Supplemental Material

Structural basis for recognition of intron branchpoint RNA by yeast Msl5 and selective effects of interfacial mutations on splicing of yeast pre-mRNAs

Agata Jacewicz, Lidia Chico, Paul Smith, Beate Schwer, and Stewart Shuman

Tables S1, S2, S3, S4

Figures S1, S2

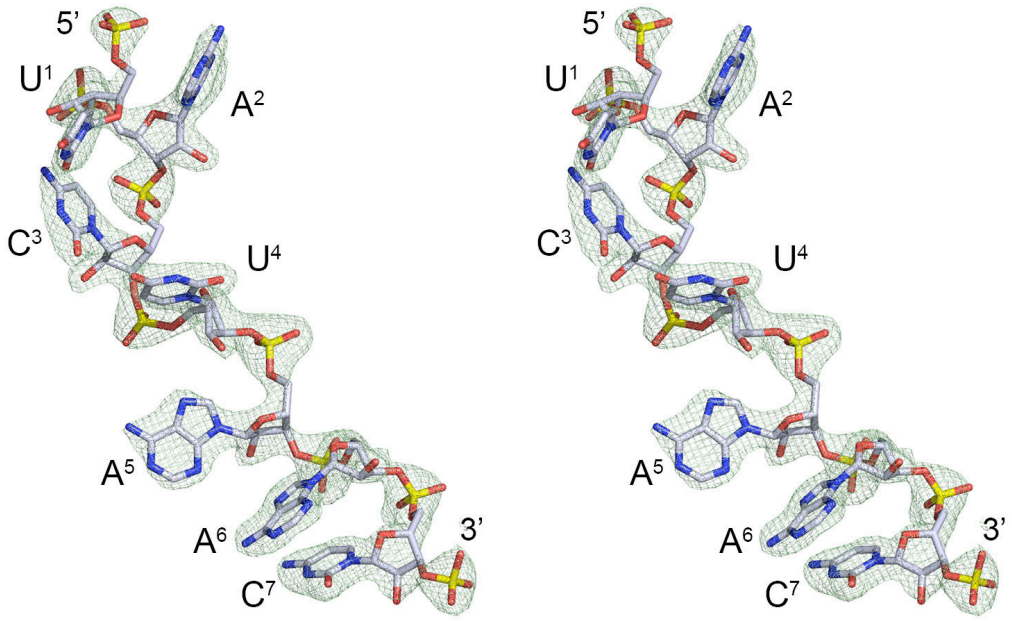
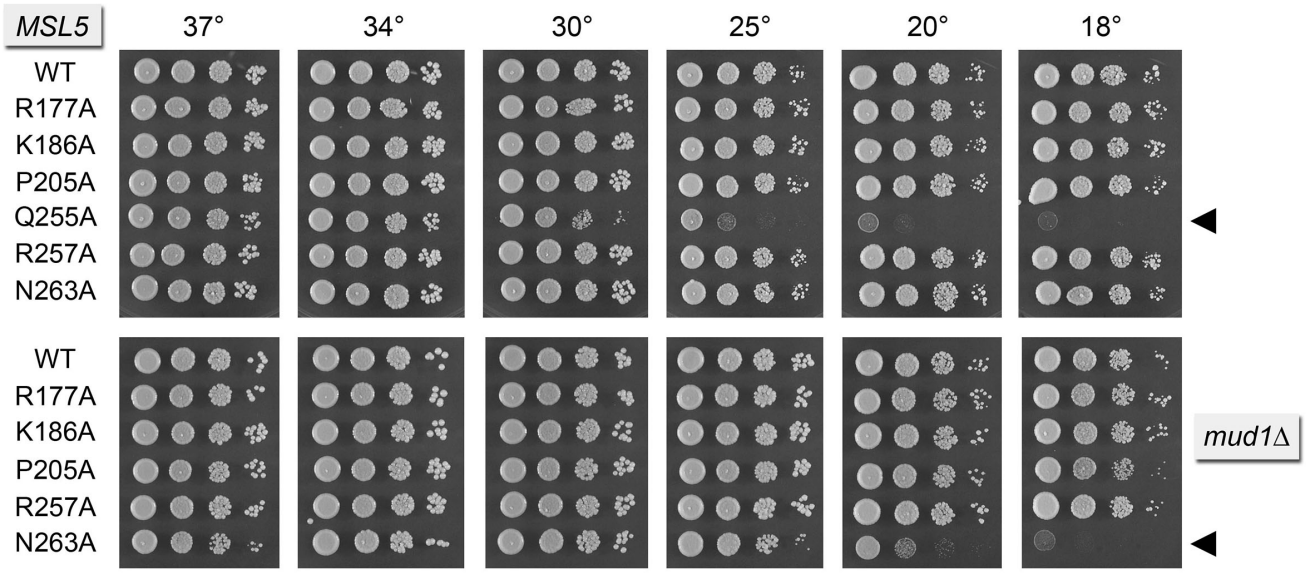


Figure S1. **Electron density for branchpoint RNA**. Stereo view of the Fo-Fc map (green mesh) of the branchpoint RNA ligand in the SeMet-Msl5-(KH-QUA2) crystal, contoured at 2 σ. The RNA modeled into the density is shown with the 5' and 3' termini indicated and each nucleobase identified.



