Supplementary Materials for

Structure of the Yeast Spliceosomal Post-catalytic P Complex

Shiheng Liu, Xueni Li, Lingdi Zhang, Jiansen Jiang, Ryan C. Hill, Yanxiang Cui, Kirk C. Hansen, Z. Hong Zhou, Rui Zhao

correspondence to: rui.zhao@ucdenver.edu; Hong.Zhou@UCLA.edu

This PDF file includes:

Other Supplementary Materials for this manuscript includes the following:

Materials and Methods

Plasmids and Yeast strains

Plasmids p358-PRP22, p358-H606A and yeast strain YBST1 (*Mat***a** *ura3-52 trp1-63 his3-Δ200 leu2Δ1 ade2-101 lys2-801 prp22::LEU2 [*p360-Prp22 (*URA3 CEN*)]) were gifts from Dr. Beate Schwer (*8, 39*). CEF1 in YBST1 was C-terminal tagged by 2xprotein A-tag using PCR-based method (*40*), resulting in yeast strain yXL332 (*Mat***a** *ura3-52 trp1-63 his3-Δ200 leu2Δ1 ade2-101 lys2-801 prp22::LEU2 Cef1-2protA::HIS3 [*p360-Prp22 (*URA3 CEN*)]). The coding sequence of calmodulin-binding peptide (CBP) tag was introduced into both p358-PRP22 and p358-H606A using PCR and NEBuilder HiFi DNA Assembly Kit (New England BioLabs). p358-CBP-PRP22 and p358-CBP-H606A plasmids were then transformed into yXL332 yeast strain. After shuffling out the wild-type Prp22 plasmid on FOA plates, the resulting strains yXL351(WT) and yXL352(H606A) were used for spliceosome purification.

Purification of spliceosome

The spliceosome was tandem-affinity purified using a protein A tag on the NTC component Cef1 and a CBP tag on the N-terminus of $Prp22^{H606A}$. Thirty liters of yXL352 yeast cells were grown in YPD medium at 30°C to an OD_{600} of around 4, then shifted to 14°C for 2 hours before harvesting. The cell pellets (~200 g) were re-suspended in 40 mL of lysis buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 0.1% NP-40, 1.5mM $MgCl_2$, 0.5 mM DTT). The cell suspension was snap-frozen into liquid nitrogen to form yeast "popcorn" and cryogenically ground using a SPEX 6870 Freezer/Mill. The frozen cell powder was thawed at room temperature and re-suspended in an additional 400 mL of lysis buffer with protease inhibitor cocktails (Roche) and 1 mM Benzamidine. The cell lysate was first centrifuged at 25,412 x g for 1 hr in a GSA rotor (Sorvall) and the supernatant was further centrifuged at 185,511 x g in a 45Ti rotor (Beckman) for 1.5 hr. The supernatant was incubated with 3 mL of IgG Sepharose-6 Fast Flow resin (GE Healthcare) overnight at 4°C. The resin was washed with IgG washing buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.05% NP40, 0.5 mM DTT, 1 mM Benzamidine and protease inhibitor cocktails), and incubated with TEV protease in 1.5 mL TEV150 buffer (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.02% NP40, 0.5 mM DTT, and 0.2 mM EDTA). The elutes were combined and supplemented with 2.5 mM $CaCl_2$ and incubated with 100 μL calmodulin affinity resin (Agilent) overnight at 4°C. The resin was washed with washing buffer (20 mM Hepes7.9, 100 mM NaCl, 1 mM MgCl2, 2 mM CaCl2, 1 mM imidazole), and eluted 6 times with 100 μL eluting buffer (20 mM Hepes7.9, 100 mM NaCl, 1 mM MgCl2, 2 mM EGTA) each time. The eluate with the optimal concentration was used for sample preparation for cryoEM imaging.

