nasmyth, K.; Jansen, R. P. Mating Type
614 Switching in Yeast Controlled by Asymmetric Localization of ASH1 mRNA. *Science* **1997**, *277*
615 (5324), 383–387. https://doi.org/10.1126/science.277.5324.383.

616 (27) Sil, A.; Herskowitz, I. Identification of Asymmetrically Localized Determinant, Ash1p, Required
617 for Lineage-Specific Transcription of the Yeast HO Gene. *Cell* **1996**, *84* (5), 711–722.
618 https://doi.org/10.1016/s0092-8674(00)81049-1.
619 (28) Takizawa, P. A.; Sil, A.; Swedlow, J. R.; Herskowitz, I.; Vale, R. D. Actin-Dependent Localization
620 of an RNA Encoding a Cell-Fate Determinant in Yeast. *Nature* **1997**, *389* (6646), 90–93.
621 https://doi.org/10.1038/38015.
622 (29) Shepard, K. A.; Gerber, A. P.; Jambhekar, A.; Takizawa, P. A.; Brown, P. O.; Herschlag, D.;
623 DeRisi, J. L.; Vale, R. D. Widespread Cytoplasmic mRNA Transport in Yeast: Identification of 22
624 Bud-Localized Transcripts Using DNA Microarray Analysis. *Proc. Natl. Acad. Sci. U. S. A.* **2003**,
625 *100* (20), 11429–11434. https://doi.org/10.1073/pnas.2033246100.
626 (30) Jambhekar, A.; McDermott, K.; Sorber, K.; Shepard, K. A.; Vale, R. D.; Takizawa, P. A.; DeRisi, J.
627 L. Unbiased Selection of Localization Elements Reveals Cis-Acting Determinants of mRNA Bud
628 Localization in Saccharomyces Cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (50), 18005–
629 18010. https://doi.org/10.1073/pnas.0509229102.
630 (31) Singh, N.; Blobel, G.; Shi, H. Hooking She3p onto She2p for Myosin-Mediated Cytoplasmic
631 mRNA Transport. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (1), 142–147.
632 https://doi.org/10.1073/pnas.1423194112.
633 (32) Xayaphoummine, A.; Bucher, T.; Isambert, H. Kinefold Web Server for RNA/DNA Folding Path
634 and Structure Prediction Including Pseudoknots and Knots. *Nucleic Acids Res.* **2005**, *33* (Web
635 Server issue), W605-610. https://doi.org/10.1093/nar/gki447.
636 (33) Kertesz, M.; Wan, Y.; Mazor, E.; Rinn, J. L.; Nutter, R. C.; Chang, H. Y.; Segal, E. Genome-Wide
637 Measurement of RNA Secondary Structure in Yeast. *Nature* **2010**, *467* (7311), 103–107.
638 https://doi.org/10.1038/nature09322.
639 (34) Tutucci, E.; Maekiniemi, A.; Snoep, J. L.; Seiler, M.; van Rossum, K.; van Niekerk, D. D.; Savakis,
640 P.; Zarnack, K.; Singer, R. H. *Cyclin CLB2 mRNA Localization Determines Efficient Protein*

641 *Synthesis to Orchestrate Bud Growth and Cell Cycle Progression*; preprint; Cell Biology, 2022.
642 https://doi.org/10.1101/2022.03.01.481833.
643 (35) Frankel, A. D. If the Loop Fits. *Nat. Struct. Biol.* **1999**, *6* (12), 1081–1083.
644 https://doi.org/10.1038/70002.
645 (36) Takizawa, P. A.; DeRisi, J. L.; Wilhelm, J. E.; Vale, R. D. Plasma Membrane Compartmentalization
646 in Yeast by Messenger RNA Transport and a Septin Diffusion Barrier. *Science* **2000**, *290* (5490),
647 341–344. https://doi.org/10.1126/science.290.5490.341.
648 (37) Müller, M.; Richter, K.; Heuck, A.; Kremmer, E.; Buchner, J.; Jansen, R.-P.; Niessing, D.
649 Formation of She2p Tetramers Is Required for mRNA Binding, mRNP Assembly, and Localization.
650 *RNA N. Y. N* **2009**, *15* (11), 2002–2012.
651 (38) Lorenz, R.; Bernhart, S. H.; Höner Zu Siederdissen, C.; Tafer, H.; Flamm, C.; Stadler, P. F.;
652 Hofacker, I. L. ViennaRNA Package 2.0. *Algorithms Mol. Biol. AMB* **2011**, *6*, 26.
653 https://doi.org/10.1186/1748-7188-6-26.
654 (39) Reuter, J. S.; Mathews, D. H. RNAstructure: Software for RNA Secondary Structure Prediction and
655 Analysis. *BMC Bioinformatics* **2010**, *11*, 129. https://doi.org/10.1186/1471-2105-11-129.
656 (40) Zuker, M. Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction. *Nucleic Acids*
657 *Res.* **2003**, *31* (13), 3406–3415. https://doi.org/10.1093/nar/gkg595.
658 (41) Seiler, M.; Mehrle, A.; Poustka, A.; Wiemann, S. The 3of5 Web Application for Complex and
659 Comprehensive Pattern Matching in Protein Sequences. *BMC Bioinformatics* **2006**, *7*, 144.
660 https://doi.org/10.1186/1471-2105-7-144.
661 (42) Schrödinger, L.; DeLano, W. PyMOL. **2020**.
662 (43) Kumar, M.; Michael, S.; Alvarado-Valverde, J.; Mészáros, B.; Sámano-Sánchez, H.; Zeke, A.;
663 Dobson, L.; Lazar, T.; Örd, M.; Nagpal, A.; Farahi, N.; Käser, M.; Kraleti, R.; Davey, N. E.;
664 Pancsa, R.; Chemes, L. B.; Gibson, T. J. The Eukaryotic Linear Motif Resource: 2022 Release.
665 *Nucleic Acids Res.* **2022**, *50* (D1), D497–D508. https://doi.org/10.1093/nar/gkab975.

