Crystal structure of an archaeal Ski2p-like protein from *Pyrococcus horikoshii* OT3

XIAODONG ZHANG,[1] TAKASHI NAKASHIMA,[1,2] YOSHIMITSU KAKUTA,[1,2] MIN YAO,[3] ISAO TANAKA,[3] AND MAKOTO KIMURA[1,2]

[1]Laboratory of Structural Biology, Graduate School of Systems Life Sciences, Kyushu University, Fukuoka 812-8581, Japan
[2]Laboratory of Biochemistry, Department of Bioscience and Biotechnology, Graduate School, Faculty of Agriculture, Kyushu University, Fukuoka 812-8581, Japan
[3]Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan

(RECEIVED July 5, 2007; FINAL REVISION September 24, 2007; ACCEPTED September 27, 2007)

Abstract

The Ski complex composed of Ski2p, Ski3p, and Ski8p plays an essential role in the 3′ to 5′ cytoplasmic mRNA degradation pathway in yeast. Ski2p is a putative RNA helicase, belonging in the DExD/H-box protein families and conserved in eukarya as well as in archaea. The gene product (Ph1280p) from the hyperthermophilic archaeon *Pyrococcus horikoshii* OT3 shows sequence homology with Ski2p, sharing 22.6% identical amino acids with a central region of Ski2p. In order to gain structural information about the Ski2p-like RNA helicase, we overproduced Ph1280p in *Escherichia coli* cells, and purified it to apparent homogeneity. Ph1280p exhibits DNA/RNA-dependent ATPase activity with an optimal temperature at ~90°C. The crystal structure of Ph1280p has been solved at a resolution of 3.5 Å using single-wavelength anomalous dispersion (SAD) and selenomethionyl (Se-Met)-substituted protein. Ph1280p comprises four subdomains; the two N-terminal subdomains (N1 and N2) fold into an RecA-like architecture with the conserved helicase motifs, while the two C-terminal subdomains (C1 and C2) fold into α-helical structures containing a winged helix (WH)-fold and helix-hairpin-helix (HhH)-fold, respectively. Although the structure of each of the Ph1280p subdomains can be individually superimposed on the corresponding domains in other helicases, such as the *Escherichia coli* DNA helicase RecQ, the relative orientation of the helicase and C-terminal subdomains in Ph1280p is significantly different from that of other helicases. This structural feature is implicated in substrate specificity for the Ski2-like helicase and would play a critical role in the 3′ to 5′ cytoplasmic mRNA degradation in the Ski complex.

Keywords: archaea; DExD/H-box protein; HhH-fold; mRNA degradation; *P. horikoshii*; RNA helicase; Ski2p; Ski complex; WH-fold; X-ray crystallography

RNA helicases are found in almost all organisms and have important roles in virtually all aspects of RNA metabolism, including transcription, pre-mRNA splicing, ribosome biogenesis, nucleocytoplasmic transport, translation, RNA decay, and organellar gene expression (for reviews, see de la Cruz et al. 1999; Rocak and Linder 2004). Elucidation of structure–function relationships of RNA helicases are, therefore, essential for a full understanding of the molecular mechanisms of these cellular processes involving RNA metabolism. RNA helicases belong to helicase superfamily II (Gorbalenya and Koonin 1993) and are further grouped into three families (DEAD-box, DEAH-box, and DEVH-box or Ski2p) on the basis of particular consensus sequences in the conserved motifs (de la Cruz et al. 1999); these RNA helicase families are often referred to as

Reprint requests to: Makoto Kimura, Laboratory of Biochemistry, Department of Bioscience and Biotechnology, Faculty of Agriculture, Graduate School, Kyushu University, Hakozaki 6-10-1, Higashi-ku, Fukuoka 812-8581, Japan; e-mail: mkimura@agr.kyushu-u.ac.jp; fax: 81-92-642-2853.

Article published online ahead of print. Article and publication date are at http://www.proteinscience.org/cgi/doi/10.1110/ps.073107008.

DExD/H-box proteins. Since the DEAD-box proteins are by far the largest family of RNA helicases; extensive genetic and biochemical studies on them have been carried out (de la Cruz et al. 1999; Rocak and Linder 2004). The family members contain eight to nine conserved sequence motifs including the eponymous DEAD motif and the newly identified Q-motif (Tanner et al. 2003; Rocak and Linder 2004; Cordin et al. 2006). These motifs are believed to be involved in binding an ATP, unwinding RNA duplexes, and/or disrupting protein–RNA or protein–protein interactions (Jankowsky et al. 2001; Rocak and Linder 2004; Cordin et al. 2006). Furthermore, structural data for several DEAD-box proteins have become available in recent years, providing a much better insight into the functions of these conserved motifs (Caruthers et al. 2000; Story et al. 2001; Carmel and Mathews 2004; Cheng et al. 2005). In general, the DEAD-box RNA helicases are composed of the helicase core domain containing roughly 400 amino acids and variable lengths of the N- or C-terminal extension(s). The helicase core domain is a dumbbell structure consisting of two compact domains connected by an extended linker; the two domains form the ATP-binding cleft and all motifs occur at the surface of the two domains (Caruthers et al. 2000; Story et al. 2001; Carmel and Mathews 2004; Cheng et al. 2005). Since the structural features of the helicase core domain are highly conserved in all DEAD-box proteins, substrate specificity might reside in the domains that flank the helicase core domain or its interactions with other factors.