RNA preparation from purified spliceosome and RT-PCR

Spliceosomes were purified from three liters of yXL351(WT) and yXL352(H606A) each, scaling down to an appropriate size following the method described above. RNA from the purified spliceosomal complex was extracted by phenol:chloroform:isopentanol at a volume ratio of 25:24:1. First-strand cDNA synthesis was performed using ProtoScript II First Strand cDNA Synthesis Kit (New England BioLabs) according to the manufacturer's instruction. The reverse transcribed products were used for PCR to detect spliced *ACT1*. Primers used to detect spliced *ACT1* and intron lariat are: ActEF: 5'-

CTCAAACCAAGAAGAAAAAGA-3' (nt -128 to –106 relative to Actin1 ATG); ActER: 5'-TCTCTTGGATTGAGCTTCATCACC-3' (nt 495 to 472 relative to *ACT1* ATG); ActLF: 5'- CACTCTCCCATAACCTCCTA-3' (*ACT1* intron nt 100 to 117); ActLR: 5'- CTCTCGAGCAATTGGGAC-3' (*ACT1* intron nt 69 to 50). The PCR products were resolved on 6% polyacrylamide non-denaturing gel.

CryoEM sample preparation and imaging

For cryoEM sample optimization, an aliquot of 2.5 μl of sample was applied onto a glow discharged lacey carbon film copper grid (300 mesh, Ted Pella). The grid was blotted and flash-frozen in liquid ethane with an FEI Vitrobot Mark IV. An FEI TF20 cryoEM instrument was used to screen grids. Optimized cryoEM grids containing intact spliceosomal complexes with optimal particle distribution and ice thickness were obtained by varying the glow discharge gas source (air or H_2/O_2) and glowing discharge time, applied sample volume, Vitrobot chamber temperature and humidity, Vitrobot bolting time and force. Our best grids were obtained using H_2/O_2 for glow discharge and with the Vitrobot sample chamber set at 12°C temperature and 100% humidity.

Optimized cryoEM grids were loaded into an FEI Titan Krios electron microscope with a Gatan Imaging Filter (GIF) Quantum LS and a post-GIF K2 Summit direct electron detector. The microscope instrument was operated at 300 kV with the GIF energy-filtering slit width set at 20 eV. Movies were acquired with Leginon (*41*) by electron counting in super-resolution mode at a pixel size of 0.68Å/pixel. A total number of 45 frames were acquired in 9 seconds for each movie, giving a total dose of 40.9 $e^-/Å^2$/movie. A total of 5,926 movies were recorded from a single optimized grid chosen for a continuous imaging session spanning a total of 6 days.

Image processing

Frames in each movie were aligned for drift correction with the GPU-accelerated program MotionCor2 (*42*). The first frame was skipped during drift correction due to concern of more severe drift/charging of this frame. Two averaged images, one with dose weighting and the other without dose weighting, were generated for each movie after drift correction. The averaged images were 2x binned to yield a calibrated pixel size of 1.36 Å for the following image processing. The averaged images without dose weighting were used only for defocus determination and the averaged images with dose weighting were used for all other steps of image processing.

The defocus value of each averaged image was determined by CTFFIND4 (*43*), generating values ranging from -1.5 to -3.6 μm. Initially, a total of 753,986 particles were automatically picked from 5,926 averaged images without reference using Gautomatch (http://www.mrc-lmb.cam.ac.uk/kzhang). The particles were boxed out in dimensions of 400 × 400 square pixels square and binned to 200 × 200 square pixels (pixel size of 2.72 Å) before further processing by the GPU accelerated RELION2.1. Several iterations of reference-free 2D classification were subsequently performed using cryoSPARC (*44*) to remove ice, contaminants, and bad particles (i.e., classes with fuzzy or un-interpretable features), yielding 444,761 good particles. These particles were then used to generate an initial model by cryoSPARC. Auto-refinement of these particles by cryoSPARC yields a map with an average resolution of 5.5 Å.

The previously refined 5.5 Å model was low-pass filtered to 60 Å to serve as an initial model for 3D classification. The 3D classification generated 5 classes, and 2 classes showed similar features as C* complex and clear density for Prp22, which were presumed as the P complex. We re-centered the particles from these 2 Classes and removed duplications based on the unique index of each particle given by RELION. 214,396 unique particles (28.4% of all particles) were generated and subjected to 3D auto-refinement by RELION2.1, resulting in a 5.44 Å map. These particles were then un-binned to 400 × 400 square pixels (pixel size of 1.36 Å) and re-centered again. The images with less than 5 particles were removed from the dataset, generating 212,219 particles (28.1% of all particles). Further auto-refinement of particles yielded a map with an average resolution of 3.3 Å, which was selected as the final map for subsequent model building.

