

# Crystal structure of the functional domain of the splicing factor Prp18

Jianzhong Jiang\*<sup>†</sup>, David S. Horowitz\*<sup>§</sup>, and Rui-Ming Xu\*<sup>§</sup>

\*W. M. Keck Structural Biology Laboratory, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724; <sup>†</sup>Department of Physics, State University of New York at Stony Brook, Stony Brook, NY 11784; and <sup>§</sup>Department of Biochemistry, Uniformed Services University of the Health Sciences, Bethesda, MD 20814

Communicated by James D. Watson, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, January 13, 2000 (received for review November 18, 1999)

The splicing factor Prp18 is required for the second step of pre-mRNA splicing. We have isolated and determined the crystal structure of a large fragment of the *Saccharomyces cerevisiae* Prp18 that lacks the N-terminal 79 amino acids. This fragment, called Prp18 $\Delta$ 79, is fully active in yeast splicing *in vitro* and includes the sequences of Prp18 that have been evolutionarily conserved. The core structure of Prp18 $\Delta$ 79 is compact and globular, consisting of five  $\alpha$ -helices that adopt a novel fold that we have designated the five-helix X-bundle. The structure suggests that one face of Prp18 interacts with the splicing factor Slu7, whereas the more evolutionarily conserved amino acids in Prp18 form the opposite face. The most highly conserved region of Prp18, a nearly invariant stretch of 19 aa, forms part of a loop between two  $\alpha$ -helices and may interact with the U5 small nuclear ribonucleoprotein particles. The structure is consistent with a model in which Prp18 forms a bridge between Slu7 and the U5 small nuclear ribonucleoprotein particles.

Splicing of mRNA precursors takes place within the spliceosome in two sequential transesterification reactions (reviewed in refs. 1–3). Five small nuclear ribonucleoprotein particles (snRNPs), U1, U2, U4, U5, and U6, participate in the assembly of the spliceosome, but the two catalytic steps of splicing appear to require only the U2, U5, and U6 snRNPs and a large number of proteins. The U2 and U6 snRNAs, together with the pre-mRNA, are thought to form the catalytic core of the spliceosome, whereas the U5 snRNP interacts with exonic sequences adjacent to both splice sites, possibly aligning the exons during splicing.

We focus on the *Saccharomyces cerevisiae* splicing factor Prp18, which is involved only in the second step of splicing (4, 5). Five other yeast proteins—Prp8, Prp16, Prp17, Prp22, and Slu7—are required for the second step of splicing (6–14) (reviewed in ref. 15). Prp18 is a 28-kD protein that lacks known motifs that might suggest how it acts in splicing (16). Prp18 is associated with the U5 snRNP (16, 17) and with the Slu7 protein (18). Slu7 is involved in the selection of 3' splice sites and interacts genetically with the U5 snRNA (13, 19, 20). Prp18 is not absolutely required for splicing *in vitro* or *in vivo*; however, yeast lacking Prp18 splice pre-mRNA extremely slowly and are temperature sensitive (16). *In vitro*, the rate of the second step is reduced nearly 100-fold by depletion of Prp18 (4).

The second step of splicing occurs in two stages. In the first stage, which requires the Prp16 and Prp17 proteins, the hydrolysis of ATP fuels a conformational change in the spliceosome. The transesterification reaction occurs in the second stage, which requires Prp18, Prp22, Slu7, and perhaps Prp8, and is ATP independent (4, 8–10, 19, 21–23). Three of the proteins needed for the second stage—Prp18, Prp22, and Slu7—are dispensable if the distance from the branch point to the 3' splice site is short (10, 18, 20).

Human homologs of all six second-step proteins known in *S. cerevisiae* have recently been identified and at least partly characterized (24–31). The homologs appear to function very similarly to their yeast counterparts, suggesting that the mechanism of the second step has been largely conserved.

Here we report the x-ray crystal structure of a large functional fragment of Prp18 that includes the sequences of Prp18 that have been evolutionarily conserved. The structure provides important insights into the function of Prp18 in the second step.

## Materials and Methods

**Protein Purification and Crystallization.** Limited proteolysis of purified Prp18 identifies a  $\approx$ 20-kDa fragment that resists further digestion by *Staphylococcus aureus* endoproteinase Glu-C (V8 protease) (Calbiochem). N-terminal amino acid sequencing and matrix-assisted laser desorption ionization mass spectrometry showed that the N terminus of the proteolytic fragment was Arg-80, and that the C-terminal end was intact. A recombinant form of this fragment (Prp18 $\Delta$ 79) was produced in *Escherichia coli*. Prp18 $\Delta$ 79 includes the initiating formyl-methionine at position 79. *E. coli* strain BL21(DE3) carrying the expressing plasmid was grown at 37°C until the OD (600 nm) reached 0.8, induced with 0.4 mM isopropyl D-thiogalactoside, and grown at 23°C for 15–18 h.

Prp18 $\Delta$ 79 was purified by ammonium sulfate fractionation hydrophobic (Phenyl-TSK, TosoHaas, Montgomeryville, PA) and cation exchange (Mono S, Pharmacia) chromatography. Concentrated Prp18 $\Delta$ 79 (5–10 mg/ml) was crystallized by the hanging-drop vapor diffusion method at 17°C. The best diffracting crystals were grown with a well solution containing 100 mM Mes pH 6.0, 10% glycerol, and 20% polyethylene glycol 8000. Crystals used for x-ray diffraction typically had dimensions of  $0.1 \times 0.1 \times 0.1$  mm<sup>3</sup>.