Ski2p is a 146-kDa putative RNA helicase with the DEVH-box sequence (Widner and Wickner 1993) and involved in 3′ to 5′ mRNA degradation in yeast (Anderson and Parker 1998). This pathway requires the 3′-exonuclease activity of the exosome, a multimeric assembly of 3′ to 5′ exonucleases, as well as Ski2p in a complex with Ski3p and Ski8p (Mitchell et al. 1997; Parker and Song 2004). This pathway also requires a GTPase, Ski7p, to bridge interaction between the exosome and Ski complex (Benard et al. 1999; van Hoof et al. 2000; Araki et al. 2001). It has been thus suggested that Ski7p recruits both the Ski complex and exosome to the 3′ end of mRNA (Araki et al. 2001). In addition to their role in general RNA decay, the exosome and Ski complex function in surveillance pathways by which aberrant mRNAs are detected and degraded. These include the nonsense-mediated mRNA decay pathway (NMD) (Maquat 2004), a pathway that eliminates mRNAs harboring premature translation termination codons (PTCs), and the nonstop decay (NSD) pathway that degrades mRNAs lacking a stop codon (van Hoof et al. 2002). Furthermore, recent studies showed that the exosome and Ski complex are also involved in the rapid degradation of 5′ fragments generated by no-go decay (Doma and Parker 2006) as well as by RISC cleavage (Orban and Izaurralde 2005).

Ski2p, Ski3p, and Ski8p form a stable heterotrimeric complex with a 1:1:1 stoichiometry (Brown et al. 2000). Ski3p is a tetratricopeptide repeat protein (Rhee et al. 1989), while Ski8p is a WD repeat-containing protein (Matsumoto et al. 1993). Ski2p did not associate with Ski3p in the absence of Ski8p, nor did Ski2p associate with Ski8p in the absence of Ski3p (Brown et al. 2000). Thus, Ski3p serves as a scaffold protein with its C terminus interacting with Ski8p and the sub-C terminus interacting with Ski2p (Wang et al. 2005). Although molecular genetic and biochemical studies on the Ski complex have elucidated the functional roles of each of its proteins, the structural basis for the specific function of Ski2p in the complex remains elusive as the three-dimensional structure of Ski2p is unknown.

It has been generally accepted that various components in archaea share many sequence and functional features with eukaryotic counterparts (Kyrpides and Woese 1998a,b). In addition, proteins from thermophilic organisms are more robust and crystallize better than those of mesophilic organisms. It was reported that the gene PH1280 of *Pyrococcus horikoshii* OT3 (Kawarabayasi et al. 2001), the translational product Ph1280p of which shows sequence similarity to Ski2p, though no homologs for Ski3p or Ski8p have heretofore been identified (Koonin et al. 2001). To gain insight into the structure–function relationships of Ski2p and also to provide structural information about the Ski2p protein family, we overproduced and crystallized full-length Ph1280p under appropriate conditions. In this paper, we described the crystal structure of Ph1280p determined at a resolution of 3.5 Å. Structural comparison revealed a unique arrangement of domains, a feature implicated in the specific function of the Ski2p-like RNA helicase in 3′ to 5′ mRNA degradation.

Results and Discussion

Sequence comparison

During the course of our structural genomic study of the hyperthermophilic archaeon *P. horikoshii* OT3, we found the gene PH1280, whose translational product Ph1280p shows sequence similarity to that of Ski2p from *Saccharomyces cerevisiae*. Ph1280p is composed of 715 amino acids with a calculated isoelectric point of 6.4. Figure 1A shows an amino acid sequence comparison of Ph1280p with that of Ski2p, aligned for maximum homology by the program CLUSTAL W (Thompson et al. 1994). Although lengths of the two proteins differ considerably, Ph1280p can be aligned with a central region of Ski2p by making a few long deletions in Ph1280p (Fig. 1B). This comparison shows that there are identical residues at 153 positions in the two sequences, that is, at 21.4% of the residues compared. The Ph1280p and Ski2p sequences have long runs of highly conserved sequences, namely Ser34–Ile60, Tyr73–Glu87, Thr121–Ile163, Leu175–Trp188, Val242–Lys259,

