Crystal structure of the human ATP-dependent splicing and export factor UAP56

Hang Shi*, Olivier Cordin†‡, C. Michael Minder*, Patrick Linder†, and Rui-Ming Xu*§

*W. M. Keck Structural Biology Laboratory, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724; †Département de Microbiologie et Médecine Moléculaire, Centre Medical Universitaire, 1, Rue Michel Servet, 1211 Geneva 4, Switzerland; and ‡Centre de Génétique Moléculaire, Centre National de la Recherche Scientifique, Avenue de la Terrasse, Gif-sur-Yvette 91198, France

Pre-mRNA splicing requires the function of a number of RNA-dependent ATPases/helicases, yet no three-dimensional structure of any spliceosomal ATPases/helicases is known. The highly conserved DECD-box protein UAP56/Sub2 is an essential splicing factor that is also important for mRNA export. The expected ATPase/helicase activity appears to be essential for the UAP56/Sub2 functions. Here, we show that purified human UAP56 is an active RNA-dependent ATPase, and we also report the crystal structures of UAP56 alone and in complex with ADP, as well as a DECD to DEAD mutant. The structures reveal a unique spatial arrangement of the two conserved helicase domains, and ADP-binding induces significant conformational changes of key residues in the ATP-binding pocket. Our structural analyses suggest a specific protein-RNA displacement model of UAP56/Sub2. The detailed structural information provides important mechanistic insights into the splicing function of UAP56/Sub2. The structures also will be useful for the analysis of other spliceosomal DExD-box ATPases/helicases.

helicase | RNA processing | export | crystallography

The spliceosome is a complex molecular machine responsible for the removal of noncoding introns and the joining of exons. Pre-mRNA splicing also plays important roles in subsequent cellular processes such as mRNA export, translation, and mRNA degradation by altering the composition of ribonucleo-protein complexes assembled on spliced mRNAs (1, 2). A number of ATP-using enzymes containing the characteristic DExD sequence motif are required for the assembly, remodeling, and disassembly of the spliceosome (3, 4).

Human UAP56 (56-kDa U2AF-associated protein) and its yeast homolog Sub2 are essential DECD-box splicing factors (5–8). UAP56 is required for the association of U2 small nuclear ribonucleoprotein with pre-mRNA (5), and Sub2 has been implicated in both ATP-independent and -dependent steps of prespliceosome assembly (6, 7). Interestingly, deletion of Mud2, the yeast homolog of U2AF65, can bypass the requirement of Sub2 (6). It has been proposed that splicing may occur through Sub2–Mud2 dependent and Sub2–Mud2 independent pathways. A possible function of Sub2 is to displace Mud2 and/or SF1, a branchpoint binding protein, from pre-mRNA before the binding of U2 small nuclear ribonucleoprotein. The functions of Sub2 require intact ATPase/helicase motifs, suggesting that ATP hydrolysis is essential for its activities (8).

UAP56/Sub2 also plays important roles in the export of mRNA from the nucleus to the cytoplasm (9–12). Reduction of cellular UAP56 levels by RNA interference in *Drosophila* or *Caenorhabditis elegans* resulted in the retention of significant fractions of mRNAs in the nucleus (13–15). UAP56/Sub2 couples transcription to mRNA export in the context of a transcription–export (TREX) complex, which is recruited to activated genes and travels the entire length of the gene with RNA polymerase II during transcription (16).

Here, we report the crystal structures of UAP56 alone and in complex with ADP as well as the structure of a DECD to DEAD

mutant in motif II. The structures provide mechanistic insights into the domain organization, ADP binding, and ATP hydrolysis of UAP56. It also allows us to propose a specific RNA-dependent "unwindase" model that UAP56/Sub2 may use in pre-mRNA splicing (17, 18). The UAP56 structure also contributes to the understanding of RNA-dependent ATPase/helicases in general, because to date only two bona fide RNA-dependent ATPase/helicase structures have been determined (19–21), neither of which is involved in pre-mRNA splicing or mRNA export.

Materials and Methods

Protein Preparations. cDNA fragments encoding the full-length, UAP56ΔN (amino acids 44–428) and a C198A mutant were generated by PCR and cloned into a pGEX vector. GST-fused recombinant proteins were expressed in *Escherichia coli* at 17°C and first purified on a glutathione-Sepharose column, followed by in-column cleavage of the GST tag with thrombin. The proteins were purified further by using a HighTrap Q Sepharose column and a Superdex-75 gel-filtration column (Amersham Pharmacia Biotech). Purified protein was concentrated to ≈20 mg/ml in a buffer containing 10 mM Hepes (pH 8.0), 100 mM NaCl, 1 mM EDTA, 1 mM EGTA, 5% glycerol, and 0.2% 2-mercaptoethanol. The stable UAP56ΔN fragment was identified by limited proteolysis of the full-length protein with trypsin or elastase. A time course with an ≈1:100 (wt/wt) protease-to-UAP56 ratio revealed stable fragments of ≈40 kDa. Protein sequencing reveals that trypsin removes 35 residues and elastase removes 42 amino acids at the N terminus.