Q255A is synthetic lethal with mud1Δ, nam8Δ, and mud2Δ

R177A, K186A, P205A, R257A and N263A grow like WT in nam8Δ and mud2Δ backgrounds

Fig. S2. **New structure-guided mutagenesis of the Msl5•branchpoint interface**. *Top panel*. Complementation of *msl5*Δ by the indicated wild-type *MSL5* and *MSL5-Ala* alleles was assayed by plasmid shuffle. His^+ transformants were streaked on agar medium containing FOA. FOA-resistant colonies with viable *MSL5* alleles were grown to mid-log phase in YPD broth and adjusted to equivalent A_{600}. Aliquots (3 µl) of serial 10-fold dilutions were spotted on YPD agar plates, which were then incubated at 18, 25, 30, 34 and 37°C. The *Q255A* allele uniquely elicited a cold-sensitive growth defect (◀ at *right*). *Bottom panel*. Complementation of *msl5*Δ in the genetic backgrounds specified (*mud1*Δ, *nam8*Δ, *mud2*Δ) by otherwise functional *MSL5-Ala* alleles was assayed by plasmid shuffle (Chang et al. 2012; Schwer et al. 2013). *Q255A* failed to yield FOA-resistant colonies at any temperature tested in the *mud1*Δ, *nam8*Δ, *mud2*Δ strain backgrounds. Spot-tests of *mud1*Δ *msl5*Δ cells complemented by the indicated *MSL5* alleles for growth at the indicated temperatures are shown. The *N263A* allele caused a cold-sensitive growth defect in the *mud1*Δ background (◀ at *right*). Other complementation results are summarized at *bottom*.

Table S1: **Diffraction data and refinement statistics**

	Native	**SeMet**
Diffraction data		
Space group	*$P12_11$*	*P422*
Cell dimensions		
a, *b*, *c* (Å)	53.38,130.97,53.37	81.95,81.95,51.49
α, β, γ (°)	90.00, 117.43, 90.00	90.00, 90.00, 90.00
Wavelength (Å)	1.072	0.97919
Resolution (Å)	45.61−1.80 (1.90−1.80)	36.65−2.18 (2.30−2.18)
Unique reflections	59297	9666
R_{merge} [a)]	0.10 (0.357)	0.108 (0.464)
I / σI	9.5 (3.4)	40.2 (11.7)
Completeness (%)	99.3 (95.4)	100.0 (100.0)
Multiplicity	5.4 (4.2)	39.1 (26.2)
Phasing statistics		
Phasing method	Phenix.Phaser	Se-SAD (4 Se sites)/Phenix.Phaser
Resolution (Å)	45.61-1.80	36.65-2.70
Signal to noise[b)]	RFZ=11.4,TFZ=41.8	Ano=0.277
Figures of merit [c)]		0.488 (Phaser)
Refinement statistics		
Resolution (Å)	45.61-1.80 (1.83-1.80)	36.65−2.20 (2.52-2.20)
Completeness	99.2 (85.5)	100.0 (100.0)
Twinning fraction	0.5	
Twinning operator	l, -k, h	
R_{work} / R_{free} [d)]	17.5 / 20.6	17.2 / 21.5
B-factors (Å^2) Overall/Wilson	34.3 / 22.7	34.9 / 27.3
TLS Groups Mean Anisotropy	0.33 (33 groups)	0.44 (4 groups)
RMS deviations		
Bond lengths (Å)	0.010	0.007
Bond angles (°)	1.065	1.050
Ramachandran plot		
Favored (%)	96.2	98.5
Outliers	Glu246, Gly247 (A chain)	none
Model contents		
Protomers / ASU	4	1
Protein residues	487	127
RNA	4 oligonucleotides	1 oligonucleotide
Water	477	87
Ligands/Ions	2 acetates, 4 sulfates, 5 glycerols	1 chloride, 1 sulfate, 1 glycerol
PDB ID	4WAN	4WAL

Standard definitions are used for all parameters. Figures in parentheses refer to data in the highest resolution shell. The refinement and geometric statistics come from Phenix.Refine.

a) R_{merge} output from Scala

b) Signal to noise for anomalous phasing (Ano) is the fraction of the data for which $D_{ano} > 3.0\sigma\ D_{ano}$ for data to 2.70 Å, as calculated in Phenix.Xtriage. Signal to noise ratios for final translation function Z-score (TFZ) and rotation function Z-score (RFZ) as output by Phenix.Phaser for the native dataset.

c) Figure of merit (FOM) value is an output from Phaser (in Phenix.Autosol) for the SeMet dataset.

d) R_{free} sets consist of ~5% of data chosen randomly against which structures were not refined.