All resolutions reported above are based on the "gold-standard" FSC 0.143 criterion (*45*). FSC curves were calculated using soft spherical masks and high-resolution noise substitution was used to correct for convolution effects of the masks on the FSC curves (*46*). Prior to visualization, all maps were sharpened by applying a negative B-factor which was estimated using automated procedures (*9*). Local resolution was estimated using ResMap (*47*). A flow chart depicting the above image processing process and overall quality of the map are presented in **figs. S2 and S3**, respectively. Data collection and reconstruction statistics are presented in **Table S1**.

Model building and refinement

Though the average resolution of 3.3Å of our final map of the P complex is the highest among all known structures of spliceosomal complexes, the local resolutions vary at different areas of the P complex. To generate atomic model for the entire P complex, we had to combine different strategies, including *de novo* model building, homologous structure modeling, and rigid-body docking of known structures.

To aid subunit assignment and model building, we took advantage of the two known yeast C* structures (PDB code: 5WSG, 4.0 Å (*5*); 5MQO, 4.2 Å (*6*)), which were fitted into the P complex density map by UCSF CHIMERA (*48, 49*). The central regions of the P complex have resolutions ranging from 3.0 to 5.0 Å (**fig. S3C**); thus protein and RNA components in these regions were rebuilt manually using COOT (*50*). Briefly, for protein subunits that match well with the densities in the P complex structure, we simply adjusted side chain conformation and, when necessary, main chain coordinates to match the density map. For protein subunits that exhibit substantial main chain mismatches or have not been identified, we