666 (44) Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M.; Barton, G. J. Jalview Version 2--a
667 Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinforma. Oxf. Engl.* **2009**, *25* (9),
668 1189–1191. https://doi.org/10.1093/bioinformatics/btp033.
669 (45) Engel, S. R.; Dietrich, F. S.; Fisk, D. G.; Binkley, G.; Balakrishnan, R.; Costanzo, M. C.; Dwight, S.
670 S.; Hitz, B. C.; Karra, K.; Nash, R. S.; Weng, S.; Wong, E. D.; Lloyd, P.; Skrzypek, M. S.;
671 Miyasato, S. R.; Simison, M.; Cherry, J. M. The Reference Genome Sequence of Saccharomyces
672 Cerevisiae: Then and Now. *G3 Bethesda Md* **2014**, *4* (3), 389–398.
673 https://doi.org/10.1534/g3.113.008995.
674 (46) Engel, S. R.; Wong, E. D.; Nash, R. S.; Aleksander, S.; Alexander, M.; Douglass, E.; Karra, K.;
675 Miyasato, S. R.; Simison, M.; Skrzypek, M. S.; Weng, S.; Cherry, J. M. New Data and
676 Collaborations at the Saccharomyces Genome Database: Updated Reference Genome, Alleles, and
677 the Alliance of Genome Resources. *Genetics* **2022**, *220* (4), iyab224.
678 https://doi.org/10.1093/genetics/iyab224.
679 (47) Darty, K.; Denise, A.; Ponty, Y. VARNA: Interactive Drawing and Editing of the RNA Secondary
680 Structure. *Bioinforma. Oxf. Engl.* **2009**, *25* (15), 1974–1975.
681 https://doi.org/10.1093/bioinformatics/btp250.
682
683
684
685

686 **Figures Legends**

687 **Figure 1. The SRS consensus model describes all four *ASH1* SRS elements. A** Left, conformational
688 prediction of $E3_{1771-1821}$ by Kinefold displays the asymmetric bulge as a strand of 5 nt ('Lbulge') juxtaposed
689 with a shorter strand ('Sbulge') both indicated in red, previously known functional motifs in dark green.
690 Right, the conformation of E3 referred to its crystal structure bound to She2p/3p (**PDB: 5NEW**) displayed
691 a rearranged the asymmetric bulge with atypical base-pairing, with a length ratio of 6 nt : 3 nt (with atypical
692 base-pairs dashed in blue, but ignored in the count), and with further positions of sequence motifs adapted
693 to analogous functional residues as studied in E2A, E2B and E3 (indicated in light green). **B** Left, Kinefold
694 prediction and 180° rotation of $E2A_{1109-1185}$ revealed an asymmetric bulge with the same 5 nt : 2 nt ratio as
695 in E3. Right, the high degree of sequence similarity (**Supplementary Fig. 1**) enabled a modeling in tight
696 accordance to the E3 conformation of the crystal structure that even comprised the uncommon base-pairing
697 C_{1122}-C_{1174} and the A_{1123}-A_{1169} base-stacking (dashed in blue). **C** Left, the optimal conformation of $E2B_{1263-}$
698 $_{1320}$ after 180 ° rotation showed internal repeats (left, boxed in orange) that could be re-modeled according
699 to Kinefold predictions (right) but with increased energy content (-10.3 kcal/mol vs.-7.4 kcal/mol). **D** SRS
700 element $E1_{623-701}$ after Kinefold prediction and 180° rotation (left) and with adapted conformation (right).
701 **E** Scheme of the features of all four modeled SRS elements, projected on $E3_{1771-1821}$. An asymmetric bulge
702 (red) is framed by two inverted repeats (yellow), with sequence motifs indicated in green. **F** Schematic
703 presentation of the SRS consensus motif with its features displayed in both configurations (IR1 boxed in
704 grey, IR2 boxed in blue, Lbulge and Sbulge in red, sequence motifs indicated by green labeling). Regarding
705 the two configurations, their different order of features (left) results in a overall spatial arrangement that is
706 conserved but with varying positioning of 5' start and 3' end in its conformations (middle). Right,
707 conformational view on the unified model of the SRS consensus motif.
708

709 **Figure 2. The modeling of the SRS consensus involves choice of appropriate conformation prediction**
710 **tool, rotation and mutual feature alignment of the four SRS elements in *ASH1*.** The overview
711 demonstrates the effects of the three steps of modeling on the extended sizes of the *ASH1* SRS elements as
712 they were characterized before. Whereas the initially known conformations displayed a huge degree of
713 heterogeneity (**A, D, H, L**), the conformation prediction tool Kinefold tolerated asymmetric bulges in each
714 SRS element (**B, E, I, M**). For E3, using the published crystal structure as a blue print, its conformational
715 prediction (**B**) was adapted, with altered conformation of the asymmetric bulge (resulting in atypical base
716 pairings, marked in red, **C**). Its final asymmetric feature is indicated by a red triangle, with the numbers of
717 nucleotides displayed for both, Lbulge and Sbulge. By later analyses, we could assign one of its functional
718 cytosines as being analogous to already known CGA sequence motifs with other SRS elements (indicated