**Data Collection and Structure Determination.** The crystal structure was solved by the method of single isomorphous replacement with anomalous scattering. A 2.2-Å resolution native data set and a 2.7-Å Os derivative data set were collected by using a MAR (MAR Research, Hamburg) imaging plate detector at 95 K and  $\lambda = 1.09$  Å. Both data sets were collected at beamline X26C at the National Synchrotron Light Source, Brookhaven National Laboratory (Upton, NY). Derivatized crystals were prepared by soaking native crystals for 12 h in the crystallization-well solution supplemented with 5 mM K<sub>2</sub>O<sub>8</sub>Cl<sub>6</sub>. All raw diffraction data were processed and scaled by using the HKL (32) software packages. The PHASES software packages (33) were used for phasing and for improving the quality of electron density map by solvent flattening (34) and by phase recombination with a partial polyaniline structure. Model building and refitting were performed by using the graphics program o (35).

The structure was refined by using CNS (36). Crystallographic *R* factor and free *R* (calculated with 10% of the full data set

Abbreviation: snRNP, small nuclear ribonucleoprotein particle.

Data deposition: The x-ray diffraction amplitudes and atomic coordinates of the protein structure have been deposited in the Protein Data Bank, [www.rcsb.org](http://www.rcsb.org) (PDB ID code 1dkv).

<sup>§</sup>To whom reprint requests should be addressed. E-mail: dhorowitz@usuhs.mil or xur@cshl.org.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

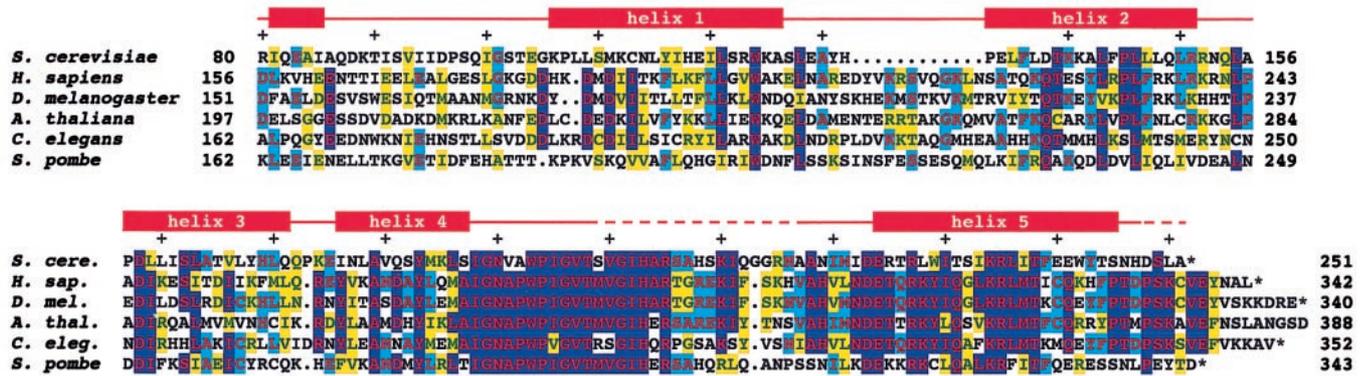


Fig. 1. Alignment of sequences of five Prp18 proteins. *S. cerevisiae* Prp18 $\Delta$ 79 was aligned with the Prp18s of *H. sapiens*, *D. melanogaster*, *A. thaliana*, *C. elegans*, and *S. pombe*. Positions at which four or more sequences are identical are shown red on blue; positions at which three sequences are identical are shown red on cyan; and positions at which three or more sequences are similar are shown green on yellow. Similarities used are D $\approx$ E, K $\approx$ R, N $\approx$ Q, A $\approx$ G, C $\approx$ S $\approx$ T, V $\approx$ L $\approx$ I $\approx$ M, and F $\approx$ Y $\approx$ W. The positions of  $\alpha$ -helices in the *S. cerevisiae* structure are indicated above the sequence. Loops 1 through 5, which precede helices 1 through 5, are indicated by red lines; broken lines identify regions where the protein was disordered in the crystal. The *A. thaliana* Prp18 extends 32 aa beyond the end shown here.

excluded from the refinement process) for the initial model were 43.9% and 43.7%, respectively, using the data in the resolution range of 8.0–3.0 Å. These values dropped to 29.1% and 39.9% after one cycle of torsion-angle simulated annealing carried out with starting temperature at 10,000 K. All data were then included in subsequent refinements with bulk solvent correction. The stereochemistry of the refined model was assessed by the PROCHECK program (37).

**Assay of Prp18 and Prp18 $\Delta$ 79 Activities.** Prp18-depleted yeast extract and hPrp18-depleted HeLa-cell nuclear extract (4, 24) were supplemented with Prp18 or Prp18 $\Delta$ 79, the mixtures were incubated on ice for 30 min, and splicing of actin pre-mRNA (yeast) or  $\beta$ -globin pre-mRNA (human) was carried out for 15 min (yeast) or 60 min (human) (38, 39).

**Sequence Alignments.** Alignments were done by using PILEUP and GAP in GCG (Ver. 10.0). Accession numbers of the sequences are *S. cerevisiae* L03536, *Homo sapiens* U51990, *Arabidopsis thaliana* AC006550 (gene F10O3.3), *Caenorhabditis elegans* Z81074 (gene F32B6.3), and *Schizosaccharomyces pombe* AL034490 (gene SPCC126.14). The *Drosophila melanogaster* sequence was assembled from 10 expressed sequence tag sequences.