ATPase Assay. The ATPase activity was monitored by using a colorimetric assay based on malachite green–molybdate as described in ref. 22, with minor modifications. The reaction buffer contained 50 mM potassium acetate, 50 mM Hepes (pH 6.5), 2 mM magnesium acetate, 2 mM dithiothreitol, and 0.1 mg/ml BSA. Reactions were incubated at 37°C in a 25-μl volume for 40 min and stopped with the addition of 5 μl of EDTA. The optimum conditions for the activity were at pH 5.5–6.5 and 2 mM magnesium concentration. The ATPase activity was ≈30% lower at conditions with pH >6.5 or magnesium concentrations >2 mM. For reactions with varying protein concentrations, 1 mM ATP (Amersham Pharmacia Biotech) and 400 ng/μl yeast RNA (Type III, Sigma) were used. Protein was incubated with RNA before adding ATP. For reactions with varying RNA concentrations, different amounts of RNA were added to the reaction mixture together with 300 ng/μl protein, followed by the addition of 1 mM ATP to start the reaction. Curve fits were

---

Table 1. Summary of crystallographic analysis

Data sets	UAP56ΔN	UAP56ΔN-ADP	UAP56ΔN (C198A)
Spacegroup	P1	P2$_1$	P1
Cell dimensions, Å	37.16 × 49.91 × 62.27	37.00 × 78.09 × 63.19	37.43 × 49.76 × 62.81
Cell angles, °	95.7 × 101.6 × 111.1	90 × 103.4 × 90	95.6 × 101.9 × 110.9
No. of proteins per asu,* V_m	1 (2.33 Å^3/Da)	1 (1.98 Å^3/Da)	1 (2.36 Å^3/Da)
Resolution, Å	1.95	2.7	2.4
Measured reflections	107,519	27,997	55,301
Unique reflections	28,304	9,164	14,805
Average I/σ	19.7	15.0	20.9
Completeness, % ($I/\sigma \geq 0$)	96.1 (87.3)	94.6 (89.1)	92.9 (75.6)
R_{merge},[†] %	4.0 (18.0)	8.5 (14.1)	6.0 (10.1)
Refinement			
Resolution range, Å	50.0–1.95	50.0–2.8	50.0–2.4
R factor/R_{free},[‡] %	21.8/25.8	21/30	24.3/29.8
No. of protein atoms	3,089	3,017 (1 MgADP)	3,095
No. of solvent molecules	4 isopropanol, 245 water	1 acetate, 85 water	1 BME, 59 water
rms deviations			
Bond lengths, Å	0.008	0.007	0.008
Bond angles, °	1.22	1.35	1.31

*asu, asymmetric unit; V_m, Matthews coefficient.

[†]$R_{merge} = \Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$, where I and $\langle I\rangle$ are the measured and averaged intensities of multiple measurements of the same reflection. The sum is over all the observed reflections.

[‡]R factor $= \Sigma\|F_o| - |F_o\|/\Sigma |F_o|$, where F_o denotes the observed structure factor amplitude and F_C denotes the structure factor calculated from the model.

obtained by using the GOGRAPH PRISM 4 (www.graphpad.com) software from three independent reactions.

Crystallographic Methods. UAP56ΔN was crystallized by hanging drop vapor diffusion in a condition containing 100 mM Hepes (pH 7.5), 5% isopropanol, and 3% 2-mercaptoethanol. Macroseeding improved the crystal size. The C198A mutant was crystallized under the same condition. The UAP56ΔN–MgADP complex was crystallized in the presence of 2 mM ADP and in a condition containing 50 mM ammonium acetate, 10 mM MgCl$_2$, 50 mM Tris (pH 7.5), and 10% 2-methyl-2,4-pentanediol. All diffraction data were collected at 100 K by using a charge-coupled device detector (Area Detector Systems, Poway, CA) at the X26C and X12C beamlines of the National Synchrotron Light Source (Brookhaven National Laboratory, Upton, NY). All data were processed by using HKL (23), and statistics are shown in Table 1.

The structure of UAP56ΔN was solved by molecular replacement with AMORE (24), by using the polyalanine models derived from the N- and C-terminal domains of yeast eIF4A [Protein Data Bank (PDB) ID codes 1FUK and 1QDE]. Refinement was carried out by using CNS (25). The graphics program O (26) was used for model building, structural alignment, and modeling. UAP56ΔN (C198A) and UAP56ΔN–MgADP structures were solved by molecular replacement with the UAP56ΔN structure as the search model. Refinement statistics are shown in Table 1. Figures were prepared with the programs MOLSCRIPT (27), RASTER3D (28), GRASP (29), and PYMOL (30).

Results and Discussion

Conservation and RNA-Dependent ATPase Activity. Human UAP56 has 428 aa, and it shares 62% sequence identity and 78% similarity with its yeast counterpart, Sub2 (Fig. 1A). The first ≈40 residues of UAP56 are not highly conserved between UAP56 and Sub2. Limited proteolytic digestion of purified UAP56 with elastase removed N-terminal 42 residues (data not shown). A UAP56 fragment lacking 43 residues at the N terminus, UAP56ΔN, was used for structural studies here. As shown in Fig. 1A, UAP56ΔN contains all of the conserved sequence motifs common to the DExD/H family of ATPases/

helicases (31), as well as the recently identified Q motif specific to the DEAD subfamily (22).

The ATPase/helicase activities of UAP56 or its homologs had not been demonstrated before this study. Here, we show that the full-length UAP56 and UAP56ΔN are active, albeit weak, RNA-dependent ATPases *in vitro* (Fig. 1 B and C). In the absence of RNA, both UAP56 and UAP56ΔN had negligible ATPase activities (Fig. 1B). The ATPase activities were stimulated greatly with the addition of total yeast RNAs (Fig. 1 B and C). UAP56ΔN has a reduced activity compared with that of the full-length, wild-type protein (≈50%). A C198A mutation, which changes the DECD box to a DEAD box, both in the context of full-length UAP56 and in UAP56ΔN, did not significantly change the ATPase activities in our assay (Fig. 1 B and C).

Overall Structure. The apo-UAP56ΔN structure was solved by molecular replacement using polyalanine models derived from the N- and C-terminal domain structures of yeast eIF4A (19). The refined 1.95-Å UAP56ΔN structure has an R factor of 21.8% and R_{free} of 25.8%. The structure also has good stereochemical properties, with 89.4% of the residues in the most favored region of the Ramachandran plot, calculated with PROCHECK (32). Detailed statistics of the crystallographic analyses are shown in Table 1.