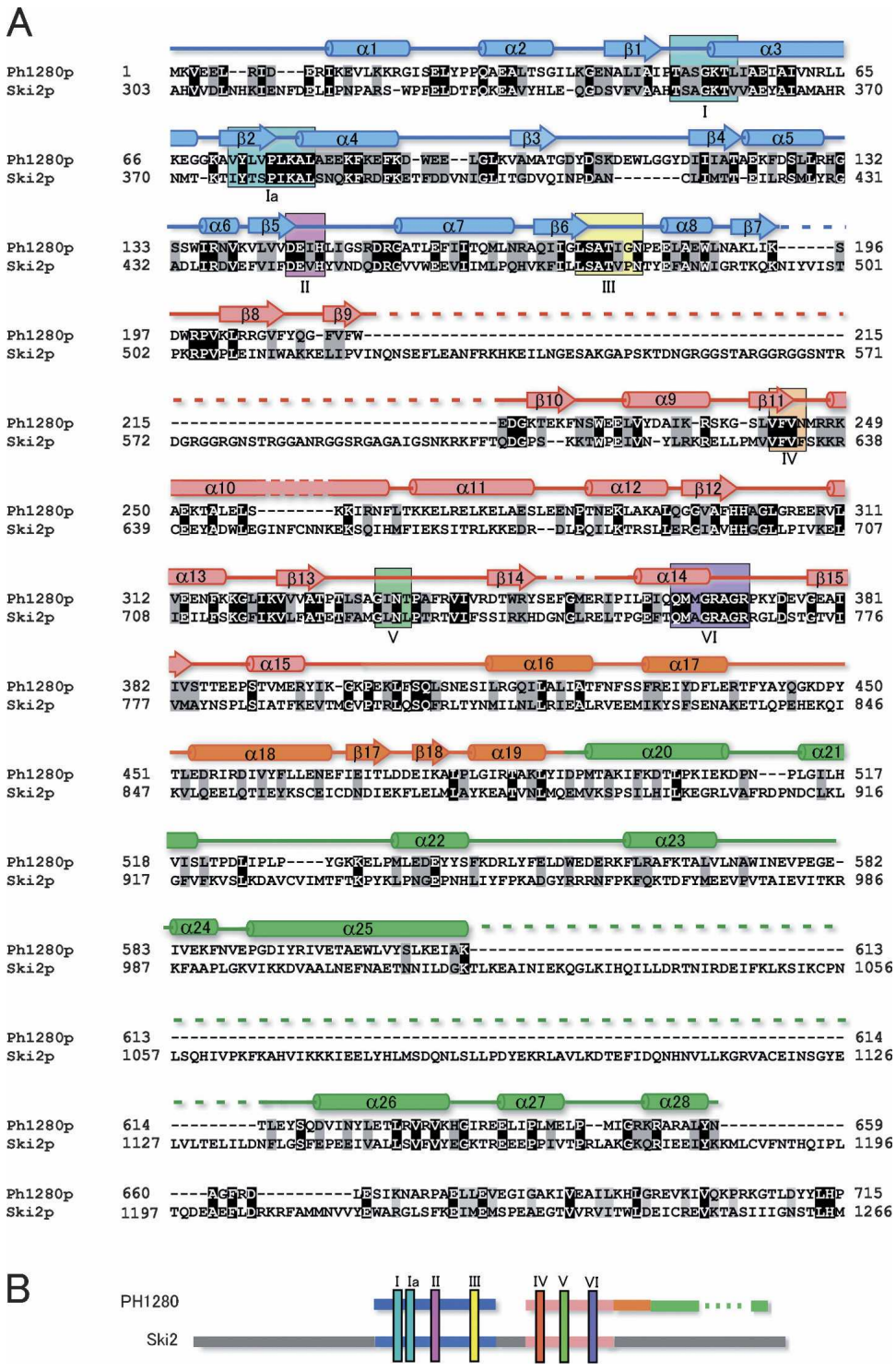

Figure 1. A sequence alignment of Ph1280p in *P. horikoshii* and Ski2p in yeast. (*A*) The two amino acid sequences are aligned by the program CLUSTAL W (Thompson et al. 1994). The Ph1280p sequence can be aligned with the central region of Ski2p, having two long insertions. Secondary structural elements as determined by the present study are also given. Ph1280p comprises four subdomains; the two N-terminal subdomains (N1 and N2) form a helicase core domain, while the C-terminal region consists of two subdomains (C1 and C2). Hence, the secondary structure in N1, N2, C1, and C2 is shown in blue, pink, orange, and green, respectively. The residues highlighted in black represent complete conservation and those in gray conservative mutation. Conserved sequence motifs (I–VI) found in the DExD/H-box family proteins are marked in boxes. (*B*) A schematic diagram of the sequence comparison of Ph1280p with Ski2p is shown. The color code for the domains is as in *A*.

Glu313–Val341, and Glu361–Ile381 in Ph1280p, all of which fall within the N-terminal two-thirds of the molecule. It is known that DExD/H-box families contain at least seven conserved sequence motifs including motif II (DExD/H). Ph1280p and Ski2p share all seven conserved sequence motifs common to the DExD/H-box proteins (de la Cruz et al. 1999) within the N-terminal two-thirds of the molecule, as shown in Figure 1. It is hence likely that the helicase core domains of Ph1280p and Ski2p fold into a similar three-dimensional structure. On the other hand, the C-terminal 300 residues are less conserved in the Ph1280p and Ski2p proteins (only 13% identity), having a long deletion in Ph1280p. Presumably, the C-terminal extra region might have an additional function in Ski2p. Nevertheless, since the two sequences show the best homology among the DExD/H-box proteins, it could be expected that the structural analysis of Ph1280p would aid in understanding the molecular mechanism of Ski2p-like helicase proteins.

Preparation and characterization of Ph1280p

For crystallographic analysis, we sought to overproduce the full-length Ph1280p (residues 1–715), the N-terminal helicase core domain (N domain: residues 1–410), and the C-terminal domain (C domain: residues 370–715) in *Escherichia coli*. The individual gene fragments were amplified from genomic DNA from *P. horikoshii* OT3 and the amplified DNA fragments were placed under the control of the phage promoter on the expression plasmid pET-22b. Expression in *E. coli* BL21 (DE3) Codon Plus RIL strain and preparation of the lysate were done as described in Materials and Methods. Since the gene products were expected to carry a His-tag sequence attached at the C terminus, the *E. coli* lysate was loaded on a His-bind resin column. The adsorbed proteins were eluted, and then further purified by gel filtration on a Sephacryl S-200 column.