## Results and Discussion

**Identification and Sequence Alignment of a Functional Domain of Prp18.** Full-length *S. cerevisiae* Prp18 did not crystallize in our initial attempts. Limited proteolysis of Prp18 (251 amino acids) yielded a stable fragment of  $\approx$ 20 kDa that appeared to represent a single large domain of Prp18. Mass spectrometry and N-terminal sequencing showed that this fragment was the C-terminal two-thirds of Prp18 lacking the first 79 amino acids. A recombinant form of the fragment, called Prp18 $\Delta$ 79, was produced in *E. coli* and used for the crystallographic studies. We show below that Prp18 $\Delta$ 79 is fully functional in pre-mRNA splicing *in vitro*.

An alignment of Prp18 $\Delta$ 79 with Prp18 proteins from five diverse organisms is shown in Fig. 1. The human and yeast Prp18 proteins are known to be functional homologs (24); the other Prp18s were identified by their sequence homology to these two proteins. Similarity is highest in the C-terminal 80 amino acids, and only in this region are the *S. cerevisiae* and *S. pombe* Prp18s strongly homologous to those of the higher eukaryotes. Searches of the recent databases confirm the previous report that the Prp18s do not appear to be related to other proteins (24).

**The Overall Structure of Prp18 $\Delta$ 79.** The recombinant Prp18 $\Delta$ 79 crystallized with a C222<sub>1</sub> symmetry and cell dimensions of  $a = 49.94$  Å,  $b = 79.17$  Å, and  $c = 171.20$  Å. There were two protein molecules per asymmetric unit. The crystal structure was determined by using a single osmium derivative and refined to 2.15 Å resolution ( $R$  factor = 20.3%, and  $R_{\text{free}} = 25.2\%$ ; see Table 1 for detailed statistics). The main chain electron density is continu-

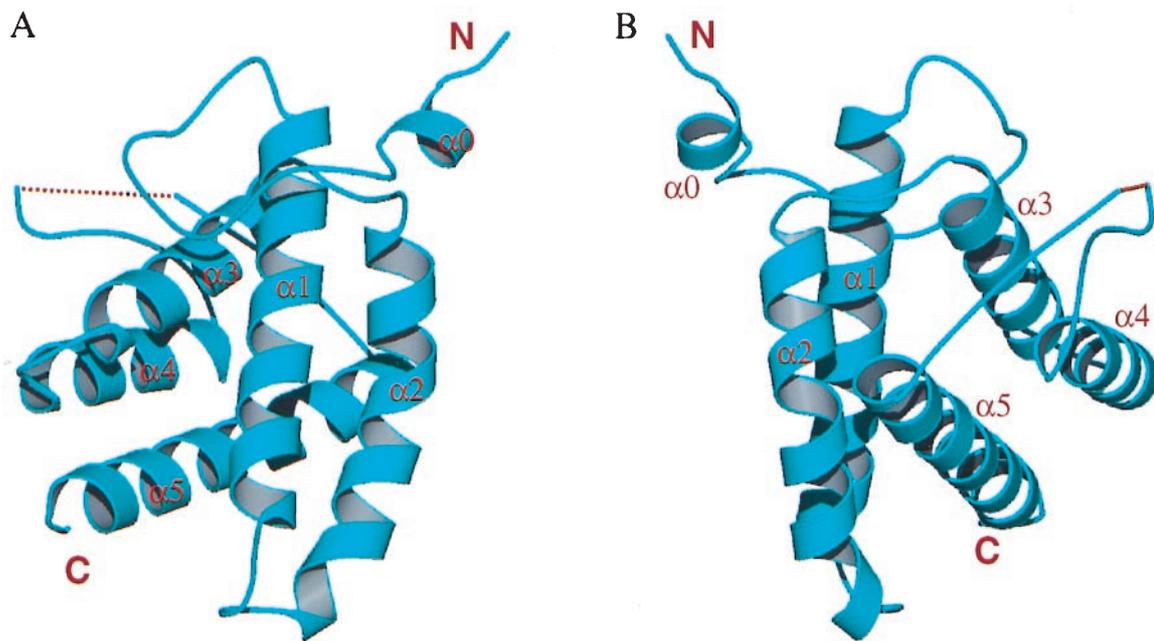
Table 1. Statistics from the crystallographic analysis

	Native	Os derivative
Diffraction data		
Resolution, Å	2.15	2.70
Observed reflections	188,477	69,049
Unique reflections	17,664	9,549
Completeness, %	93.2	99.1
Average $I/\sigma$	16.9	10.8
$R_{\text{merge}}$ , %*	4.1	6.1
Phasing		
Os sites		2
Phasing power <sup>†</sup>		
Isomorphous		2.00
Anomalous		2.44
Overall figure of merit	0.501	
Refinement		
Resolution range, Å	40.0–2.15	
$R$ factor, $R_{\text{free}}$	20.3% (25.2%)	
Number of protein atoms	2,427	
Number of water molecules	332	
Average $B$ -factor	33.2 Å <sup>2</sup>	
rms deviations		
Bond lengths	0.005 Å	
Bond angles	0.99°	
Dihedrals	19.3°	
Improper	0.70°	

\* $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$ , where  $I$  is the measured intensity and  $\langle I \rangle$  is the averaged intensity of multiple measurements of the same reflection. The summation is over all observed reflections.

<sup>†</sup>Phasing power =  $\text{rms}(\langle F_H \rangle / E)$ , where  $F_H$  is the calculated structure factor of the heavy atoms and  $E$  is the residual lack of closure.

<sup>‡</sup> $R$  factor =  $\sum ||F_O| - |F_C|| / \sum |F_O|$ , where  $F_O$  denotes the observed structure factor amplitude and  $F_C$  denotes the structure factor calculated from the model.