built atomic models *de novo*. To do so, sequence assignment was mainly guided by visible densities of amino acid residues with bulky side chains, such as Trp, Tyr, Phe, and Arg. Other residues including Gly and Pro also helped the assignment process. Unique patterns of sequence segments containing such residues were utilized for validation of residue assignment. For the central regions of U2/U5/U6 snRNA and intron (except for the intron stem), we simply adjusted the base conformation and, when necessary, main chain coordinates to match the density map. For the ligated exons (nt -13 to +9 and +15 to +21 with respect to the exon-exon junction), the well-defined nucleotide densities in the central region, along with different sizes of purine and pyrimidine bases, facilitated the RNA model building process. RNA model building in these regions was

performed *de novo* in COOT. The RNA components were subsequently adjusted using RCrane (*51*) and ERRASER (*52*).

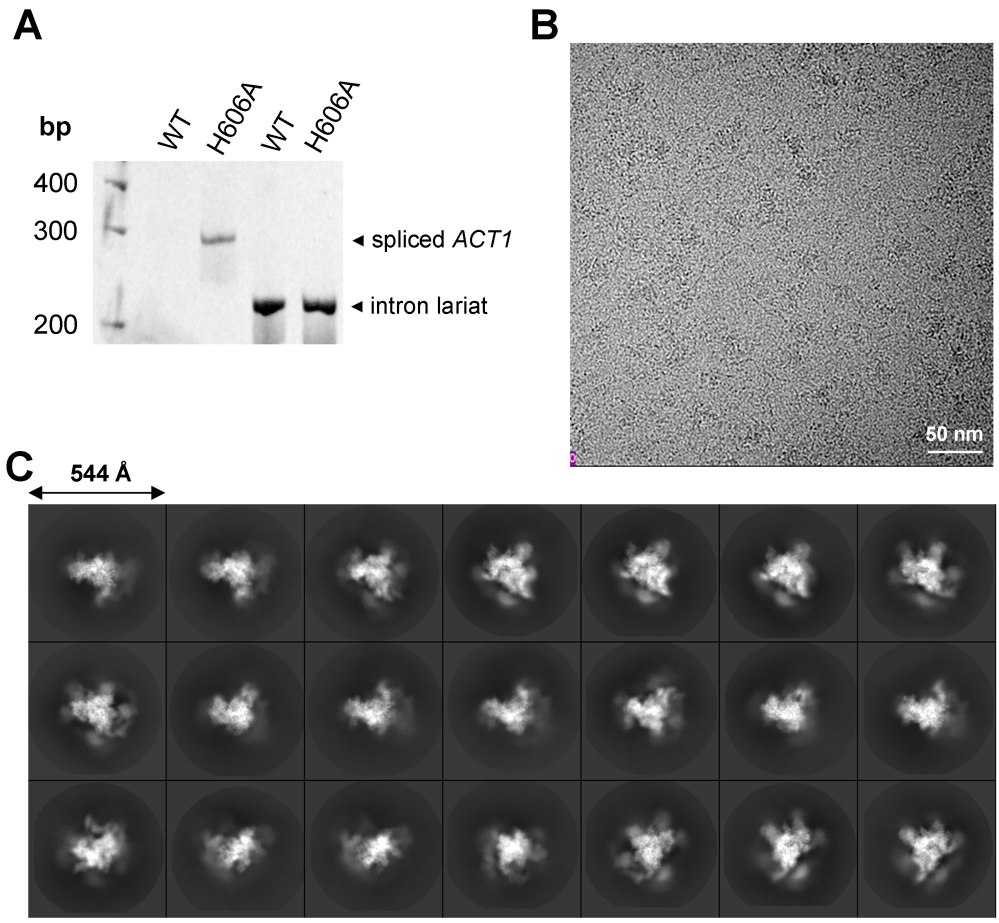
Models built for the protein and RNA subunits in these regions include: Prp8 and Snu114; NTR consisting of Prp45, Prp46, Slt11 (aa. 3-147), Cwc2, Bud31, Cwc15; NTC consisting of Cef1, N-terminal region of Clf1 (aa. 12-291), Syf2, C-terminal region of Syf1 (641-809); splicing factors consisting of Cwc21, C-terminal region of Cwc22 (287-487), Prp17, Slu7, Prp18, the C-terminal tail of Prp22 (aa.1106-1145); the core regions of U2, U5, U6 snRNAs (**Fig. 2**), intron (except for the intron stem), and ligated exons (nt -13 to +9 and +15 to +20 with respect to the exon-exon junction). Because its sequence is unknown, the helices belonging to protein UNK were traced with poly-alanine chains. They interact with Prp22, Prp8, Cef1, and Syf1, which could secure the 3' exon and potentially regulate the activity of Prp22.