UAP56ΔN comprises two canonical, RecA-like α/β helicase domains (33) (Fig. 2A). The N-terminal domain contains a seven-stranded parallel β-sheet in the spatial order of β7–β1–β6–β5–β2–β4–β3. Eight α-helices packed against the β-sheet on both sides as follows: αA, αB, αC, and αD on one side, and αE, αF, αG, and αH on the other side (Fig. 2A). The nomenclature for secondary structure elements is defined in Fig. 1A. The C-terminal domain also contains seven parallel β-strands, forming a twisted β-sheet in the spatial order of β10–β11–β9–β12–β13–β8–β14. There are five helices in the C-terminal domain as follows: αI and αJ on one side of the β-sheet, and αK, αL, and αM on the opposite side. The overall folds of the N- and C-terminal domains resemble that of other superfamily II (SF2) ATPases/helicases (34).

A structure-homology search on the Dali server (www.ebi.ac.uk/dali) revealed that the closest matches to the N- and
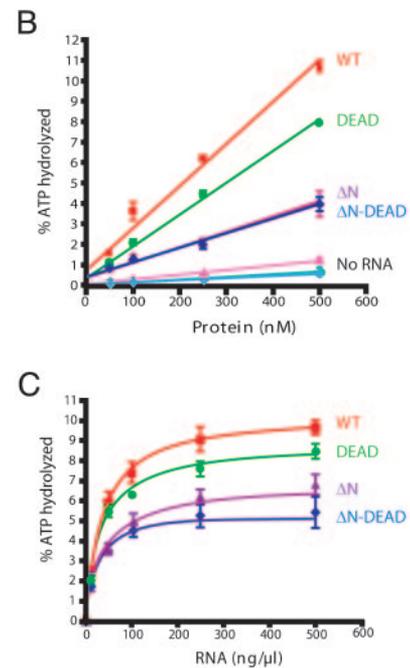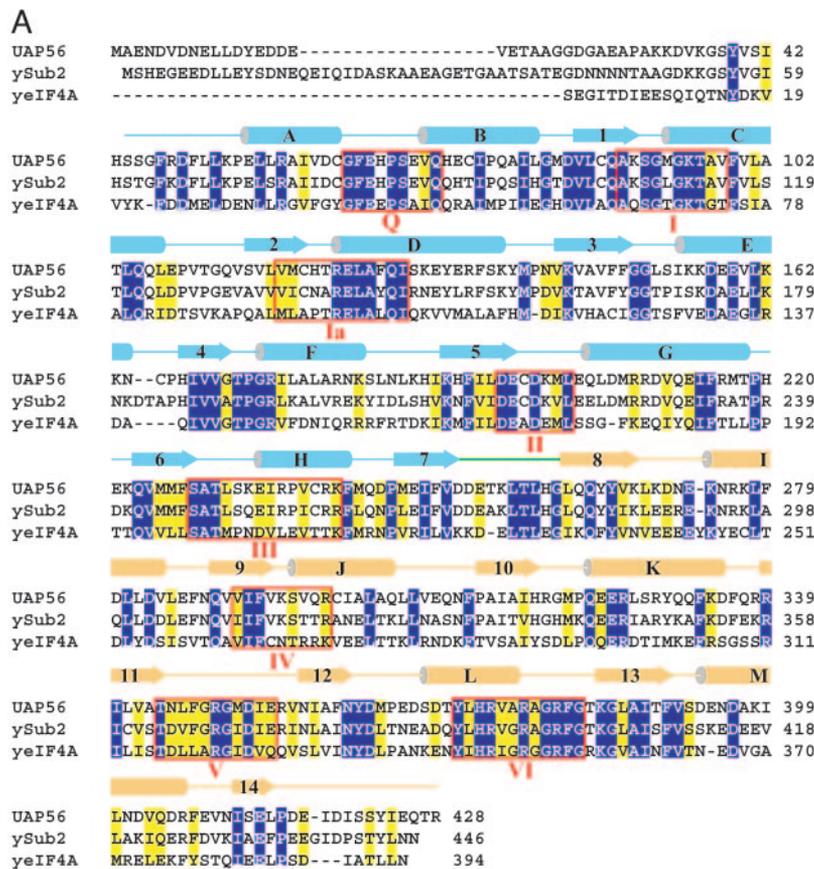
Fig. 1 (A) Sequence alignment

```
             A                    B              1            C
UAP56    MAENDVDNELLDYEDDE------------------VETAAGGDGAEAPAKKDVKGSYVSI  42
ySub2    MSHEGEEDLLEYSDNEQEIQIDASKAAEAGETGAANTSATEGDNNNNTAAGDKVGSYVGI  59
yeIF4A   ----------------------------------------SEGITDIEESQIQTNMYDKV  19

UAP56    HSSGFRDFLLKPELRAIVDCGFEHPSEVQHECIPQAILEMDVICQAKSGMGKTAVFVLA 102
ySub2    HSTGFKDFLLKPELSRAIIDCGFEHPSEVQHTIPQSIHGTDVICQAKSGLGKTAVFVLS 119
yeIF4A   VYK-FDMBELDENLRGVFGYGPEEPSAIQRAMPMPIEGHDVIACAQSGTGKTATSIA    78
                 Q                                      I
           2       D             3              E
UAP56    TLQQLEPVTGQVSVLVMCHTRELAFQISKEYERFSKYMPNVKVAVFFGGLSIKKDEEVLK 162
ySub2    TLQQLDPVPGEVAVVVICNARELAYQIRNEYLRFSKYMPDVKTAVFYGGTPISKDAEPLK 179
yeIF4A   ALQRIDTSVKAPQALMLAPTRELALQIQKVVMALAFHM-DIRVHACIGGTSFVEDAEGLR 137
               II

           4       F             5             G
UAP56    KN--CPHIVVGTPGRILALARNKSLNLKHIKHFILDECDKMLEQLDMRRDVQEIFRMTEH 220
ySub2    NKDTAPHIVVATPGRLKALVREKYIDLSHVKNPVIDECDKVLEELDMRRDVQEIFRATER 239
yeIF4A   DA----QIVVGTPGRVFDNIQRRRFRTDKIKMFILDEADEMLSSG-FKEQIYQIFTLLP  192
                               II

           6       H             7         8             I
UAP56    EKQVMMFSATLSKEIRPVCRKMQDEMEIFVDDETKLTLHGLQQYYVKLKDNE-KNRKLF 279
ySub2    DKQVMMFSATLSQEIRPICRRLQNELEIFVDDEAKLTLHGLQQYYIKLEERE-KNRKLA 298
yeIF4A   TTQVLLSATMPNDVLEVTTKFMRNEPVRILVKKD-ELTLEGIKQFYVNVEEEYKYECFT 251
                    III

           9       J           10             K
UAP56    DLLEVLEFNQVVIFVKSVQRCIALAQLLVEQNFPAIAIHRGMPQEERLSRYQQFKDFQRR 339
ySub2    QLLDLEFNQVIIFVKSTTRANELTKLLNASNFPAITVHGHMKQEERIARYKAFKDFEKR 358
yeIF4A   DLYESISVTQAVIFCNTRRKVEELTTKLRNDKTVSAIYSDLPQQERDTIMKEFRSGSSR 311
                 IV

          11         12             L         13             M
UAP56    ILVATNLFGRGMDIERVNIAFNYDMPEDSDTYLHRVARAGRFGTKGLAITFVSDENDAKI 399
ySub2    ICVSTDVFGRGIDIERINLAINYDLTNEADLLLHRVGRAGRAGTKGLAISFVSSKEDEEV 418
yeIF4A   ILISTDLLARGIDVQQVSLVINYDLPANKENYIHRIGRGGRFGRKGVAINVVTN-EDVGA 370
                V                              VI

          14
UAP56    LNDVQDRFEVNISELPLE-IDISSYIEQTR 428
ySub2    LAKIQERFDVKIAEFPEEGIDPSTYLNN    446
yeIF4A   MRELEKFYSTQIEPLKSD---IATLLN     394
```