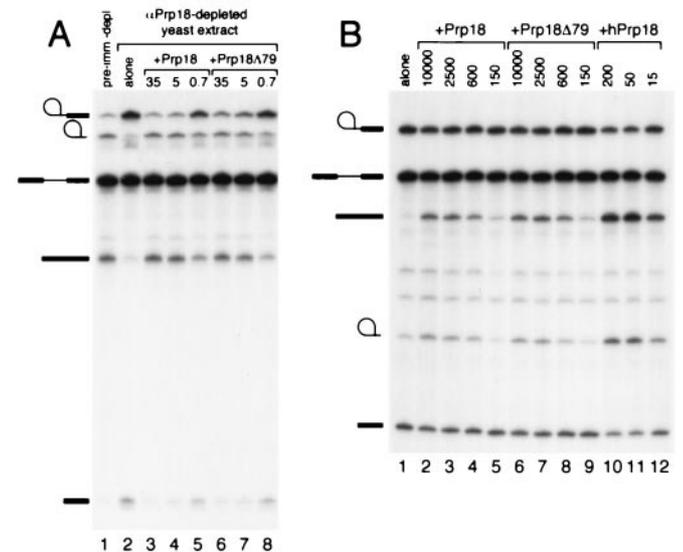
**Fig. 2.** Overall structure of Prp18 $\Delta$ 79. (A) Front view of the Prp18 $\Delta$ 79 structure shown in a ribbon representation. The protein core has a five-helix X bundle structure. The N and C termini and all helices are labeled. The dotted line represents the disordered segment of loop-5. (B) Another view of the Prp18 $\Delta$ 79 structure (rotated 210° with respect to a vertical axis) showing that helices  $\alpha$ 3,  $\alpha$ 4, and  $\alpha$ 5 are antiparallel and located on the same side of the  $\alpha$ 1– $\alpha$ 2 plane. Loop-5 connecting helices  $\alpha$ 4 and  $\alpha$ 5 are exposed. The figure is produced with MOLSCRIPT (40) and RASTER3D (41).

ous except for a 20-aa segment (Thr-198–Ala-217), which was disordered in both of the molecules in the asymmetric unit and could not be modeled reliably. The refined model has good stereochemistry: the Ramachandran plot shows 94.4% of the nonglycine nonproline residues are within the most favored region, 4.8% in the additional allowed regions, and 0.7% in generously allowed regions. The two molecules in the asymmetric unit have essentially identical conformations. The two protein molecules can be superimposed with a rms deviation of 0.55 Å by using C $\alpha$  atoms of well-ordered segments of the polypeptides (amino acids 84–197 and 218–246) for least-squares superimposition.

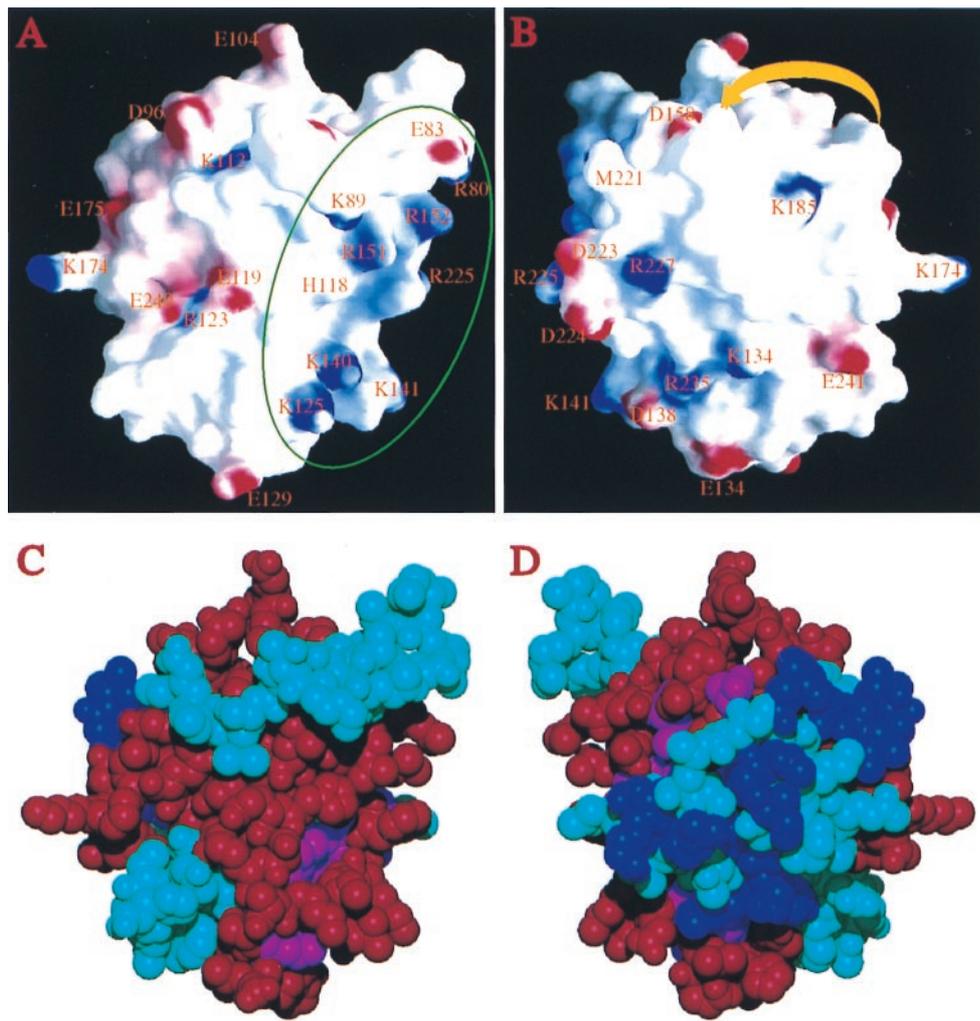
The structure of Prp18 $\Delta$ 79 is shown in Fig. 2. The core structure is formed by five  $\alpha$ -helices, designated  $\alpha$ 1 through  $\alpha$ 5. Helix  $\alpha$ 1 is connected to an additional short N-terminal helix,  $\alpha$ 0, by a 20-residue loop (loop-1; flanking loops are designated 1 through 5). The five helices,  $\alpha$ 1– $\alpha$ 5, pack into a compact globular structure.  $\alpha$ 1 and  $\alpha$ 2 are antiparallel, and they are packed at approximately a 50° angle with respect to  $\alpha$ 3,  $\alpha$ 4, and  $\alpha$ 5 (Fig. 2A). The latter three helices are packed on one side of the  $\alpha$ 1– $\alpha$ 2 plane, and they form a three-helix bundle (Fig. 2B).