719 in red). For E2A, E1 and E2B, it was the rotation of the initial predictions that harmonized the respective
720 conformations with E3 (**F, J, N**). For E2A, due to its remarkable sequence identity with E3, we could also
721 introduce the same atypical base pairings (marked in red, **G**). As mentioned before, E2B represented an
722 outlier: it was a alternative conformation prediction that could be harmonized with E3, suggesting the need
723 of a further intramolecular rearrangement to align the inverted boxes (marked in yellow, **N**, **O**).
724
725 **Figure 3. The configuration switch in E3 displays conserved function and surface plasmon resonance**
726 **analysis confirms the optimal C_1-C_2 distance of 6 nt in minimal constructs of E3. A** Representative
727 results of electrophoretic mobility shift assays (EMSA) of triplicates to analyze complex formation of wild-
728 type and variant constructs of radioactive ASH1 $E3_{51nt}$ RNA with the She2p-She3p fusion protein ($She2p_{6-}$
729 $_{246}$-$(GGSGG)_2$-$She3p_{331-405}$) in increasing concentration (0.02 - 5 μM): Wildtype $E3_{1771-1821}$ (left), construct
730 with switch of $E3_{1771-1821}$ from config-I to config-II (middle) and control M6 mutational variant of $E3_{1771-}$
731 $_{1821}$ with decreased SRS function[8]. **B** Results of surface plasmon resonance analysis (SRP), performed in
732 triplicates to analyze the minimal E3 constructs and alterations in their C_1-C_2 distance. The minimal
733 construct $E3_{28nt}$wt was re-designed to create constructs with differing C1-C2 distances (5 and 7 nt) and was
734 incubated with She2p-She3p fusion protein (125 nM).
735
736 **Figure 4. Computational screen for the SRS consensus motif across 24 known She2p/3p interacting**
737 **transcripts results in 30 predicted instances in 17 transcripts. A** Computational screen utilized both,
738 feature screening of the SRS consensus motif and conformational evaluation (**Materials and Methods**)
739 and was applied on 24 sequences known for She2p/3p interaction but with varying experimental confidence
740 (**Supplementary Tab. 1**). It identified 30 SRS predictions in 17 sequences, here visualized as bar charts,
741 with their coding sequences (rectangles) and 3' UTR extensions (200 nt, black lines). SRS predictions were
742 indicated by their configurations (config-I: red, config-II: blue). **B** For each configuration, the occurrence
743 of the AA_L dinucleotide was specified. **C** Schematic overview on the high-confidential set of SRS elements.
744 Similar to the SRS elements in *ASH1*, predictions in *IST2*, *EAR1* and *WSC2* matched published, well-
745 characterized SRS elements (nomenclature, see **Tab. 2**) and could be arranged accordant to the model.
746 Color scheme is adopted of **Figure 1**, with visualization of AA_L dinucleotides, boxed in orange).
747 Uncommon base pairings had detected in the E3-She2p/She3p crystal structure were consistently mapped
748 to E2A that displayed high sequence similarity with E3 after rotation (blue-dashed). Of note, similar
749 uncommon base-pairings may also arise in E1 and *IST2*-SRS-2.
750

751 **Figure 5. The configuration switch in the minimal SRS constructs of *EAR*-SRS2, *IST*-SRS2, *KSS1*-**
752 **SRS1 and *TCB3*-SRS1 displays conserved SRS function in electrophoretic mobility shift assays.**
753 Representative results of electrophoretic mobility shift assays (EMSA) to analyze interaction of radioactive
754 wild-type and designed RNA constructs with the She2p-She3p fusion protein (She2p$_{6\text{-}246}$-(GGSGG)$_2$-
755 She3p$_{331\text{-}405}$) in increasing concentration (0.02 - 5 µM). All minimal constructs were designed switched in
756 configuration (from config-II to config-I). Their final loop sequence GAAA referred to the artificial loop
757 sequence of the E3 constructs in [8] that was shown not to modify SRS function. **A** EMSA experiments of
758 well-known SRS elements. Left, *EAR1*-SRS2$_{1572\text{-}1621}$ (config-II) was studied as wildtype construct *EAR1*-
759 SRS2$_{50nt}$wt and as config-switched, minimal construct *EAR1*-SRS2$_{26nt\text{-}}$switch (config-I). Right, *IST2*-
760 SRS2$_{2716\text{-}2777}$ (config-II) was studied as wildtype construct *IST2*-SRS2$_{62nt}$wt and as config-switched,
761 minimal construct *IST2*-SRS2$_{29nt\text{-}}$switch (config-I). **B** EMSA experiments of new, predicted SRS elements.
762 Left, *KSS1*-SRS1 (config-II) was studied as wildtype construct *KSS1*-SRS1$_{72nt}$wt and as config-switched,
763 minimal construct *KSS1*-SRS1$_{29nt\text{-}}$switch (config-I). Right, *TCB3*-SRS1 (config-II) was studied as config-
764 switched, minimal construct *TCB3*-SRS1$_{38nt\text{-}}$switch (config-I).
765
766 **Figure 6. *EAR1*-SRS2$_{28nt}$switch is partly sensitive to alterations of its functional motifs but tolerates**
767 **alterations in its conformational features and in the AA$_L$ dinucleotide of the Lbulge as showed by**
768 **surface plasmon resonance analysis. A** Schemes of the minimal constructs of *EAR1*-SRS2$_{28nt}$switch that
769 were re-designed to create constructs with altered conformational features (indicated in red in
770 conformations). The final loop sequence GAAA referred to the artificial loop sequence of the E3 constructs
771 in [8] that was shown not to alter SRS function **B** Substitution scheme to analyze the relevance of cytosines
772 in the sequence motifs of the SRS consensus model. Alterations are marked in red in substitution tables that
773 refer to both, sequence positions and terminologies of resulting constructs. Since two cytosines (C_{1601} and
774 C_{1602}) precluded an unambiguous assignment of the C_1 position, both had to be considered. A third cytosine
775 C_{1600}, marked by a star, was not considered in alterations due to its uncommonly short distance to C_2. **C**
776 Substitution scheme to analyze the relevance of the AA$_L$ dinucleotide in the Lbulge. For each alterations,
777 construct name (blue), conserved sequence in the Lbulge (black) and alterations (red) are displayed. **D**
778 Results of surface plasmon resonance analysis (SRP), performed in triplicates to analyze the minimal
779 EAR1-SRS2 constructs and alterations in their C_1-C_2 distance. E3 RNA was incubated with She2p-She3p
780 fusion protein (125 nM).
781
782