Fig. 1. Sequence conservation and ATPase activity of UAP56. (*A*) Alignment of human UAP56, yeast Sub2, and eIF4A sequences (white over blue, identical; black over yellow, similar). Conserved sequence motifs are enclosed in red rectangles. (*B*) ATPase activities of wild-type (WT) and variant UAP56 proteins. ATPase activities were not detectable without total yeast RNA (No RNA). With 0.4 μg/μl total yeast RNA, both the WT and a DEAD mutant exhibited concentration-dependent ATPase activities. UAP56ΔN (ΔN) and its DEAD mutant (ΔN-DEAD) have reduced activities. (*C*) RNA concentration dependence of UAP56 ATPase activities. A fixed protein concentration (0.3 μg/μl) was used in the assays.

C-terminal domains of UAP56ΔN are the corresponding domains of an archaeal DEAD-box protein mjDEAD (35), the yeast translation initiation protein eIF4A (19, 36, 37), and a prokaryotic nucleotide excision repair helicase UvrB (38–40).

Fig. 2. Overall structure. The structure of UAP56ΔN is shown as a ribbon model (*A*) and in a surface representation (*B*), viewed from a similar direction. (*A*) Cyan, N-terminal domain; brown, C-terminal domain; green, interdomain linker. Locations of conserved helicase sequence motifs are labeled. (*B*) Blue, positively charged electrostatic potential; white, neutral electrostatic potential; red, negatively charged electrostatic potential. A large ATP-binding cleft is formed between the N- and C-terminal helicase domains.

The N-terminal domain of UAP56ΔN can be superimposed with the corresponding domain of mjDEAD (PDB ID code 1HV8; Z score = 30.9), eIF4A (PDB ID code 1QVA; Z score = 30.5), and UvrB (PDB ID code 1D2M; Z score = 12.8) with rms deviations of 1.4, 1.6, and 3.0 Å, respectively. The rms deviations between the C-terminal domain of UAP56ΔN and that of eIF4A (PDB ID code 1FUK; Z score = 21.2), mjDEAD (Z score = 18.8), and UvrB (Z score = 13.1) are 2.0, 2.1, and 2.7 Å, respectively.