Searches for structural similarity by using Prp18 $\Delta$ 79 as a query against protein structure and folding databases resulted in no significant matches. Visual inspection of representative protein structures belonging to the five-helix folds confirmed that Prp18 $\Delta$ 79 has a novel fold. By analogy with the classification of four-helix bundle motifs (42), we term the Prp18 $\Delta$ 79 fold the five-helix X-bundle. Helices  $\alpha$ 1 and  $\alpha$ 2 make direct contacts with  $\alpha$ 3 and  $\alpha$ 5, whereas  $\alpha$ 4 is separated from  $\alpha$ 1 and  $\alpha$ 2 by  $\alpha$ 3 and  $\alpha$ 5 (Fig. 2B). Helix  $\alpha$ 2 is bent in the middle because of the presence of a conserved proline, Pro-145. The short N-terminal helix  $\alpha$ 0 is located on the opposite side of the  $\alpha$ 1– $\alpha$ 2 plane, and it makes little contact with helices  $\alpha$ 1 and  $\alpha$ 2. The N terminus of Prp18 $\Delta$ 79 points away from the core structure, suggesting that the N-terminal 79 amino acids not included in this study should not interfere with the folding of the C-terminal domain, namely Prp18 $\Delta$ 79, in the full-length protein. The conformation of  $\alpha$ 0

observed in the structure might be stabilized by protein–protein interactions in the crystal lattice, as it makes extensive interactions with the same region of symmetry-related molecules.



**Fig. 3.** Activity of Prp18 $\Delta$ 79. (A) Prp18 or Prp18 $\Delta$ 79 was added to Prp18-depleted yeast extract, and splicing of actin pre-mRNA was assayed. Lane 1 shows splicing in extract depleted with preimmune serum. Lane 2 shows splicing in extract depleted by anti-Prp18 with no added Prp18. For the splicing reactions displayed in lanes 3–8, the amount of Prp18 or Prp18 $\Delta$ 79 added is shown at the top of each lane in femtomol per microliter of yeast extract. The positions of lariat intermediate, lariat intron, pre-mRNA, mRNA, and exon1 are shown at the left of both A and B. (B) Prp18, Prp18 $\Delta$ 79, or hPrp18 was added to hPrp18-depleted HeLa-cell nuclear extract, and splicing of  $\beta$ -globin pre-mRNA was assayed. Lane 1 shows splicing in hPrp18-depleted extract alone. For lanes 2–12, the amount of Prp18, Prp18 $\Delta$ 79, or hPrp18 added is shown at the top in femtomol per microliter of nuclear extract.



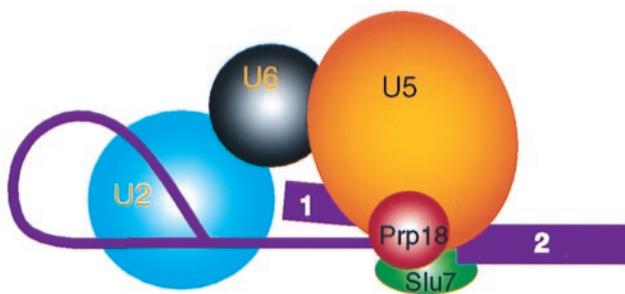
**Fig. 4.** Surface characteristics of the Prp18 $\Delta$ 79 structure. (A) Electrostatic potential on the protein surface, viewed from a similar direction as in Fig. 2A. Positive charges are shown in blue, negative charges in red. Charged residues are labeled by their standard single-letter code followed by the residue number in the sequence. The green circle identifies the proposed Slu7-interacting area. His-118 on  $\alpha$ 2 is also labeled, although histidines are not included in calculating the electric potential on the surface (charges were assigned only to Lys and Arg, Asp and Glu). The surface was calculated by using a probe radius of 1.4 Å, and the potential is displayed at a  $-15 k_B T$  to  $+15 k_B T$  scale, where  $k_B$  is the Boltzmann constant. The region on the right side concentrated with positively charged residues is the putative Slu7-binding region. (B) Rear view (the protein is rotated  $180^\circ$  with respect to a vertical axis). The yellow arrowheaded ribbon indicates schematically the location of the disordered segment in loop-5. The direction of the arrowhead indicates the amino to carboxyl direction of the segment. The figures were generated with the program GRASP (44). (C) Corey–Pauling–Koltun representation of the Prp18 $\Delta$ 79 structure viewed from the same direction as in A. Residues implicated in interacting with Slu7 by the two-hybrid method are colored red; conserved residues (*S. cerevisiae* residues shared by three or more other species as shown in Fig. 1) are colored blue; conserved residues that lie within the Slu7 interacting region are colored magenta, and all others are colored cyan. (D) Rear view. Most invariant residues are located on this side of the surface (the disordered segment of loop-5 is not shown, but it is also expected to be located on this side of the protein). These residues are not involved in interacting with Slu7.