Figure 1

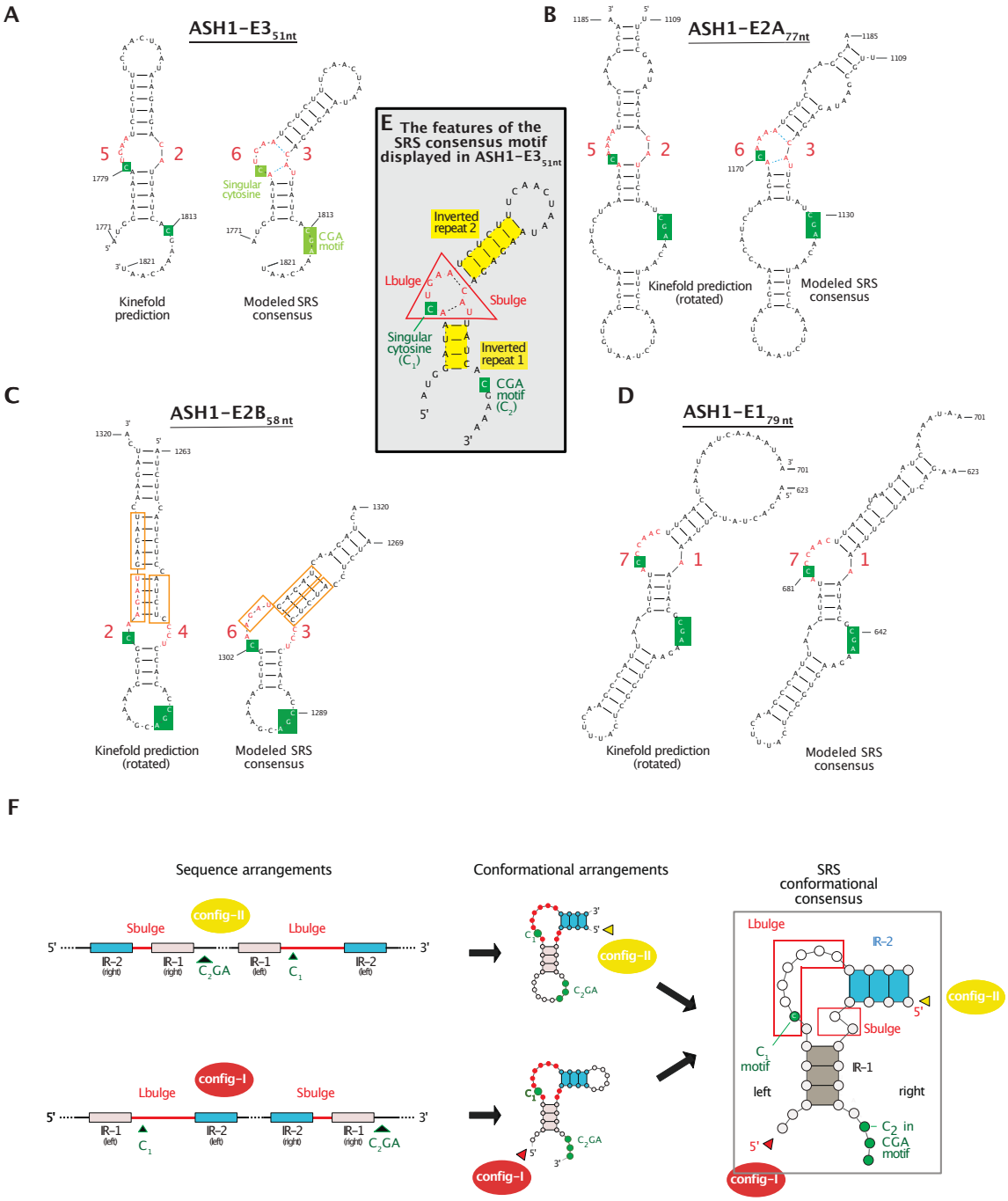

A
ASH1-E3 51nt
Kinefold prediction
Modeled SRS consensus
Singular cytosine
CGA motif
B
ASH1-E2A 77nt
Kinefold prediction (rotated)
Modeled SRS consensus
E The features of the SRS consensus motif displayed in ASH1-E3 51nt
Inverted repeat 2
Lbulge
Sbulge
Singular cytosine (C1)
Inverted repeat 1
CGA motif (C2)
C
ASH1-E2B 58 nt
Kinefold prediction (rotated)
Modeled SRS consensus
D
ASH1-E1 79 nt
Kinefold prediction (rotated)
Modeled SRS consensus
F
Sequence arrangements
Conformational arrangements
SRS conformational consensus
config-II
config-I
Sbulge
Lbulge
IR-2
IR-1
C2GA
C1
C1 motif
left
right
C2 in CGA motif