The two domains are joined covalently by a nine-residue interdomain linker (Fig. 2*A*). Besides the linker, the two domains contact each other mainly by means of a portion of αK in the C-terminal domain and αA and αD in the N-terminal domain (Fig. 2). The "open" conformation of the two domains creates a large pocket between them. Amino acids belonging to the conserved ATPase/helicase motifs are located either inside or on the protein surface adjacent to the pocket (Fig. 2*A*). The location of these conserved residues suggests that ATP binds in the pocket between the two domains.
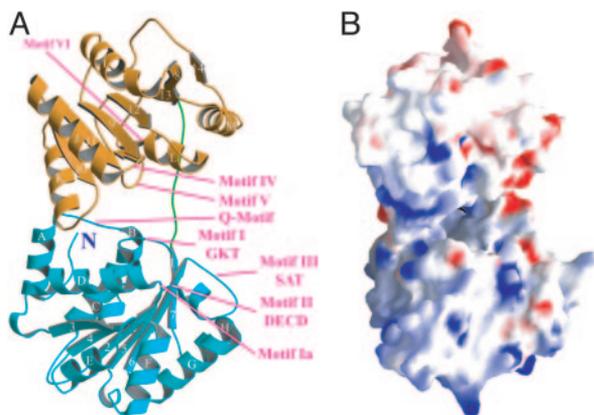
ATP-Binding Pocket and UAP56–ADP Interactions. To understand the exact mode of UAP56–ATP interaction and the mechanism of ATP hydrolysis, we attempted but failed to cocrystallize UAP56ΔN with ATP or its nonhydrolyzable analogs (adenosine 5′-[β,γ-imido]triphosphate or adenosine 5′-[γ-thio]triphosphate). Soaking the apo-UAP56ΔN crystals with ATP or its nonhydrolyzable analogs resulted in rapid dissolving of UAP56ΔN crystals. We cocrystallized UAP56ΔN with MgADP and solved a 2.7-Å structure by molecular replacement. A
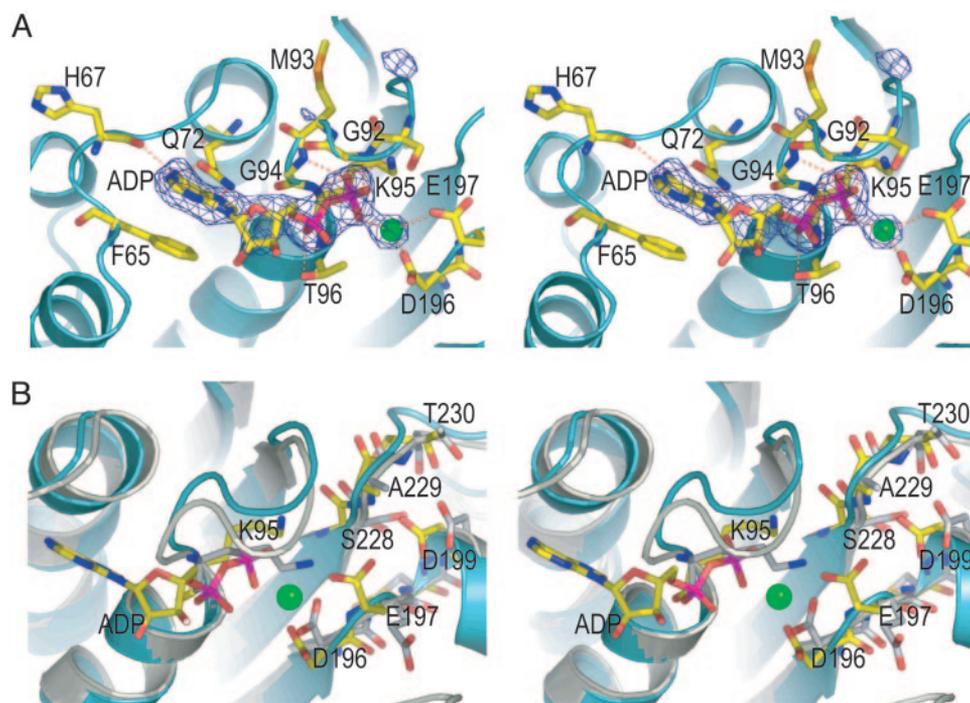
Fig. 3. MgADP binding site. (*A*) A stereoview of UAP56–MgADP interactions (protein is shown in a ribbon representation; ADP and residues contacting MgADP are shown in stick models; and magnesium ion is depicted as a green sphere). An omit difference (MgADP is omitted) electron density map is shown at 3.0 σ cutoff level. Red dashed lines denote hydrogen bonds or ionic bonds when the magnesium ion is involved. (*B*) Superposition of apo- and MgADP-bound structures. White ribbons represent the apostructure of UAP56; salmon-colored ribbons represent the MgADP-bound UAP56. ADP and selected residues are shown in a stick representation. Residues from the apostructure are colored white.

difference electron density map clearly shows the presence of an ADP molecule and a magnesium ion (Fig. 3*A*). The two helicase domains in the complex structure adopt an interdomain conformation very similar to that in the apostructure, although UAP56ΔN and the UAP56ΔN–MgADP complex were crystallized under different conditions and belong to different space-groups (P1 vs. P2₁). In the complex structure, residues 347–351, which are parts of motif V, and five residues at the C terminus (amino acids 424–428) are disordered.

As expected, ADP is bound in the pocket between the N- and C-terminal domains. The adenine base is bound in a region formed by a loop encompassing residues 65–69 (Phe-65, Glu-66, His-67, Pro-68, and Ser-69) and residues Phe-47, Gln-72, and Ala-97. In particular, the N6 atom of the adenine ring forms a hydrogen bond with the carbonyl group of His-67, and the N6 and N7 atoms of the base also interact with the Oε1 and Nε2 atoms of Gln-72 (from the Q-motif), respectively (Fig. 3*A*). Additionally, the adenine ring stacks with the phenyl ring of Phe-65. The side chains of Phe-65 and Gln-72 appear to be the major determinants of ADP binding, because other residues, such as Phe-47 and Ala-97, contact the base through van der Waals interactions.

The ribose moiety is bound in a manner with the 2′ and 3′ hydroxyl groups pointing away from the N-terminal domain. The pyrophosphate moiety makes extensive interactions with residues in Motif I (Fig. 3*A*). The α phosphate of ADP directly interacts with the conserved Thr-96 by means of a hydrogen bond (Fig. 3*A*). The oxygen atoms of the β phosphate make three hydrogen bonds with the backbone amide groups of Gly-92, Gly-94, and Lys-95, the latter two residues being part of the invariant GKT triplet of motif I. The O2 atom of the β phosphate makes an additional hydrogen bond to the Nζ atom of Lys-95. The O3 atom interacts with a magnesium ion, which is coordinated by Glu-197 (2.7 Å apart). Although the magnesium ion is

further away from Asp-196 (3.5 Å apart), the presence of this negatively charged residue contributes to the binding of the magnesium ion.