Resolution for the periphery of the complex was more varied, ranging from 6 Å to 10 Å (**fig. S3C**), insufficient for *de novo* atomic modeling. Therefore, the following proteins and RNA regions were rigidly docked into the low-pass filtered map of the 3.3 Å map: Slt11(aa. 203-266), C-terminal region of Clf1 (aa. 295-556), N-terminal region of Cwc22 (aa. 14-262), U5 Sm ring, Prp19/Snt309, intron stem. Focused classification of U2 Sm and N-terminal region of Syf1 (aa. 21-634) improved the density enough (**fig. S2**) to enable docking of these proteins from yeast C* structures (PDB code: 5MQO, 4.2 Å) and fitted into the density using CHIMERA (*49*).

Focused classification of Prp22 improved the density so that we could build the atomic model for Prp22 C-terminal domain (CTD, residues 901 to 1100) and trace the main chain for RecA1 and RecA2 domains of Prp22. More importantly, we could trace the main chain for nt +10 to +14 of ligated exons. Combined with the previous atomic model, 21 nucleotides downstream of the exon-exon junction were built, which directly go into the active site of Prp22.

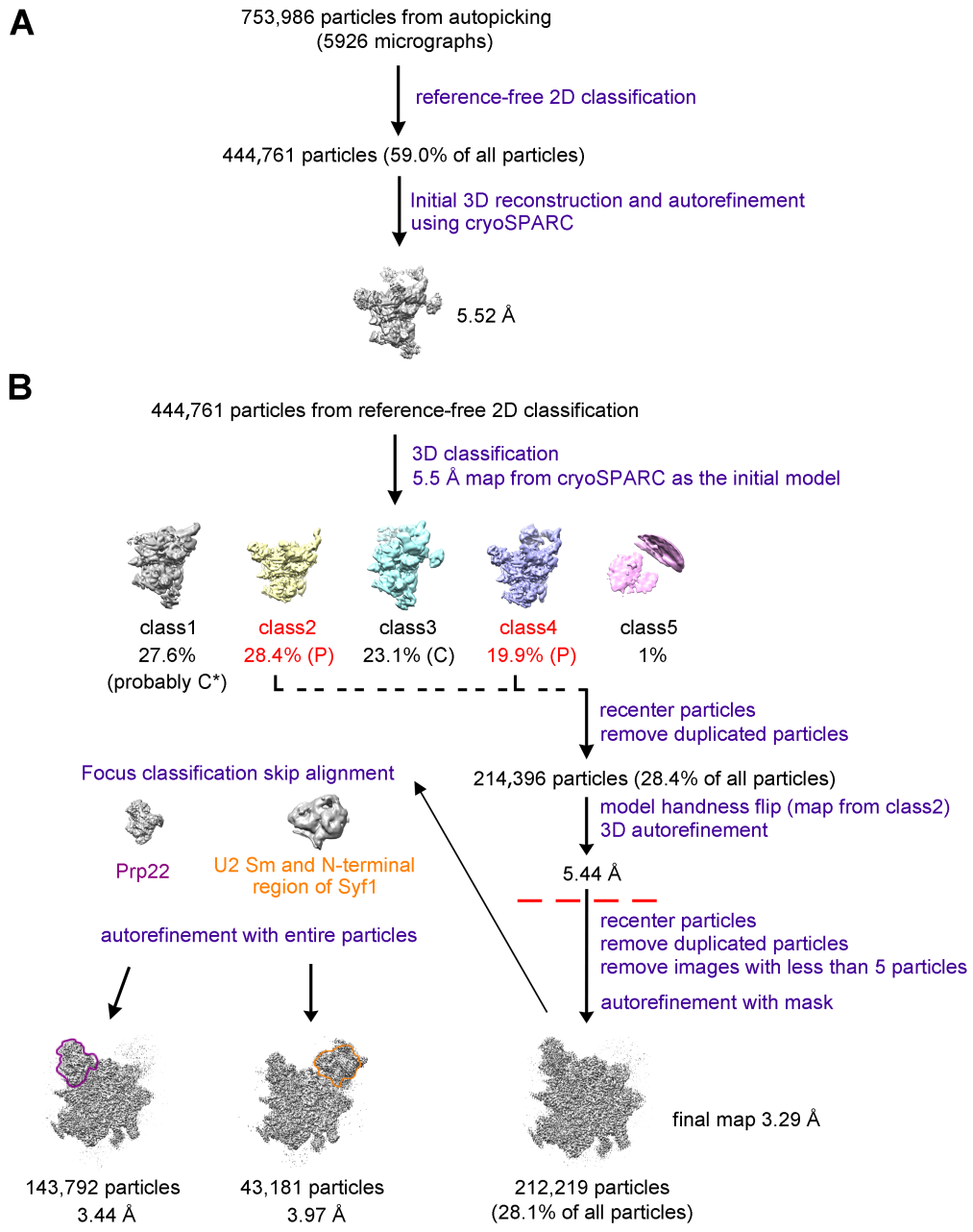
The model was refined using PHENIX in real space (*53*) with secondary structure and geometry restraints. Over-fitting of the overall model was monitored by refining the model in one of the two independent maps from the "gold-standard" refinement using FSC=0.25 criterion (**fig. S3D**). Refinement statistics of the P complex were summarized in **Table S1**. The structure of P complex was also validated through examination of the Morprobity scores (*54*) and Ramachandran plots (**Table S1**). Representative densities for the proteins and RNA are shown in **fig. S4**.

Fig. S1.