Figure 2

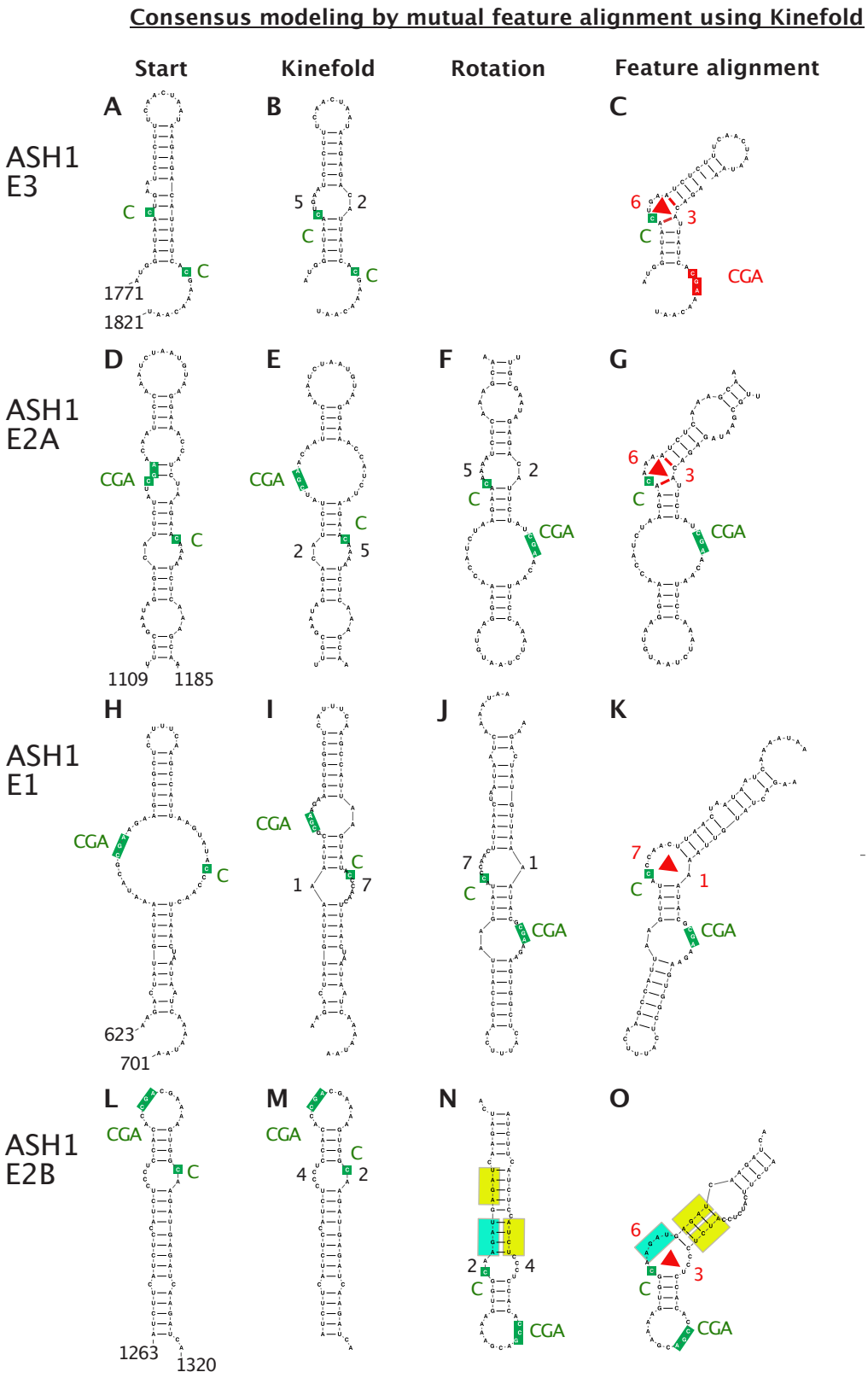
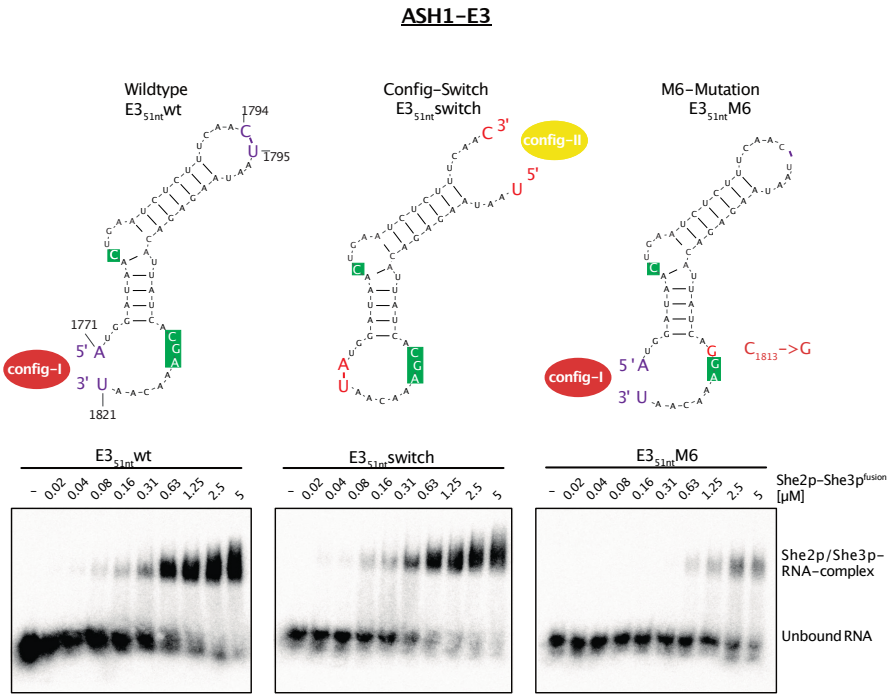
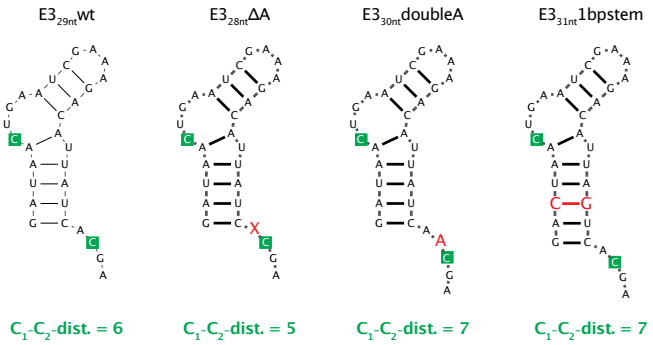

Consensus modeling by mutual feature alignment using Kinefold
Start
Kinefold
Rotation
Feature alignment
ASH1 E3
ASH1 E2A
ASH1 E1
ASH1 E2B
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
C
CGA
1771
1821
1109
1185
623
701
1263
1320

Figure 3

A


ASH1-E3
Wildtype
E3 51nt wt
1794
1795
1771
1821
5' A
3' U
config-I
Config-Switch
E3 51nt switch
C 3'
U 5'
config-II
A
U
M6-Mutation
E3 51nt M6
C1813->G
5' A
3' U
config-I
E3 51nt wt
E3 51nt switch
E3 51nt M6
- 0.02 0.04 0.08 0.16 0.31 0.63 1.25 2.5 5
She2p-She3p fusion [µM]
She2p/She3p-RNA-complex
Unbound RNA

B


E3 29nt wt
E3 28nt ΔA
E3 30nt doubleA
E3 31nt 1bpstem
C1-C2-dist. = 6
C1-C2-dist. = 5
C1-C2-dist. = 7
C1-C2-dist. = 7