Although there are two Prp18 $\Delta$ 79 molecules per asymmetric unit, only the monomeric form was detected in solution using both dynamic-laser scattering and gel filtration (data not shown). In the Prp18 $\Delta$ 79 crystals, the contact area between the two molecules in the asymmetric unit is rather small; a total area of  $792 \text{ \AA}^2$  is buried by the protein contact, about half that seen for specific and stable protein–protein interactions (43). The data suggest that Prp18 $\Delta$ 79 is a monomer in solution and that the dimeric interaction in the asymmetric unit is most likely a result of crystallization.

**Activity of Prp18 $\Delta$ 79 in pre-mRNA Splicing.** The activity of Prp18 $\Delta$ 79 was tested by complementation of Prp18-depleted extracts. Yeast extract depleted of Prp18 is defective in second-step splicing, and addition of recombinant Prp18 to the depleted extract restores splicing (Fig. 3A, lanes 2–5) (4). Recombinant

Prp18 $\Delta$ 79 also restored full splicing activity to Prp18-depleted yeast extracts (Fig. 3A, lanes 6–8). The activities of Prp18 and Prp18 $\Delta$ 79 were very similar at all concentrations tested. Prp18 $\Delta$ 79 was also assayed in hPrp18-depleted HeLa-cell nuclear extracts (24). Both full-length yeast Prp18 and Prp18 $\Delta$ 79 were active in the complementation assay (Fig. 3B), and their activities were quite similar. Prp18 $\Delta$ 79, therefore, is an evolutionarily conserved functional domain of Prp18. The N-terminal domain of Prp18 does not appear to play a role in splicing *in vitro*, despite some weak conservation of amino acid sequences near the N terminus (not shown) (24).

**Interaction Between Prp18 and Slu7.** Prp18 interacts with Slu7 genetically and physically (8, 13, 18). The crystal structure shows how Slu7 can interact with one positively charged face of Prp18. The region of Slu7 required for interacting with Prp18 has been



**Fig. 5.** A schematic drawing showing the proposed model of the Prp18 function in the second step of splicing. Pre-mRNA is shown in purple; exons 1 and 2 are indicated. Prp18 is shown in red and in contact with both Slu7 (green) and U5 snRNP (orange); the latter two contact the 3' splice site directly. U2 (cyan) and U6 (black) snRNPs are shown to indicate their involvement in the second step splicing reaction.

mapped to a 25-aa segment (amino acids 200–224) by the two-hybrid method (18). This segment is rich in negatively charged residues (a total of 12 charged residues, 9 negative and 3 positive). The portion of Prp18 needed for interaction with Slu7 has been mapped to a fragment encompassing amino acids 98 to 179, also by the two-hybrid assay (D.S.H., T. Niccoli, B. Schwer, and A. R. Krainer, unpublished data). This region encompasses a small part of loop-1, and  $\alpha$ -helices 1, 2, and 3, but does not include the most conserved part of Prp18 (Fig. 1). The charge distribution on the protein surface is shown in Fig. 4*A* and *B*.  $\alpha$ -Helices  $\alpha$ 1 and  $\alpha$ 2 (amino acids 106–126 and 133–151) are contiguous and form a positively charged surface, consistent with a role in interacting with Slu7 (Fig. 4*A*). Helix  $\alpha$ 3 (amino acids 157–171) is almost entirely buried inside Prp18 and does not appear to contact Slu7 directly. Nine positively charged residues are all located on the same face of Prp18 (Fig. 4*A*). Within the mapped interaction region, His-118 and Lys-125 on  $\alpha$ 1, Lys-140 and Lys-141, Arg-151, and Arg-152 on  $\alpha$ 2 are positively charged. In addition, Arg-80 on  $\alpha$ 0, Lys-89 on loop-1, and Arg-225 on  $\alpha$ 5 are located on the same side of the protein surface (Fig. 4*A*). These charged residues are only partly conserved. However, by superimposing the sequences of other Prp18 proteins on the yeast Prp18 structure, we find that all of the Prp18s except that of *S. pombe* have a positively charged surface formed by helices  $\alpha$ 1 and  $\alpha$ 2 (not shown). Interestingly, the region of Slu7 involved in interacting with Prp18 is also not well conserved in other organisms (including humans, *C. elegans*, and *A. thaliana*) except for its negatively charged nature (data not shown; ref. 29). It is worth noting that the *S. pombe* Slu7 differs significantly in the putative Prp18-binding region, which may account for the corresponding differences in the *S. pombe* Prp18. The sequences and structures strongly suggest that charge complementarity is an important part of the Prp18–Slu7 interaction. However, one should bear in mind that the two-hybrid studies of the interaction of Slu7 with Prp18 both used deletions of the proteins; the results, therefore, identify a minimal region needed for interaction, but do not preclude other contacts between the proteins.

**The Conserved Region of Prp18.** The Prp18 proteins are most closely related in their C-terminal 80 amino acids. (Fig. 1). Especially striking is the nearly invariant block of amino acids from 188 through 203 (*S. cerevisiae* numbering). Even *S. pombe* Prp18, the most distantly related of the Prp18s, is highly conserved in this region. There do not appear to be any other proteins that have this sequence, which appears to be the “signature” of Prp18s.