DExD/H-box proteins are known to have DNA- and/or RNA-dependent ATPase/helicase activity. To evaluate the structural integrity of the resulting proteins and also to confirm a domain arrangement of Ph1280p, we tested the ATPase activity of the resulting proteins by measuring the free phosphate released with a colorimetric assay based on malachite green, as described in Materials and Methods. As shown in Figure 2A,B, Ph1280p is an active DNA- and RNA-dependent ATPase in vitro. In the absence of DNA or RNA, Ph1280p had negligible ATPase activity. The ATPase activity was stimulated greatly with the addition of herring sperm DNA or total yeast RNA, with maximum activity at ~90°C. The result showed that the purified Ph1280p folds into an active form with the DNA/RNA-dependent ATPase activity.

The DExD/H motif is known to be responsible for the ATPase activity. To further verify the DNA/RNA-dependent ATPase activity of Ph1280p, two mutants, D145N and D145A, in which Asp145 at DEIH sequence was replaced with Asn and Ala, respectively, were prepared, and the resulting mutants were again characterized with respect to the DNA/RNA-dependent ATPase activity. The result indicated that both mutations significantly reduced the ATPase activity (Fig. 2C), demonstrating that the purified Ph1280p has the DNA/RNA-dependent ATPase activity and that DExD/H motif plays a crucial role in the ATPase activity of Ph1280p.

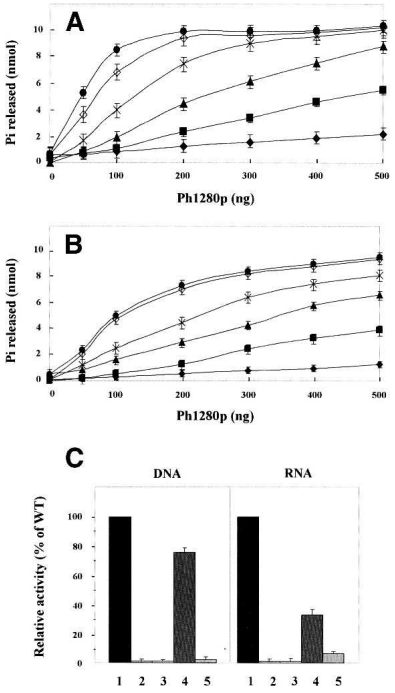

Figure 2. DNA- and RNA-dependent ATPase activity. (*A*) DNA-dependent ATPase activity of Ph1280p. (*B*) RNA-dependent ATPase activity of Ph1280p. ATPase activity was measured in the presence or absence of herring sperm DNA or total yeast RNA for 15 min at different temperatures (37°C, ◆; 50°C, ■; 60°C, ▲; 70°C, ×; 80°C, ◇; and 90°C, ●), as described in Materials and Methods. (*C*) DNA- and RNA-dependent ATPase activities of the mutants D145A (lane *2*) and D145N (lane *3*), the helicase core domain (residues 1–410) (lane *4*), and the C domain (residues 370–715) (lane *5*) were measured at 90°C, as described under Materials and Methods, and the activities of the domains were expressed relative to that of Ph1280p as 100% (lane *1*). The experiments were done at least in triplicate and bars indicate SD.

Next, the DNA/RNA-dependent ATPase activities of the helicase core domain (N domain: residues 1–410) and C domain (residues 370–715) were examined at 90°C, and then their ATPase activities were compared with those of full-length Ph1280p (Fig. 2C). As we expected, the helicase core domain retained, albeit to a lesser extent, both DNA- and RNA-dependent ATPase activities, while the C domain exhibited no activity. Interestingly, it was found that deletion of the C domain more strongly affected the RNA-dependent ATPase activity than DNA-dependent ATPase

activity, as shown in Figure 2C. This finding suggests the C domain in Ph1280p to be favorably involved in interaction with RNA.

To address this assumption, we measured RNA binding potency of the C domain (residues 370–715) by gel shift assay, as described in Materials and Methods. The gel shift assay revealed that the C domain formed a complex with single-stranded RNA (ssRNA) (Fig. 3A), whereas the incubation with double-stranded RNA (dsRNA) showed a little band shift (Fig. 3B). The result suggests that the C domain may play an important role in the RNA-dependent ATPase activity of Ph1280p, preferentially interacting with ssRNA.

Crystallization and structural determination

Crystallization was carried out at 20°C by hanging-drop vapor diffusion. The crystals of full-length Ph1280p and the C domain were grown under several conditions of Wizard I and II, but almost all crystals diffracted to resolutions of up to 8 Å. In contrast, the crystallization trails have yielded no crystals of the N domain. To obtain crystals suitable for X-ray analysis, the crystallization was optimized through sitting-drop vapor diffusion. The best crystal of the full-length Ph1280p was grown at 20°C from 0.1 M phosphate-citrate buffer (pH 4.2) containing 0.2 M Li_2SO_4 and 10% 2-propanol. This condition grew crystals in the space group $P4_322$ with unit cell constants $a = b = 133.74$ Å, $c = 137.64$ Å, and $\alpha = \beta = \gamma = 90°$ that diffracted to a resolution of up to 3.5 Å. The crystals contain a single molecule in the asymmetric unit, with a solvent content of 67.2% and a V_M value of 3.75 $Å^3Da^{-1}$, which is in the range of common protein crystals.