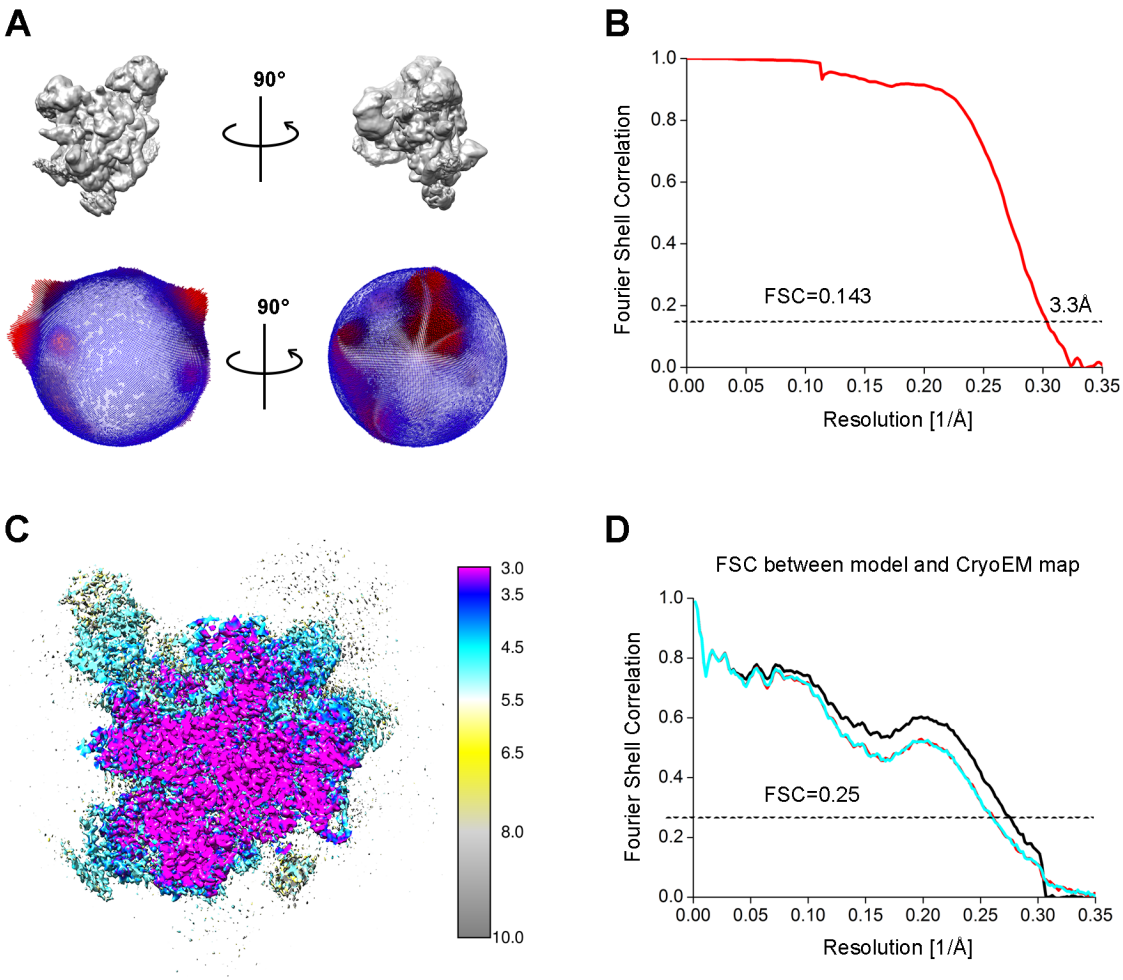
Characterization of the Purified P Complex.
(A) Spliceosome purified from yeast strain carrying the Prp22^{H606A} mutant at non-permissive temperature has substantial accumulation of spliced ACT1 gene compared to spliceosome purified from the WT strain using exactly the same protocol, demonstrated by RT-PCR of RNA extracted from these purified spliceosomes. Both the WT and Prp22^{H606A} spliceosomes contain intron lariat. (B) A drift-corrected cryoEM micrograph of the P complex. (C) Representative 2D class averages of the P complex obtained in RELION.