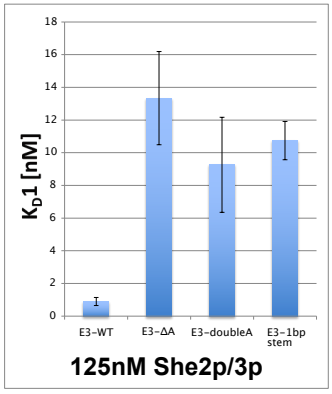
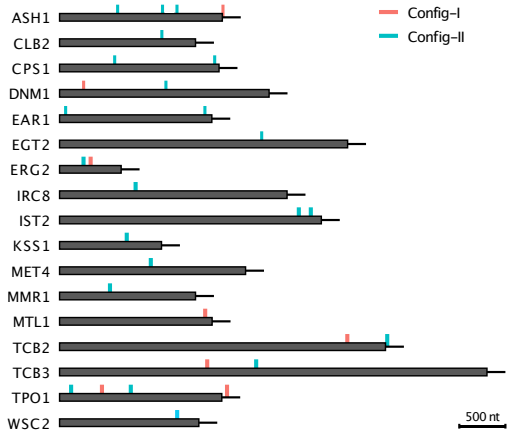
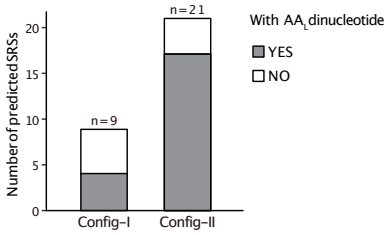

KD1 [nM]
0
2
4
6
8
10
12
14
16
18
E3-WT
E3-ΔA
E3-doubleA
E3-1bp stem
125nM She2p/3p

Figure 4

A


ASH1
CLB2
CPS1
DNM1
EAR1
EGT2
ERG2
IRC8
IST2
KSS1
MET4
MMR1
MTL1
TCB2
TCB3
TPO1
WSC2
Config-I
Config-II
500 nt

B


Number of predicted SRSs
0
5
10
15
20
n=9
n=21
Config-I
Config-II
With AA dinucleotide
YES
NO

C

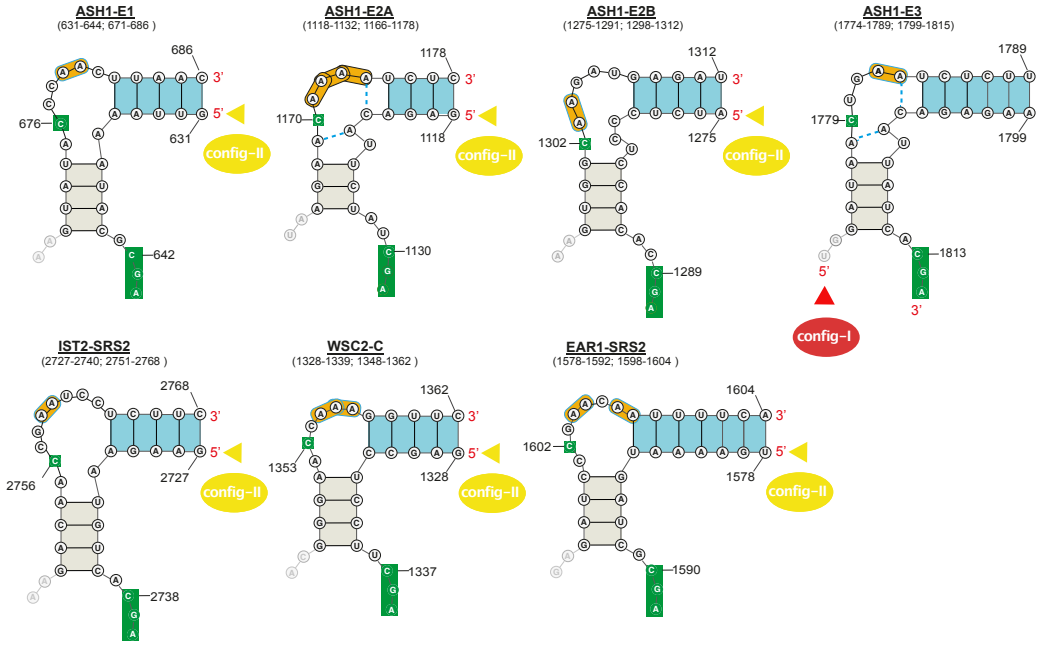
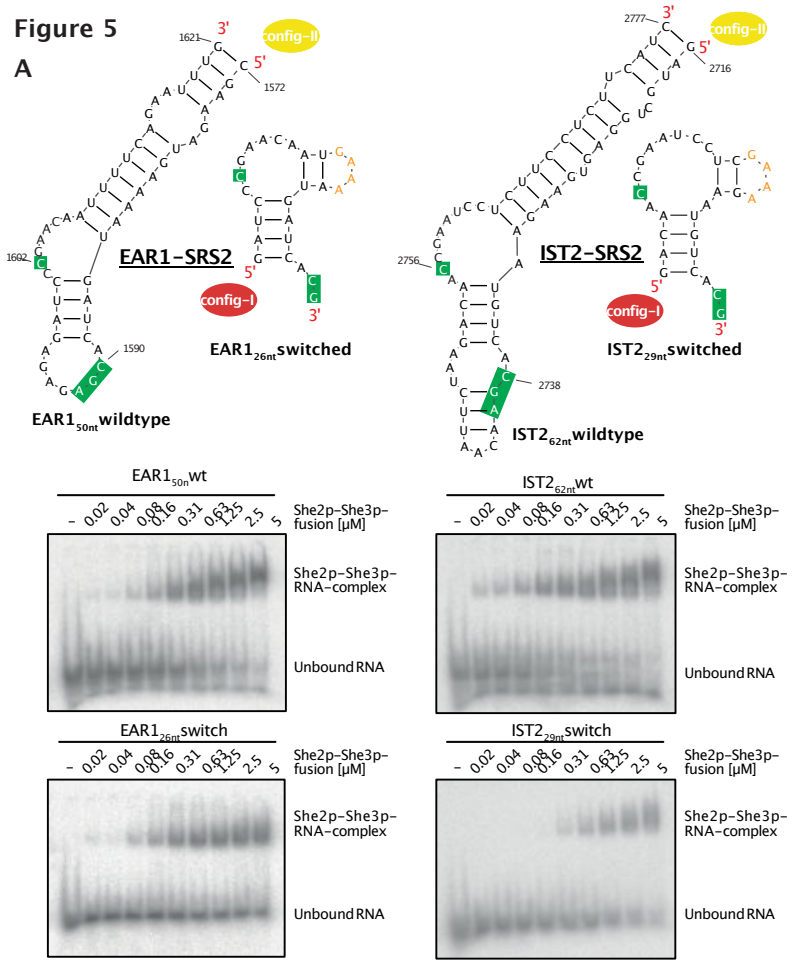