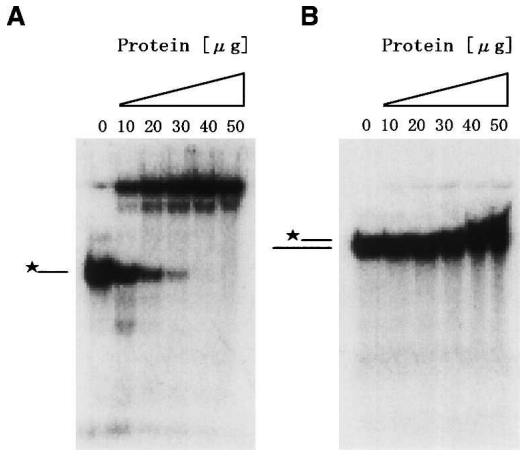

Figure 3. Gel shift assay of the C domain in Ph1280p. ssRNA (*A*) and dsRNA (*B*) used for the gel shift assay were prepared as described in Materials and Methods. ^{32}P-labeled ssRNA or dsRNA were incubated with the indicated concentrations of the C domain and separated on a 20% polyacrylamide gel that was run at 4°C. *–, free ssRNA; *=, free dsRNA.

The structure was solved by single-wavelength anomalous dispersion (SAD) phasing using the selenomethionine (Se-Met)-derivatized Ph1280p. The heavy atom sites were located by using the automated Patterson search routine in the program SHELX and were later found to correspond to eight of the 12 possible methionines in the molecule. Phase calculation for these sites yielded an interpretable electron density map, which was further improved with the programs SOLVE and RESOLVE. The path of the polypeptide backbone could be unambiguously traced as α-helices and β-strands in the initial electron density map calculations, but the identity of the amino acids was not revealed at this resolution. We could, however, build and refine an atomic model for the helicase core domain of Ph1280p with the help of a model generated from the known crystal structure of the *E. coli* DNA helicase RecQ (Bernstein et al. 2003): The helicase domains of Ph1280p and RecQ share 85 identical residues, that is, 21.9% identity. The homology model allowed an unambiguous fit of the amino acid residues in the helicase core domain (Fig. 4A). The α-carbon chain in the C-terminal domain could be primarily traced using a combination of RESOLVE and manual rebuilding, but a model of Ph1280p generated with the program Robetta (http://www.robetta.org) further allowed us to model for the C-terminal region (Fig. 4B). Amino acid residues at the C-terminal domain could be resolved up to position 659, but the identity of the remaining amino acids was not revealed at this resolution. The resulting model for Ph1280p further served to improve electron density maps. By iterating this process, a current model could be built and refined to an R_{factor} of 29.1% (R_{free} was 34.4%) at 3.5 Å resolution. The data collection and refinement statistics are summarized in Table 1.

Structure description

The current model includes 659 residues out of 715 residues in Ph1280p, lacking ~50 residues at the C terminus. The protein comprises four subdomains; the two N-terminal subdomains (N1 and N2) form a conserved RecA-like helicase core domain that contains seven conserved motifs involved in ATPase activity and DNA/RNA binding (Story and Steitz 1992), while the two C-terminal subdomains (C1 and C2 domains) predominantly fold into α-helical structures (Fig. 5A). A topological diagram showing the secondary structure of Ph1280p is shown in Figure 5B. The N1 subdomain (residues 1–195) is a typical RecA-like helicase domain consisting of a seven-stranded parallel β-sheet (β1–β7) surrounded on one side by four α-helices (α1–α4) and on the other side by four α-helices (α5–α8). The N2 subdomain (residues 196–395), like the N1 subdomain, folds into an eight-stranded parallel β-sheet (β8–β15) surrounded on one side by four α-helices (α9–α12) and

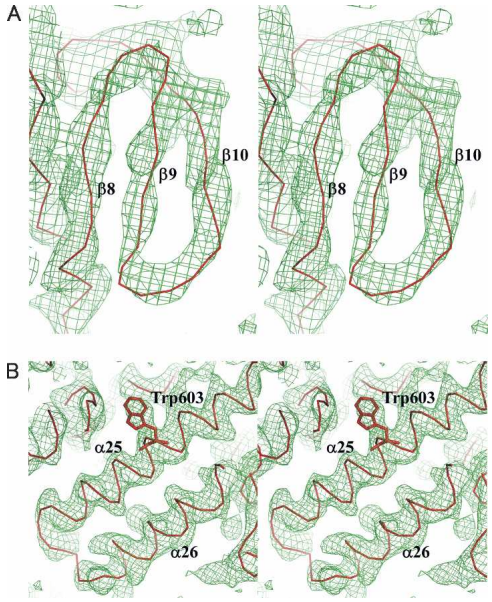

Figure 4. A representative electron density map. (*A*) 2Fo−Fc simulated annealing omit map for residues 210–217, at contour level 1.0 σ. The Cα trace from strand β8 to β10 is shown in red. (*B*) 2Fo−Fc simulated annealing omit map for residues 591–614, at contour level 1.0 σ. The Cα trace from strand α25 to α26 is shown in red. Trp603 is shown in stick representation.

on the other side by three α-helices (α13–α15). The N1 and N2 subdomains are connected by a short peptide linker (residues 196–201), which joins the last β-strand, β7, of the N1 subdomain and the first β-strand, β8, of the N2-subdomain. These two subdomains are arranged so as to create the putative ATP and RNA-binding sites observed in other DExD/H-box proteins.