Fig. S2



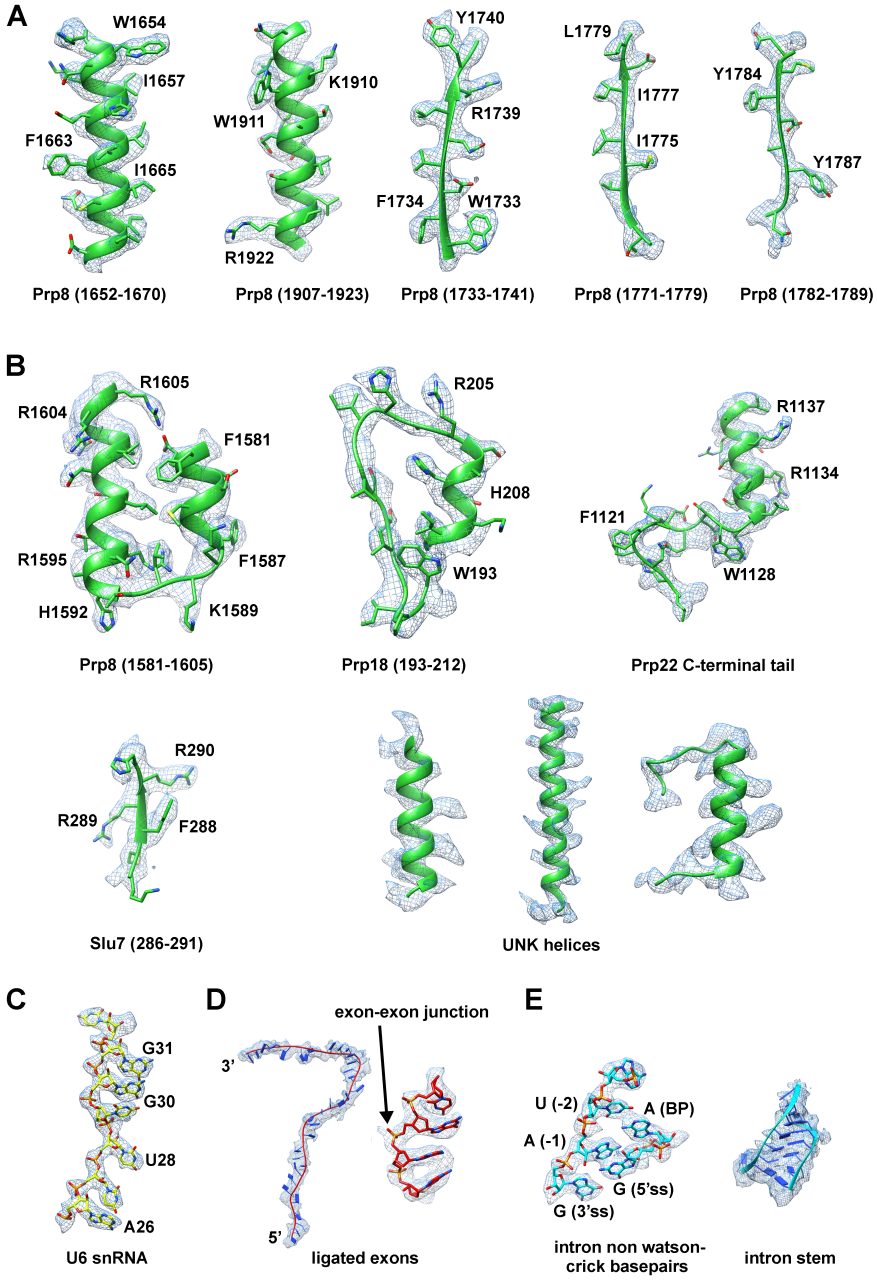
Workflow for the cryoEM Data Processing.
For data processing above the red dash line, the particle images were binned to a pixel size of 2.72 Å. The rest of data processing was performed with a pixel size of 1.36 Å. Please refer to Experimental Procedures for more details.

Fig. S3



Angular Distribution and Fourier shell correlation (FSC) of the P Complex Structure. (A) Angular distribution for all particles used for the final 3.3 Å map of the P complex. (B) FSC as a function of spatial frequency demonstrating the resolution for the final reconstruction of the P complex. (C) Resmap local resolution estimation for the core map. (D) FSC curves of the final refined model (without Prp19 and Snt309) versus the overall 3.3 Å map (black) and the maps used for gold-standard FSC (cyan and red).