ASH1-E1
(631-644; 671-686)
ASH1-E2A
(1118-1132; 1166-1178)
ASH1-E2B
(1275-1291; 1298-1312)
ASH1-E3
(1774-1789; 1799-1815)
IST2-SRS2
(2727-2740; 2751-2768)
WSC2-C
(1328-1339; 1348-1362)
EAR1-SRS2
(1578-1592; 1598-1604)
config-II
config-I
3'
5'

Figure 5

A


EAR1-SRS2
IST2-SRS2
config-II
config-I
EAR1$_{26nt}$switched
EAR1$_{50nt}$wildtype
IST2$_{29nt}$switched
IST2$_{62nt}$wildtype
EAR1$_{50n}$wt
IST2$_{62nt}$wt
EAR1$_{26nt}$switch
IST2$_{29nt}$switch
- 0.02 0.04 0.08 0.16 0.31 0.63 1.25 2.5 5
She2p-She3p-fusion [µM]
She2p-She3p-RNA-complex
Unbound RNA

B

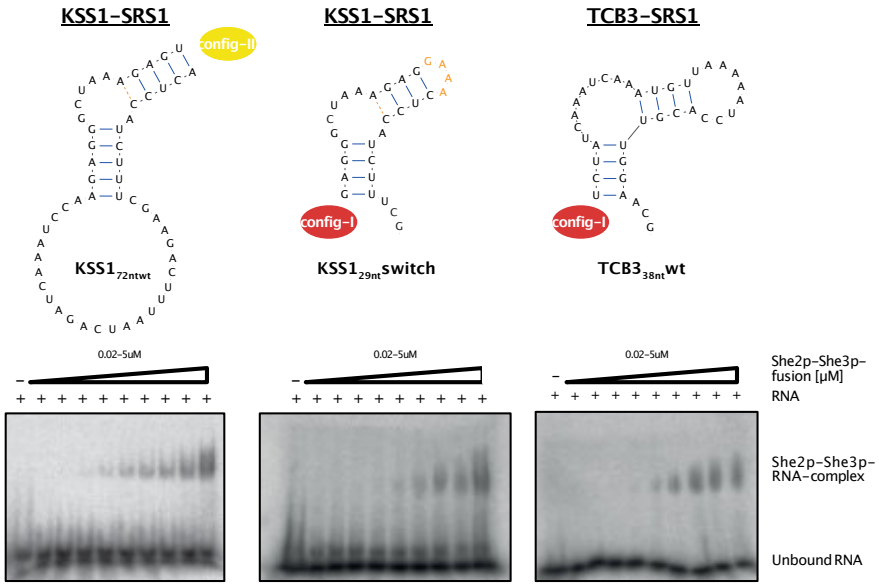
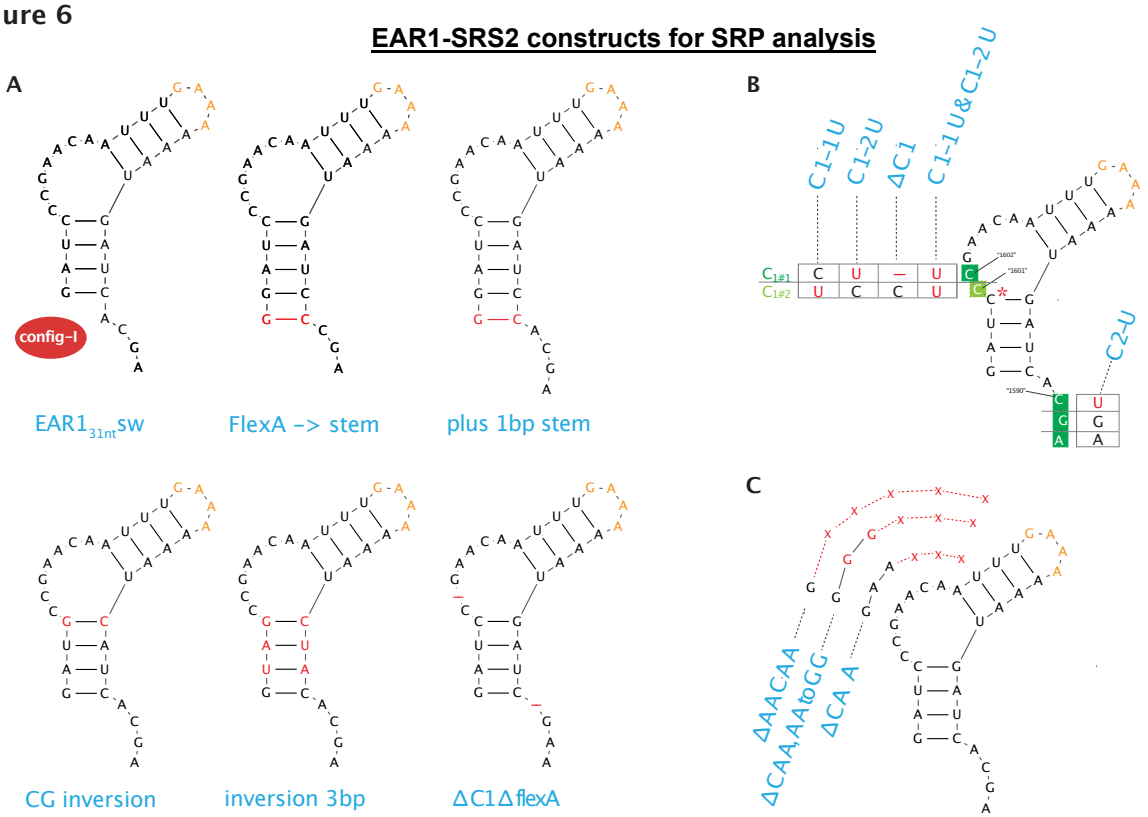