The polypeptide chain enters the C1 subdomain (residues 414–491), being folded into a compact α/β structure containing a four-helix bundle (α16–α19) inserted by a short antiparallel β-sheet (β17 and β18). The first three α-helices are connected through a series of loops, forming a helix-turn-helix (HTH)-fold. The fourth α-helix (α19) in the C1 subdomain is followed by the first α-helix (α20) in the C2 subdomain (residues 492–659), where the peptide chain extensively folds into a helix-hairpin-helix (HhH)-fold, having nine α-helices (α20–α26); the structure is completed by two long α-helices (α25 and α26) followed by two short helices (α27 and α28).

Structural comparison

A structure homology search on the Dali server revealed a helicase core domain (N1 and N2) of Ph1280p to have similarity to the corresponding domains of the *E. coli* RecQ (Dali Z-score = 15.5) (Bernstein et al. 2003) and the *Methanococcus jannaschii* DEAD-box protein (Dali Z-score = 19.4) (Story et al. 2001). The helicase domain of Ph1280p could be superimposed on the corresponding domains of RecQ (PDB ID 1OYW) and the *M. jannaschii* protein (PDB ID 1HV8), giving root-mean-squared deviation (RMSD) values of 2.39 Å for 340 Cα atoms and 3.84 Å for 363 Cα atoms, respectively, as shown in Figure 6A. When individual domains were superimposed, the N1 subdomain and N2 subdomain could be superimposed on the corresponding domains of the *M. jannaschii* proteins, giving RMSD values of 2.05 Å for 209 Cα atoms and 2.41 Å for 154 Cα atoms, respectively, whereas the comparison with the domains of RecQ gave RMSD values of 2.49 Å for 209 Cα atoms and 1.89 Å for 132 Cα atoms, respectively. This finding indicates that, in structure and spatial arrangement, the helicase subdomains (N1 and N2) of Ph1280p are highly conserved with the corresponding domains in the DExD/H-box protein families. Since conserved helicase sequence motifs in Ph1280p line the cleft walls in an arrangement similar to that observed in other DExD/H-box protein structures, it is likely that the helicase core domain (residues 1–395) in Ph1280p plays an essential role in RNA, unwinding in a manner identical to that of other DExD/H-box helicases. It was reported that motifs I, II, and VI are involved in ATP binding and hydrolysis, and that aspartic acid in motif II forms interactions with β- and γ-phosphates through Mg^{2+} ions (Caruthers and McKay 2002). Indeed, the mutations of Asp145 abolished the DNA- and RNA-dependent ATPase activity of Ph1280p, as shown in Figure 2C. In this context, it is of interest that closer inspection revealed a possible electron density map

Table 1. *Data collection and refinement statistics*

Data set	Se-Met peak
Data collection	
Space group	$P4_322$
Unit cell dimensions	$a = b = 133.74$ Å, $c = 133.64$ Å, $\alpha = \beta = \gamma = 90°$
Resolution (Å)	50–3.51 (3.63–3.51)
Wavelength (Å)	0.97904
No. of observed reflections	186,427
No. of unique reflections	16,159
I/σ (I)	25.4
Multiplicity	11.5
Completeness (%)	99.9 (99.9)
R_{sym} (%)[a]	13.7 (47.4)
Refinement statistics	
R_{cryst}/R_{free} (%)[b]	29.1/34.4
Average B factor (Å^2)	76.5
RMSD for bond length (Å)	0.012
RMSD for bond angles (°)	1.91

Values in parentheses are for the highest resolution shell.
[a] $R_{sym} = \sum|I_j - \langle I\rangle| / \sum I_j$, where I_j is the intensity of the observation and $\langle I\rangle$ is the mean intensity of the reflection.
[b] $R = \sum||F_o| - |F_c||/\sum|F_o|$, where R_{cryst} is calculated using the 95% of reflections used in refinement and R_{free} is calculated using the 5% held aside.

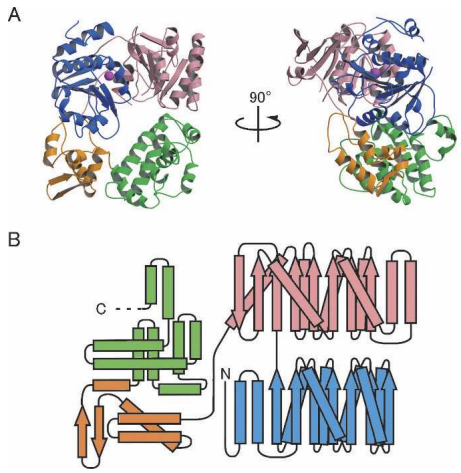

Figure 5. Overall structure of the archaeal Ski2p-like protein Ph1280p in *P. horikoshii*. (*A*) Orthogonal views of a ribbon diagram of the crystal structure of Ph1280p. The protein comprises four subdomains; the N-terminal subdomains N1 (blue) and N2 (pink) form the helicase core domain, whereas the C-terminal subdomains C1 (orange) and C2 (green) predominantly fold into the α-helical structure, having WH-fold and HhH-fold, respectively. A possible Mg^{2+} ion bound to the side chain of Asp145 in motif II is shown as a magenta sphere. (*B*) A topological diagram showing the secondary structure of Ph1280p. α-Helices and β-strands are indicated by rectangles and arrows, respectively. Color coding is the same as in Figure 1. All figures were prepared using Molscript (Kraulis 1991) and Raster3D (Merritt and Bacon 1997).

for a single Mg^{2+} ion bound to the side-chain carboxyl group of Asp145 in motif II (DEIH) in Ph1280p.