Table S2: **Structural Homologs of Msl5-(145-258)**

Protein	pdb ID	Z score	Root mean square deviation
human Quaking (QKI)	4JVH	14.5	2.5 Å at 115 Cα positions
Caenorhabditis elegans GLD-1	4JVY	13.8	2.4 Å at 111 Cα positions
human SF1	1K1G	10.9	3.1 Å at 109 Cα positions
human ZBP1	3KRM	9.4	3.3 Å at 76 Cα positions
Caulobacter crescentus PNPase	4AIM	9.2	3.1 Å at 70 Cα positions
human NOVA-2	1EC6	9.2	2.9 Å at 77 Cα positions
human FUBP-1	4LIJ	8.8	1.9 Å at 69 Cα positions
human hnRNP K	2ZZ1	8.5	2.0 Å at 71 Cα positions

Table S3: **Msl5-(KH-QUA2) contacts to branchpoint RNA pU^{1}pA^{2}pC^{3}pU^{4}pA^{5}pA^{6}pC^{7}p**

Essential amino acids (alanine lethal)	
Arg172	U^{4}pA^5 phosphate via main-chain NH *and* NH1
Arg190	C^7 base N3, O2
Leu259	Cδ2 vdw to C^3 ribose C5'; U^4 ribose O4',C4'
Important amino acids (alanine sick and cold-sensitive)	
Gln255	U^4 base, Oε–>N3 and Cβ vdw to O2
Leu256	vdw to U^1 ribose C1', O4'; C^3 ribose (O4'); U^4 base (N3)
Inessential amino acids (alanine fully viable)	
Asn163	Nδ via water to A^5 base N6, N7 and to C^{3}pU^4 phosphate Oδ via water to A^5 base N6; Nδ vdw to C^{3}pU^4 phosphate main-chain O via water to U^4 base O4
Val165	A^5 base, Cγ–>C5, C6, N1
Leu169	A^6 base, Cδ2–>C4, N9; A^5 base, Cδ1–>N9, ribose O4'
Leu176	A^6 base, vdw Cδ–>C2, N3, vdw Cδ–> A^6 ribose O2'
Arg177	A^6 ribose O2' from Nε, A^{6}pC^7 phosphate from terminal NH2
Lys186	Nζ via waters to C^7 base O2 and C^7 ribose O2'; Cε vdw to C^{7}p phosphate
Ile189	A^6 base, vdw Cβ–>N1; main chain O–>N6, main chain NH–>N1
Lys196	A^5 base, vdw Cγ–>C2; Nζ via water to C^{3}pU^4 phosphate
Pro205	C^7 base C2, N3 (from Cγ); C7 ribose O2' (Cβ)
Lys252	Nζ –> U^1 base O2 and C^3 base O2
Arg253	NH1 to pU^1 phosphate
Thr265	main chain O vdw to A^2 base N7,C8
Arg267	A^2 base; main chain N vdw to N7, main chain O-> N6
Others with contacts	
Gly166	main-chain O to U^4 base N3
Leu167	U^4 base, vdw from Cδ2 to O4
Gly170	main-chain O to U^4 ribose O2'; main-chain N vdw to U^4 ribose O2' Cα vdw to U4 base O2
Pro171	U^4 base O2 vdw from main chain N and Cδ
Ile187	main-chain O via water to C^7 base O2
Lys199	Nζ via waters to A^{5}pA^6 phosphate and A^5 ribose O2'
Gly193	main chain O via water to A^5 base N1
Asn249	U^1 base via main chain NH–>O4
Ala260	A^2 ribose, vdw from Cβ to O2'
Leu266	Cδ1 vdw to A^2 base N6

Table S4: **Oligonucleotide primers for RT-PCR analysis of pre-mRNA splicing**

Primer	Strand	DNA Sequence
SUS1 exon1	Sense	5' TGGATACTGCGCAATTAAAGAGTC
SUS1 exon3	Antisense	5' TCATTGTGTATCTACAATCTCTTCAAG
DYN2 exon1	Sense	5' ATGAGCGATGAAAATAAGAGTACGC
DYN2 exon3	Antisense	5' TTATGCTGTTTTGAAAACTAAAAAC
GLC7 exon1	Sense	5' CCTGGTCAACAAGTTGATCTAGAAG
GLC7 exon2	Antisense	5' CTAAGGATTGTTTACCACGGTCGAC
YFR045w exon1	Sense	5' AAAAAGGGCAGTAAGAACAGCACCC
YFR045w exon2	Antisense	5' ATGATTTCGAAAGCGGTACCCTTCG
PMI40 exon1	Sense	5' ATGTCCAACAAGCTGTTCAGGTTAG
PMI40 exon2	Antisense	5' GGTACCCATCCATAACTCTGCATAT
MATa1 exon1	Sense	5' AAAGTCGACATGGATGATATTTGTAGTATGGCGG
MATa1 exon3	Antisense	5' GGGATCCCTTATTTAGATCTCATACGTTTATTT
MATa1 intron1	Sense	5' CATTTCAAGGATAGCCTTTGAATCA
MATa1 intron2	Antisense	5' GTTAGTATAGGATATATTTAAGTTTGATTCTC