KSS1-SRS1
KSS1-SRS1
TCB3-SRS1
config-II
config-I
KSS1$_{72ntwt}$
KSS1$_{29nt}$switch
TCB3$_{38nt}$wt
0.02-5uM
She2p-She3p-fusion [µM]
RNA
She2p-She3p-RNA-complex
Unbound RNA

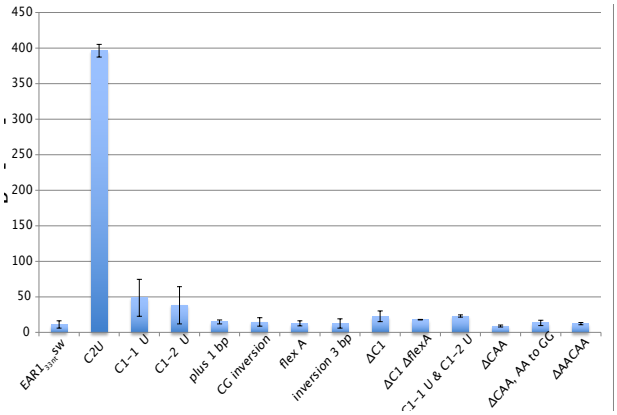
Figure 6

EAR1-SRS2 constructs for SRP analysis



D

SRP: EAR1-SRS2 constructs vs. 125nM She2p/3p



792

INFO-BOX: Summary of the SRS consensus motif

1. Two **inverted repeats** (IR-1, IR-2) flank an **asymmetric bulge** (consisting of **Lbulge** and **Sbulge**)
2. The conserved **singular cytosine C_1** resides in one of the first three positions of the Lbulge
3. The conserved **C_2GA triplet** resides in a further bulge, at the opposite strand of C_1, separated by an inverted repeat in downstream direction on this strand in a **C_1-C_2 distance of six positions**
4. The two **configurations** are marked by the two different 5' entry sites into the kinked SRS element conformation.

793
794
795
796
797
798

TABLE 1 Overview on the four localization elements of *ASH1*. Positions and configuration classification are listed according to the SRS consensus motif

SRS name	SRS config I	SRS config II	SRS AA motif in Lbulge	SRS start	SRS end
E2		II	yes	631	686
E2A		II	yes	1118	1178
E2B		II	yes	1275	1312
E3	I		yes	1774	1817

799
800
801
802

TABLE 2 Overview on the *SRS instance collection.* The columns list info at transcripts and SRS instance level; SRS instance are classified according to the SRS consensus motif (including positions and configuration); the last two columns contain information to the identification of SRS elements in fragment size or as a detailed description in previous publications and this study

mRNA #	mRNA name	mRNA gene ID	SRS instance also known as	SRS instance collection name (*)	SRS start	SRS end	SRS config I	SRS config II	SRS as terminal element (***)	SRS with AA motif	SRS fragment reference	SRS element reference
1	*ASH1*	*YKL185W*	E1	ASH1-E1	631	686		II		AA	several	several
			E2A	ASH1-E2A	1118	1178		II		AA	several	several
			E2B	ASH1-E2B	1275	1312		II	terminal	AA	several	several
			E3, U_{min}	ASH1-E3	1774	1817	I			AA	several	several
2	*CLB2*	*YPR119W*		CLB2-SRS1	1111	1145		II		AA		Tutucci et al
3	*CPS1*	*YJL172W*		CPS1-SRS1	599	636		II		AA		
				CPS1-SRS2	1684	1738		II		AA		
4	*DNM1*	*YLL001W*		DBM1-SRS1	262	302	I					
				DBM1-SRS2	1155	1189		II				
5	*EAR1*	*YMR171C*		EAR1-SRS1	110	167		II		AA		
				EAR1-SRS2	1578	1614		II	terminal	AA	Olivier et al	Olivier et al this study
6	*EGT2*	*YNL327W*		EGT2-SRS1	2194	2235		II		AA		
7	*ERG2*	*YMR202W*		ERG2-SRS1	261	314		II				
				ERG2-SRS2	338	380	I			AA		
8	*IRC8*	*YJL051W*		IRC8-SRS1	826	870		II		AA		
9	*IST2*	*YBR086C*		IST2-SRS1	2597	2626		II		AA		
				IST2-SRS2	2727	2768		II	terminal	AA	Olivier et al	Olivier et al this study
10	*KSS1*	*YGR040W*		KSS1-SRS1	730	781		II		AA		this study
11	*MET4*	*YNL103W*		MET4-SRS1	991	1051		II				
12	*MMR1*	*YLR190W*		MMR1-SRS1	548	590		II				
13	*MTL1*	*YGR023W*		MTL1-SRS1	1581	1651	I					
14	*TCB2*	*YNL087W*		TCB2-SRS1	3123	3175	I			AA		
				TCB2-SRS2	3556	3592		II		AA		
15	*TCB3*	*YML072C*		TCB3-SRS1	1603	1642	I			AA		this study
				TCB3-SRS2	2134	2173		II		AA		
16	*TPO1*	*YLL028W*		TPO1-SRS1	175	240		II		AA		
				TPO1-SRS2	461	527	I					
				TPO1-SRS3	774	822		II				
				TPO1-SRS4	1818	1862	I					
17	*WSC2*	*YNL283C*	WSC2C	WSC2-C	1326	1373		II		AA		Jambhekar et al

Remarks

SRS instances studies in this work are indicated in **red**.

(*) SRS names in SRS collection — previously existing SRS terminologies are kept
in case of non-existing previous terminologies, the nomenclature contains "-SRS" suffixes (**)

(**) Numbering with suffixes — the numbering follows their positions in 5'-3' RNA sequence.
SRS elements that were discovered later, are numbered with subsequent numbers

(***) SRS as terminal element — indicates such SRS instances that reside at the terminal region of a RNA stem fold