The Dali server search using the entire length of the C domain of Ph1280p failed to detect any structures similar to other DExD/H-box proteins. However, individual comparison of the C1 and C2 subdomains have revealed that the C1 subdomain shows a topology similar to a winged helix (WH)-fold occurring in a variety of DNA- and RNA-binding proteins, such as the second subdomain in the RecQ-Ct domain (Bernstein et al. 2003) as well as in the DNA-binding domain Zα in double-stranded RNA adenosine deaminase (ADAR1) (Schwartz et al. 1999), as shown in Figure 6B. In contrast, the C2 subdomain shows no similarity to any other known structures. The C1 subdomain of Ph1280p can be superimposed on the RecQ-Ct subdomain and DNA-binding domain Zα in ADAR1, giving 2.44 Å for 65 Cα atoms and 1.89 Å for 53 Cα atoms, respectively. The WH proteins constitute a subfamily within the large ensemble of HTH proteins (Gajiwala and Burley 2000). The WH-fold is a compact α/β structure consisting of two wings, three α-helices, and three β-strands, arranged in the order H1–S1–H2–H3–S2–W1–S3–W2 (Fig. 6B). The N-terminal half of the motif is largely helical, whereas the C-terminal half is composed of two of the three strands forming the twisted antiparallel β-sheet and the two large loops or wings, W1 and W2. The wing W1 connects strands S2 and S3, while the wing W2 extends from strand S3 to the C terminus (Fig. 6B). Although the C1 subdomain of Ph1280p lacks the first strand S1 and the second wing W2 folds into the α-helix (α19) instead of a loop structure in the canonical WH-fold, the arrangement of the secondary structure and relative orientation of the β-sheet (β17 and β18) and the third α-helix (the recognition helix) (α18) are similar to those of the canonical WH-fold (Fig. 6B). The present study revealed that the C domain (residues 370–715) is involved in the RNA-dependent ATPase activity (Fig. 2C)

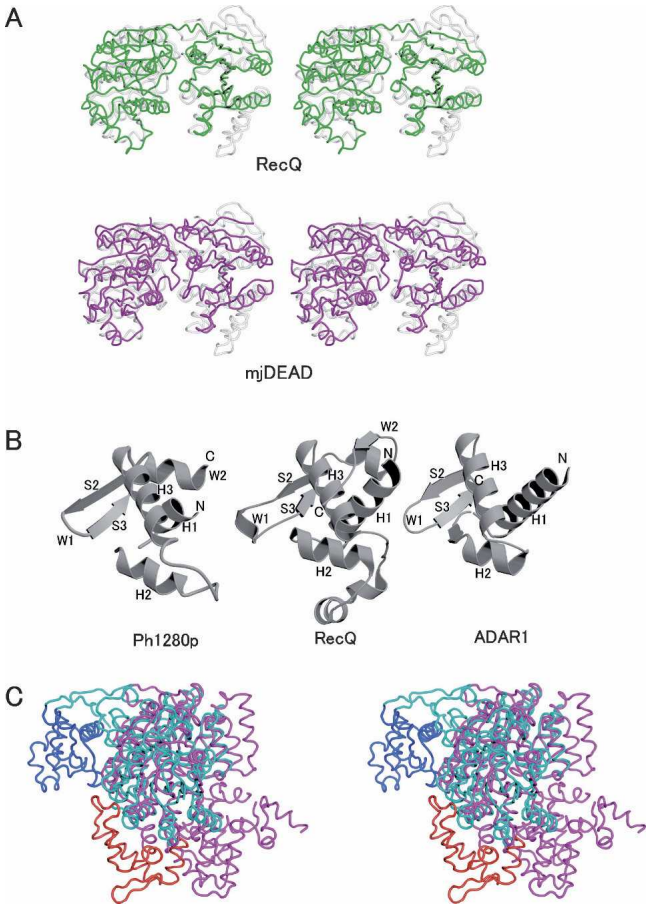

Figure 6. Structural comparison of the subdomains of Ph1280p with the corresponding domains of other DExD/H-box helicases. (*A*) The N-terminal helicase core domain of Ph1280p (gray) is superimposed on the corresponding domains in the *E. coli* DNA helicase RecQ (green) (PDB code 1OYW) and the *M. jannaschii* DEAD-box protein (magenta) (mjDEAD) (PDB code 1HV8). (*B*) The WH-folds of the C1 subdomain in Ph1280p, RecQ, and ADAR1 are shown in a similar orientation. The secondary structures are labeled on the basis of a canonical WH-fold, arranged in the order H1–S1–H2–H3–S2–W1–S3–W2 (Gajiwala and Burley 2000). (*C*) Superimposition of full-length Ph1280p (pink) on the *E. coli* RecQ structure (cyan). In this comparison, the C1 domain in Ph1280p and the corresponding domain in RecQ are shown in red and blue, respectively. The two structures were superimposed by fixing the structures of the helicase core domains. The structure of Ph1280p is in approximately the same orientation as the structure on the *right* in Figure 3A.

and preferentially binds to ssRNA (Fig. 3A). It is, therefore, likely that the C1 subdomain of Ph1280p may play an important role in RNA-dependent ATPase/helicase activity of Ph1280p, serving as a ssRNA recognition site.