Fig. S4

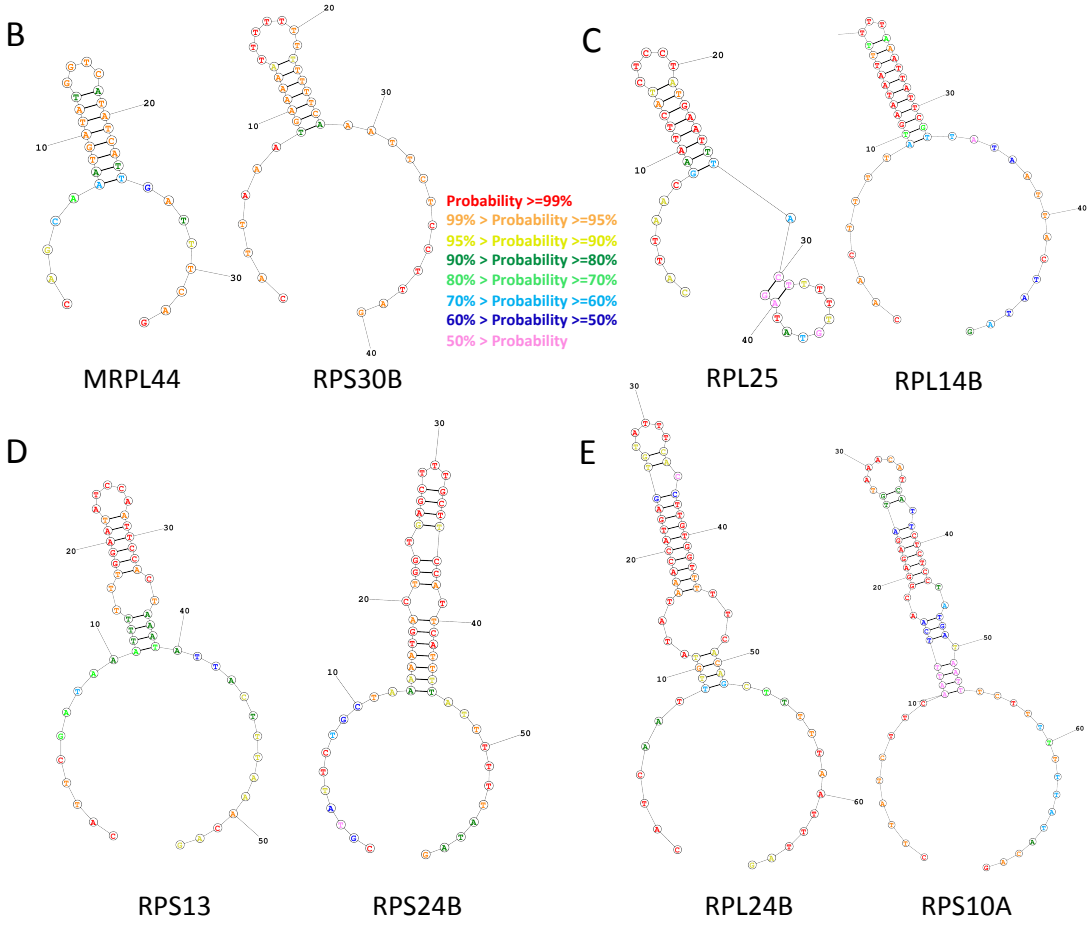


Representative cryoEM Density Maps of the P Complex.
The cryoEM density maps are shown for (A) selected regions of Prp8 that are present in both the C* and P complex structures; (B) selected regions of Prp8, Prp18, Prp22, Slu7, and protein UNK that are present in the P complex structure but absent in the C* complex structure; (C) selected region of U6 snRNA; (D) ligated exon (the map in the left panel is filtered to 5 Å); (E) regions around the 5' ss and 3' ss where non-Watson-Crick interactions are observed as well as the intron stem. The map for the intron stem is filtered to 5 Å.

Fig. S5

A

BP-3'ss distance(nt)	number of introns	Number of introns with stem-like structures between BP and 3' ss
1-20	58	6
21-30	75	32
31-40	82	70
41-50	40	34
51-60	16	16
>60	25	25



Many Yeast Intron-containing Genes Contain Secondary Structures in the region between BP and 3' ss.
(A) Number of introns predicted to have secondary structures in the region between BP and 3' ss using RNAstructure 6.0 (https://rna.urmc.rochester.edu/RNAstructureWeb/). Program NUPACK (www.nupack.org/partition/new) generated similar results. (B-E) Representative secondary structure predictions for introns with BP to 3' ss distances between 31-40 (B), 41-50 (C), 51-60 (D), and over 60 (E).

Table S1.

CryoEM data collection and refinement statistics.

Data collection	
EM equipment	FEI Titan Krios
Voltage (KV)	300
Detector	Gatan K2
Pixel size (Å)	1.36
Electron dose ($e^-/Å^2$)	40.9
Defocus range (μm)	-1.5 ~ -3.6
Reconstruction	
Software	RELION2.1
Number of used particles	212,219
Accuracy of rotation (°)	1.13
Accuracy of translation (pixels)	0.59
Map sharpening B-factors ($Å^2$)	-113.4
Final Resolution (Å)	3.3
Model building	
Software	COOT
Refinement	
Software	PHENIX &
Resolution	3.3
Average Fourier shell correlation	
R-factor	0.35
Model composition	
Protein residues	9433
RNA nucleotides	433
Validation	
R.m.s deviations	
Bonds length (Å)	0.01
Bonds angle (°)	1.31
Ramachandran plot statistics (%)	
Preferred	93.61
Allowed	6.06
Outlier	0.34
Molprobity score	1.63

Refinement was performed for all components except Prp19, Snt309, U5 Sm, U2 Sm, Msl1, and Lea1.

Movie S1

CryoEM structure of the yeast spliceosomal post-catalytic P complex.