The binding of MgADP induces structural rearrangement near the pyrophosphate moiety and the magnesium binding site. Upon ADP binding, the motif I loop (amino acids 90–94) undergoes significant conformational changes, moving in a direction away from the center of the N-terminal domain (Fig. 3*B*). This conformational change is necessary to accommodate the binding of ADP, because the pyrophosphate moiety is inserted between the motif I loop and the body of the N-terminal domain. The side chains of several invariant residues surrounding ADP, notably Lys-95 and Glu-197, undergo large conformational changes to accommodate the binding of MgADP (Fig. 3*B*). In the absence of MgADP, Lys-95 contacts Asp-196 by means of ionic/hydrogen bond interactions. The binding of MgADP displaces the above interactions and gives rise to the Mg²⁺-Asp-196 and β-phosphate-Lys-95 bondings shown in Fig. 3. Glu-197 points away from the magnesium binding site when MgADP is not present (Fig. 3*B*).

A DECD to DEAD Mutant Destabilizes Motif III. UAP56 contains a characteristic DECD box, distinct from a large number of DEAD-box RNA helicases. Glu-197 in the DECD motif binds the magnesium ion and can serve as the catalytic base in ATP hydrolysis (Fig. 3*A*). Cys-198 and Asp-199 stabilize the SAT loop in Motif III: Asp-199 makes several hydrogen bonds with the backbone amide groups of the SAT loop, as well as with the hydroxyl group of Ser-228 (Fig. 3*B*). Cys-198 is buried, and the sulfhydryl group makes a hydrogen bond with the carbonyl of Phe-147. It was reported that changing the DECD motif of Sub2 to a more common DEAD motif renders the yeast inviable (8). Our *in vitro* assay revealed that the C198A mutation of UAP56 only has a mild effect on the RNA-stimulated ATPase
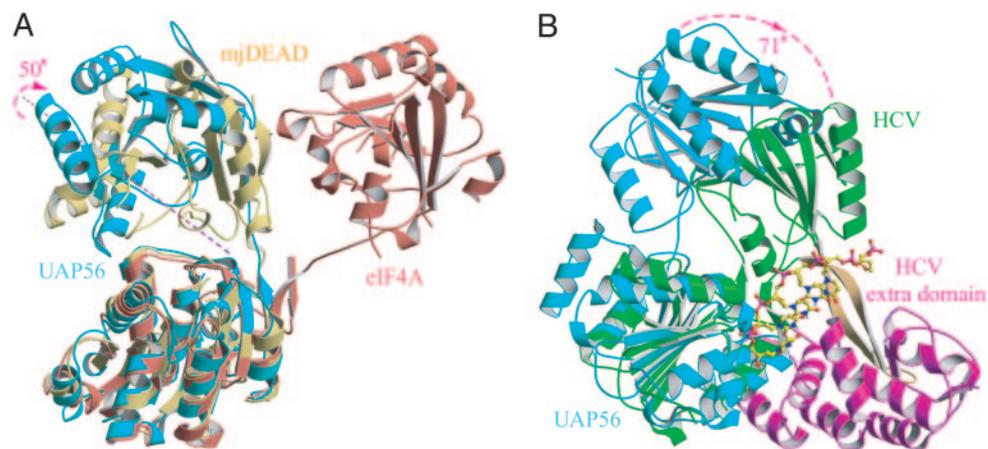
Fig. 4. Domain organizations. (*A*) The UAP56 structure (cyan) is superimposed with the structures of yeast eIF4A (red; PDB ID code 1FUU) and mjDEAD (yellow; PDB ID code 1HV8). The N-terminal domains were aligned. The dashed magenta line indicates the axis around which a rotation of ≈50° will superimpose the C-terminal domain of UAP56 with that of mjDEAD. (*B*) Superposition of UAP56 with hepatitis C virus helicase (PDB ID code 1A1V) (helicase domains shown in green; extra domain shown in magenta and beige), viewed from a perpendicular direction (from left to right) with respect to that in *A*. The magenta dashed arc indicates a rotation of ≈71° between the two C-terminal domains. The single-stranded DNA oligo bound to the viral helicase is shown in a stick model.

activity (Fig. 1 *B* and *C*). It is possible that the C198A mutation induces structural changes distant from the ATPase catalytic site, thus interfering with other functions of UAP56. To learn the potential structural changes, we crystallized the DEAD variant of UAP56ΔN under the same condition in which the apoform of UAP56ΔN crystals was obtained. The mutant and the wild-type structures were virtually identical, except that the helix (αH) immediately following the SAT motif (motif III) has poor electron density (data not shown), suggesting that the helix is destabilized in the mutant structure. Motif III differs both in length and amino acid composition among various helicases, and a consensus function of this motif is not clear. In PcrA (41), this motif has been implicated in interdomain interactions. In hepatitis C virus NS3 (19–21, 42), UvrB (38–40), and SecA (43), motif III is involved in interacting with the protein-specific extra domains that form part of the ligand-binding site. Because αH is located on the surface and exposes a large accessible area (see Fig. 5*A*), it is likely that it is involved in interacting with RNA or other protein factors.

Structural Comparison. The overall fold of UAP56 domains is very similar to that of all other SF2 helicases, including the recently determined separate domain structures of UAP56 (44). The N- and C-terminal domain structures in this study (the apoform) superimpose with the corresponding domains in ref. 44 at rms deviations of 1.14 and 2.1 Å, respectively. The main differences in the N-terminal domain are as follows: (*i*) the citrate-bound motif-I loop in ref. 44 resembles the ADP-bound conformation in our structure; and (*ii*) the helix αH (motif III) is shifted by ≈2.1–4.2 Å between the structures of apo-UAP56ΔN and that of ref. 44. The structural flexibility of αH is consistent with our observation that this helix is partially destabilized in the C198A mutant. Our structure, which contains both domains, does not support the proposal that the N-terminal domain of UAP56 forms a dimer (44). The two C-terminal domains superimpose with an rms deviation of 2.1 Å. Substantial conformational differences occur in two segments spatially close to each other, i.e., residues 346–355 (motif V) and 378–383 (motif VI). This region is stabilized by crystal packing in the C-terminal domain structure in ref. 44, and this region is flexible in our structure with the two domains together. As described earlier, the motif V loop is disordered in the structure of the UAP56ΔN–MgADP complex. All together, the structural differences suggest that motifs III, V, and VI are likely to be dynamic in solution.