The highly conserved residues in yeast Prp18 are distributed

in a region encompassing the C-terminal end of  $\alpha$ 4, loop-5, and  $\alpha$ 5. This region occupies a large area on the protein surface, and it is spatially distinct from the surface that likely interacts with Slu7. In Fig. 4*C* and *D*, the conserved residues are highlighted in blue, the residues implicated in Slu7 interaction in red, and residues that are both conserved and Slu7-interacting in magenta. The surface formed by these conserved residues is on the opposite face from the putative Slu7-binding area (Fig. 4*D*). This conserved surface can be divided into two parts, one of which is rigid and the other, flexible. Exposed residues located on helices  $\alpha$ 4 and  $\alpha$ 5 form the rigid part. This part of the surface is stable because  $\alpha$ 4 and  $\alpha$ 5 are rigidly embedded in the structure as they are part of the core of the protein. The second, flexible part consists of the 36-aa loop-5, which contains the most conserved residues in Prp18. Part of loop-5, residues 198–217, is disordered and appears to be entirely exposed to the solvent. The ordered portions of loop-5 are also located on the protein surface. The N- and C-terminal portions of loop-5 are stabilized by interaction with residues located on helices  $\alpha$ 3,  $\alpha$ 4 and the C-terminal portion of loop-1. In view of its accessibility and high degree of conservation, loop-5 is likely to function in protein–protein interactions.

The conserved surface of Prp18 is likely to be the site for interaction of Prp18 with other components of the spliceosome. Both the conservation and the flexibility of the conserved surface point to this conclusion; regions of proteins that are involved in ligand binding are often flexible in their unbound form. The conservation of this surface contrasts with the evolutionary divergence of sequences in the Prp18–Slu7 interaction region. Several causes could be responsible for restricting changes in this region: (*i*) The conserved sequences in Prp18 may interact with more than one factor; (*ii*) Prp18 could interact with a factor that is highly conserved; (*iii*) the conserved amino acids may perform a specific function during the second step.

The functional results imply that the conserved portion of Prp18 may be strictly required for activity. Yeast Prp18 can restore splicing activity to HeLa cell nuclear extracts that have been depleted of hPrp18 (24). Yeast Prp18 has much lower affinity for the human spliceosome than does hPrp18 (Fig. 3). However, when sufficient yeast Prp18 is present to drive the binding of the human spliceosome to yeast Prp18, yeast Prp18 is only a few-fold less active in splicing. Yeast Prp18 likely fails to make specific contacts with the human spliceosome; for example, the portion of Prp18 that interacts with Slu7 is not conserved. The activity of yeast Prp18 in human splicing is almost certainly because of the highly conserved C-terminal part of the protein, strongly arguing that this portion of Prp18 is functionally indispensable.

**Interaction Between Prp18 and the U5 snRNP.** Yeast Prp18 is associated with the U5 snRNP, and we surmise that the conserved surface of Prp18 interacts with the U5 snRNP. Immunoprecipitation experiments and, more recently, purification of intact U4/U6·U5 snRNP particles both show that Prp18 is part of the U5 snRNP, although Prp18 binding to U5 is relatively weak, and Prp18 dissociates from the snRNP at moderate salt concentrations (16, 17). Association of Prp18 with the U5 snRNP may be transient; hPrp18 is not stably bound to the U4/U6·U5 snRNP and is stably associated with spliceosomes only during the second step (24). Yeast *PRP18* interacts genetically with loop-I of U5 snRNA (13). Loop-I is highly conserved and interacts with exonic sequences at the 5' and 3' splice sites and is specifically required for the second step of splicing in yeast (45, 46). The biochemical and genetic results show that the U5 snRNP and Prp18 interact, and it seems likely to us that the critical conserved interactions of Prp18 are with the U5 snRNP. The structural data do not suggest how Prp18 interacts with the U5 snRNP.

**A Model of Prp18 Action.** A model of how Prp18 could function in the second step can be proposed based on the crystal structure and the above arguments. The crystal structure reveals that Prp18 has distinct interaction sites, suggesting that Prp18 can interact with the U5 snRNP and Slu7 simultaneously. Both the U5 snRNP and Slu7 interact with the 3' splice site (7, 23, 45, 47); no stable association of Slu7 with U5 snRNP has been reported, although they do interact genetically (13). Prp18 may function as a bridge between Slu7 and U5 snRNP, making their association more stable at the 3' splice site. A schematic drawing of the model is shown in Fig. 5.

The proposed model can account for many of the observed properties of Prp18. Schwer and coworkers have shown that both Prp18 and Slu7 are needed only if the distance between the 3' splice site and the branch point is more than 12 nt (18, 20). When the branch point to 3' splice site distance is short, U5 can bind to the 3' splice site sufficiently tightly to allow splicing. When the distance is large, Slu7 and Prp18 are required to stabilize the binding of U5 to the 3' splice site. Prp18, acting as a bridge between U5 and Slu7, stabilizes the binding of the U5 snRNP to the 3' splice site. If Slu7 is present at high levels, then it could stabilize the binding of the U5 snRNP itself if it makes some contact with the U5 snRNP. This model, in which Prp18

stabilizes a conformation of the spliceosome, is consistent with the temperature sensitivity of *PRP18*-knockout strains.

We have determined the crystal structure of a fully functional portion of yeast Prp18. This domain is conserved from yeast to human, and the structure will be the basis for understanding the function of Prp18 in pre-mRNA splicing. We propose that Prp18 functions to stabilize the interaction between Slu7 and the U5 snRNP at the 3' splice site during the second step of splicing. Only a limited amount of structural information on proteins involved in pre-mRNA splicing is currently available. To our knowledge, the structure of Prp18 $\Delta$ 79 that we report is the first of a splicing factor involved specifically in the catalytic steps of splicing. Further structural information on second-step proteins will allow more precise placement of Prp18 in the spliceosome. This type of structural information will be vital for understanding the mechanism of pre-mRNA splicing.