The most striking difference among the SF2 helicases is the spatial organization of the two helicase domains. Among the SF2 structures, only two are well documented DExD/H box RNA-dependent ATPases/helicases. They are the hepatitis C virus NS3 RNA helicase domain (20, 21) and yeast eIF4A (19). UAP56, eIF4A, NS3, and mjDEAD all have distinct interdomain conformation (Fig. 4): the two eIF4A domains are not in contact (Fig. 4*A*), and the two domains in NS3 and UvrB are closer to each other (a "closed" conformation). When the two N-terminal domains are superimposed, the C-terminal domain of NS3 is rotated ≈71° with respect to the C-terminal domain of UAP56 (Fig. 4*B*), whereas the corresponding rotation involving UvrB is ≈79° (data not shown). The two mjDEAD domains are in an intermediate conformation, with an ≈50° rotation between the
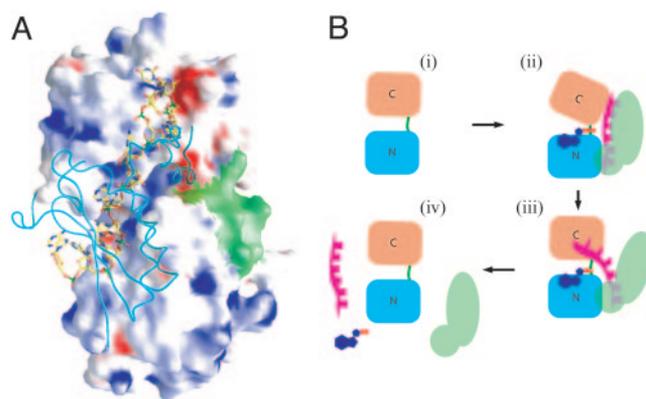

Fig. 5. A model of UAP56 mechanisms. (*A*) The branch point-containing RNA in the SF1–RNA complex is modeled into the putative RNA-binding channel of UAP56. The main chain of SF1 is shown as a cyan worm model, and the RNA is shown in a stick model. The exposed surface area of motif III, which is partially destabilized in the DEAD mutant, is colored green. (*B*) A "spring-loaded" unwindase/ATPase model of UAP56. Steps *i* and *iv* represent initial (apo) and product-bound (MgADP) conformation. In the presence of ATP, interactions with an RNA-bound U2AF65 or SF1 are likely to cause a reposition of the two helicase domains resembling the "productive" ATPase conformation (*ii*) as in UvrB. The RNA (shown in magenta) is sandwiched between the two proteins. Because both RNA and the RNA-binding protein (shown in green) contact both helicase domains, the large domain movement accompanying ATP hydrolysis rips attached RNA from the RNA-binding protein (*iii*).

two C-terminal domains when UAP56ΔN and mjDEAD are superimposed through the N-terminal domains (Fig. 4A). The rotation is around an axis approximately passing through the Cα atoms of Thr-344 and Gln-337 of UAP56 (Fig. 4A). The open conformation of UAP56 is likely to represent the apoconformation and a product-bound conformation, whereas a closed conformation similar to that of UvrB resembles that of a productive ATPase, because the two domains need to be in contact during ATP hydrolysis.

Structural Implications. UAP56, eIF4A and mjDEAD have distinct domain organizations, and they all comprise only the two essential helicase domains. Other SF2 family members, such as UvrB, NS3, PcrA, Rep, and SecA, contain extra nonconserved domains. The two helicase domains in the proteins with the extra domains have more stable and similar spatial organizations. These proteins with extra domains also have more robust and processive ATPase/helicase activities in the absence of other protein factors. In all cocrystal structures of SF2 helicases solved to date, namely NS3 (21), PcrA (41), and Rep (45), single-stranded DNA is bound in a cleft between the helicase domains and the extra nonconserved domains. In addition, the equivalent site has been proposed to be a protein-binding site in SecA (43). Thus, it appears that a conserved ligand-binding site in the SF2 family of helicases is located at the interface between the helicase domains and separate, protein-specific domains. UAP56, eIF4A, and mjDEAD lack extra domains, so it is likely that the equivalent roles of the extra domains are provided by their respective interacting proteins. The relatively low ATPase activities of UAP56 may be due to the absence of suitable substrates, possibly including both RNA and protein factors; two candidate proteins are U2AF65 and SF1.

We explored this hypothesis further by molecular modeling. Because no structure of an RNA-bound U2AF65 has been determined, we examined the possibility with SF1 by using the NMR structure of the SF1 RNA-binding domain complexed with a branch point-containing RNA (46). Although SF1 is not known to stably interact with UAP56, they are in close proximity because both proteins can interact simultaneously with U2AF65 (8, 46). As shown in Fig. 5A, the RNA in the SF1 complex fits well in the putative RNA-binding cleft of a modeled UAP56 structure, with the two domains in a conformation similar to that of UvrB and NS3 (Fig. 4B). The exposed RNA backbone faces UAP56, similar to the situation in the NS3–ssDNA complex (21). In this model, only one loop in SF1 (amino acids 188–192) clashes with UAP56. However, this loop is one of the most flexible regions in the NMR structure (46), suggesting that it is capable of undergoing large conformational changes (the averaged NMR structure is used for modeling here), thus the clashes may be avoided when the loop takes an alternate conformation. It is interesting to note that the SF1 domain also contacts both domains in UAP56, and the interaction area in UAP56 is near the exposed surface of αH (Fig. 5A), which is destabilized in the C198A mutant structure. Upon ATP hydrolysis, the two domains of UAP56 return to the product-bound conformation as observed in the crystal structures. Because the RNA contacts both domains in UAP56, the large domain movement accompanying ATP hydrolysis is likely to rip the RNA apart from SF1 by stretching and/or constraining the RNA structure. The release of separated RNA and protein from UAP56 and the exchange of ATP with the reaction product ADP should prepare UAP56 for the next round of reaction. A schematic drawing of the proposed model is shown in Fig. 5B.

1. Reed, R. (2003) *Curr. Opin. Cell Biol.* **15,** 326–331.
2. Maquat, L. E. (2004) *Nat. Rev. Mol. Cell Biol.* **5,** 89–99.
3. Staley, J. P. & Guthrie, C. (1998) *Cell* **92,** 315–326.
4. Silverman, E., Edwalds-Gilbert, G. & Lin, R. J. (2003) *Gene* **312,** 1–16.
5. Fleckner, J., Zhang, M., Valcarcel, J. & Green, M. R. (1997) *Genes Dev.* **11,** 1864–1872.
6. Kistler, A. L. & Guthrie, C. (2001) *Genes Dev.* **15,** 42–49.
7. Libri, D., Graziani, N., Saguez, C. & Boulay, J. (2001) *Genes Dev.* **15,** 36–41.
8. Zhang, M. & Green, M. R. (2001) *Genes Dev.* **15,** 30–35.
9. Luo, M. L., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M. & Reed, R. (2001) *Nature* **413,** 644–647.
10. Jensen, T. H., Boulay, J., Rosbash, M. & Libri, D. (2001) *Curr. Biol.* **11,** 1711–1715.
11. Gatfield, D., Le Hir, H., Schmitt, C., Braun, I. C., Kocher, T., Wilm, M. & Izaurralde, E. (2001) *Curr. Biol.* **11,** 1716–1721.
12. Strasser, K. & Hurt, E. (2001) *Nature* **413,** 648–652.
13. Herold, A., Teixeira, L. & Izaurralde, E. (2003) *EMBO J.* **22,** 2472–2483.
14. Longman, D., Johnstone, I. L. & Caceres, J. F. (2003) *RNA* **9,** 881–891.
15. MacMorris, M., Brocker, C. & Blumenthal, T. (2003) *RNA* **9,** 847–857.
16. Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A. G., Aguilera, A., Struhl, K., Reed, R. & Hurt, E. (2002) *Nature* **417,** 304–308.
17. Will, C. L. & Luhrmann, R. (2001) *Science* **291,** 1916–1917.
18. Schwer, B. (2001) *Nat. Struct. Biol.* **8,** 113–116.
19. Caruthers, J. M., Johnson, E. R. & McKay, D. B. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 13080–13085.
20. Yao, N., Hesson, T., Cable, M., Hong, Z., Kwong, A. D., Le, H. V. & Weber, P. C. (1997) *Nat. Struct. Biol.* **4,** 463–467.
21. Kim, J. L., Morgenstern, K. A., Griffith, J. P., Dwyer, M. D., Thomson, J. A., Murcko, M. A., Lin, C. & Caron, P. R. (1998) *Structure (London)* **6,** 89–100.
22. Tanner, N. K., Cordin, O., Banroques, J., Doere, M. & Linder, P. (2003) *Mol. Cell* **11,** 127–138.
23. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276,** 307–326.
24. Navaza, J. (2001) *Acta Crystallogr. D* **57,** 1367–1372.
25. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., et al. (1998) *Acta Crystallogr. D* **54,** 905–921.
26. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991) *Acta Crystallogr. A.* **47,** 110–119.
27. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950.
28. Merritt, E. A. & Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524.
29. Nicholls, A., Sharp, K. A. & Honig, B. (1991) *Proteins* **11,** 281–296.
30. DeLano, W. L. (2002) *The PYMOL Molecular Graphics System* (DeLano Scientific, San Carlos, CA).
31. Gorbalenya, A. E. & Koonin, E. V. (1993) *Curr. Opin. Struct. Biol.* **3,** 419–429.
32. Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291.
33. Story, R. M. & Steitz, T. A. (1992) *Nature* **355,** 374–376.
34. Caruthers, J. M. & McKay, D. B. (2002) *Curr. Opin. Struct. Biol.* **12,** 123–133.
35. Story, R. M., Li, H. & Abelson, J. N. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 1465–1470.
36. Johnson, E. R. & McKay, D. B. (1999) *RNA* **5,** 1526–1534.
37. Benz, J., Trachsel, H. & Baumann, U. (1999) *Structure* **7,** 671–679.
38. Machius, M., Henry, L., Palnitkar, M. & Deisenhofer, J. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 11717–11722.
39. Theis, K., Chen, P. J., Skorvaga, M., Van Houten, B. & Kisker, C. (1999) *EMBO J.* **18,** 6899–6907.
40. Nakagawa, N., Sugahara, M., Masui, R., Kato, R., Fukuyama, K. & Kuramitsu, S. (1999) *J. Biochem. (Tokyo)* **126,** 986–990.
41. Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S. & Wigley, D. B. (1999) *Cell* **97,** 75–84.
42. Cho, H. S., Ha, N. C., Kang, L. W., Chung, K. M., Back, S. H., Jang, S. K. & Oh, B. H. (1998) *J. Biol. Chem.* **273,** 15045–15052.
43. Hunt, J. F., Weinkauf, S., Henry, L., Fak, J. J., McNicholas, P., Oliver, D. B. & Deisenhofer, J. (2002) *Science* **297,** 2018–2026.
44. Zhao, R., Shen, J., Green, M. R., MacMorris, M. & Blumenthal, T. (2004) *Structure (London)* **12,** 1373–1381.
45. Korolev, S., Hsieh, J., Gauss, G. H., Lohman, T. M. & Waksman, G. (1997) *Cell* **90,** 635–647.
46. Liu, Z., Luyten, I., Bottomley, M. J., Messias, A. C., Hounginou-Molango, S., Sprangers, R., Zanier, K., Kramer, A. & Sattler, M. (2001) *Science* **294,** 1098–1102.

BIOCHEMISTRY