We thank Ying Zhang for help with protein purification, Ryuji Kobayashi for protein sequencing and analysis with mass spectrometry, and Michael Murray, Beate Schwer, and Adrian Krainer for comments on the manuscript. This work was supported in part by the W. M. Keck foundation and by National Institutes of Health (NIH) Grant GM55874 to R.M.X. and by Uniformed Services University of the Health Sciences Grant C071EJ and NIH Grant GM57267 to D.S.H.

- Krämer, A. (1996) *Annu. Rev. Biochem.* **65**, 367–409.
- Staley, J. P. & Guthrie, C. (1998) *Cell* **92**, 315–326.
- Burge, C. B., Tuschl, T. H. & Sharp, P. A. (1999) in *RNA World II*, eds Gesteland, R. F., Cech, T. R. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 525–560.
- Horowitz, D. S. & Abelson, J. (1993) *Genes Dev.* **7**, 320–329.
- Vijayraghavan, U. & Abelson, J. (1990) *Mol. Cell. Biol.* **10**, 324–332.
- Umen, J. G. & Guthrie, C. (1995) *Genes Dev.* **9**, 855–868.
- Teigelkamp, S., Whittaker, E. & Beggs, J. D. (1995) *Nucleic Acids Res.* **23**, 320–326.
- Jones, M. H., Frank, D. N. & Guthrie, C. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 9687–9691.
- Schwer, B. & Guthrie, C. (1991) *Nature (London)* **349**, 494–499.
- Schwer, B. & Gross, C. H. (1998) *EMBO J.* **17**, 2086–2094.
- Collins, C. A. & Guthrie, C. (1999) *Genes Dev.* **13**, 1970–1982.
- Siatecka, M., Reyes, J. L. & Konarska, M. M. (1999) *Genes Dev.* **13**, 1983–1993.
- Frank, D., Patterson, B. & Guthrie, C. (1992) *Mol. Cell. Biol.* **12**, 5197–5205.
- Vijayraghavan, U., Company, M. & Abelson, J. (1989) *Genes Dev.* **3**, 1206–1216.
- Umen, J. G. & Guthrie, C. (1995) *RNA* **1**, 869–885.
- Horowitz, D. S. & Abelson, J. (1993) *Mol. Cell. Biol.* **13**, 2959–2970.
- Gottschalk, A., Neubauer, G., Banroques, J., Mann, M., Lührmann, R. & Fabrizio, P. (1999) *EMBO J.* **18**, 4535–4548.
- Zhang, X. & Schwer, B. (1997) *Nucleic Acids Res.* **25**, 2146–2152.
- Frank, D. & Guthrie, C. (1992) *Genes Dev.* **6**, 2112–2124.
- Bryson, A. & Schwer, B. (1996) *RNA* **2**, 707–717.
- Schwer, B. & Guthrie, C. (1992) *EMBO J.* **11**, 5033–5039.
- Ansari, A. & Schwer, B. (1995) *EMBO J.* **14**, 4001–4009.
- Umen, J. G. & Guthrie, C. (1995) *RNA* **1**, 584–597.
- Horowitz, D. S. & Krainer, A. R. (1997) *Genes Dev.* **11**, 139–151.
- Ono, Y., Ohno, M. & Shimura, Y. (1994) *Mol. Cell. Biol.* **14**, 7611–7620.
- Zhou, Z. & Reed, R. (1998) *EMBO J.* **17**, 2095–2106.
- Ortlepp, D., Lagerbauer, B., Mullner, S., Achsel, T., Kirschbaum, B. & Lührmann, R. (1998) *RNA* **4**, 1007–1018.
- Ben Yehuda, S., Dix, I., Russell, C. S., Levy, S., Beggs, J. D. & Kupiec, M. (1998) *RNA* **4**, 1304–1312.
- Chua, K. & Reed, R. (1999) *Genes Dev.* **13**, 841–850.
- Lindsey, L. A. & Garcia-Blanco, M. A. (1998) *J. Biol. Chem.* **273**, 32771–32775.
- Luo, H. R., Moreau, G. A., Levin, N. & Moore, M. J. (1999) *RNA* **5**, 893–908.
- Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **277**, 307–326.
- Furey, W. & Swaminathan, S. (1997) *Methods Enzymol.* **277**, 590–620.
- Wang, B.-C. (1985) *Methods Enzymol.* **115**, 90–112.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47**, 110–119.
- Brünger, A. T., Adams, P. D., Clore, G. M., Gross, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., et al. (1998) *Acta Crystallogr. D* **54**, 905–921.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993) *J. Appl. Crystallogr. A* **42**, 140–149.
- Lin, R.-J., Newman, A. J., Cheng, S.-C. & Abelson, J. (1985) *J. Biol. Chem.* **260**, 14780–14792.
- Mayeda, A. & Krainer, A. R. (1999) in *RNA-Protein Interactions*, ed. Haynes, S. R. (Humana, Totowa, NJ), pp. 315–321.
- Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946–950.
- Merritt, E. A. & Bacon, D. J. (1997) *Methods Enzymol.* **277**, 505–524.
- Harris, N. L., Presnell, S. R. & Cohen, F. E. (1994) *J. Mol. Biol.* **236**, 1356–1368.
- Janin, J. & Choithia, C. (1990) *J. Biol. Chem.* **265**, 16027–16030.
- Nicholls, A., Sharp, K. & Honig, B. (1991) *Proteins* **11**, 281–296.
- Newman, A. J. & Norman, C. (1992) *Cell* **68**, 743–754.
- O'Keefe, R. T., Norman, C. & Newman, A. J. (1996) *Cell* **86**, 679–689.
- Newman, A. J., Teigelkamp, S. & Beggs, J. D. (1995) *RNA* **1**, 968–980.