Conclusion

The archaeal Ski2p-like protein Ph1280p comprises four subdomains; the two N-terminal subdomains (N1 and N2) form the helicase core domain, whereas the two C-terminal subdomains (C1 and C2) fold into WH-fold and HhH-fold, respectively. The individual structures of the helicase core domain and the C1 subdomain of Ph1280p are highly conserved with respect to the corresponding domains in the DExD/H-box proteins, such as the *E. coli* RecQ. The spatial arrangement of the domains of Ph1280p is, however, significantly distinct from that of RecQ (Fig. 6C). When the helicase core domain of Ph1280p is superimposed onto the equivalent domain of RecQ, the positioning of the C-terminal domain relative to the helicase core domain of Ph1280p is substantially different from that of RecQ (Fig. 6C); the wing W1 of the C1 subdomain in Ph1280p must rotate approximately $\alpha = 36°$, $\beta = 110°$, and $\gamma = -37°$ in eulerian angles to superimpose on the equivalent W1 of the RecQ-Ct subdomain in RecQ, and the two wings must be ~41 Å apart from each other. The unique structural arrangement of the helicase and the C-terminal domains in Ph1280p probably plays a critical role in the specialized Ski2p function in the Ski complex. Further crystallographic studies on Ph1280p will provide more insight into the molecular mechanism of the helicase activity of the Ski2p-like RNA helicase.

Materials and Methods

Materials

Restriction enzymes were purchased from MBI Fermentas. The plasmid vectors used were pGEM^{TM}-T EASY and pET-22b from Promega and Novagen, respectively. *E. coli* strain JM109 was used as a host for cloning. *E. coli* strain BL21 Codon Plus (DE3) RIL (Stratagene) was used as a host for the expression of recombinant proteins. Crystal Screen I and II, Index, PEG/ION, and Crystal Screen Lite were obtained from Hampton Research. Wizard I and II were obtained from Emerald BioSystems.

Overproduction in E. coli

We attempted to overproduce the full-length Ph1280p, N-terminal helicase domain (residues 1–410), and C-terminal domain (residues 370–715) for crystallographic analysis. The genes encoding Ph1280p and individual fragments were amplified as single DNA fragments from genomic DNA of *P. horikoshii* OT3 (http://gib.genes.nig.ac.jp/single/main.php?spid=Phor_OT3) and the amplified DNA fragments were placed under the control of the phage promoter on the expression plasmid pET-22b using *Nde* I and *Not* I restriction sites. In these constructs, the gene products were expected to carry a His-tag sequence attached at the C terminus. Expression in *E. coli* strain BL21 Codon Plus (DE3) RIL was induced overnight with 1 mM IPTG at 20°C, as described in the manufacturer's manuals.

Purification of the proteins

The *E. coli* cells were centrifuged at 8000 rpm for 10 min at 4°C. The cell pellet was washed several times in Buffer A (50 mM Tris-HCl, pH 8.0, containing 200 mM NaCl) and then sonicated in Buffer A. After centrifugation at 12,000 rpm for 10 min, the supernatant was heated for 30 min at 70°C. The thermostable proteins were purified by affinity chromatography using a His-bind resin column (5 mL; Novagen), as described in the manufacturer's manual. The proteins were loaded onto the nickel-charged His-bind resin column, previously equilibrated with 15 mL of the binding buffer (20 mM Tris-HCl, pH 7.9, containing 5 mM imidazole and 0.5 M NaCl). After washing with two column volumes of the wash buffer (20 mM Tris-HCl, pH 7.9, containing 60 mM imidazole and 0.5 M NaCl), the adsorbed proteins were eluted with the elution buffer (20 mM Tris-HCl, pH 7.9, containing 1 M imidazole and 0.5 M NaCl), and then further purified by gel filtration on a Sephacryl S-200 column (1.5 × 70 cm) equilibrated with Buffer A. The expression and preparation of Se-Met-derivatized Ph1280p were carried out exactly as described for the native Ph1280p, except that host cells were from the *E. coli* B834 strains (Novagen) grown in minimal medium.

ATPase assay

The ATPase activity was monitored by using a colorimetric assay based on malachite green-molybdate as described (Chan et al. 1986; Pugh et al. 1999), using a BIOMOL GREEN (Biomol) phosphate detection kit. The reaction buffer contained 25 mM Tris-HCl (pH 7.5), 20 mM NaCl, 2 mM 2-mercaptoethanol, and 3 mM $MgCl_2$. Reactions were incubated in 20 μL for 15 min at different temperatures (37°C, 50°C, 60°C, 70°C, 80°C, and 90°C) and stopped with the addition of 5 μL of EDTA. For reactions with varying protein concentrations, 0.55 mM ATP and 0.05 mg/mL of herring sperm DNA (Promega) or yeast RNA (Sigma) were used.

Site-directed mutagenesis

Site-directed mutagenesis of Asp145 in Ph1280p was done using a Quick mutagenesis kit (Stratagene). The mutant proteins D145A and D145A, in which Asp145 was replaced with Ala and Asn, respectively, were purified by the same manner as those described for wild-type Ph1280, and the resulting mutants were characterized with respect to the DNA- and RNA-dependent ATPase activity, as described above.

Gel shift assay

The chemically synthesized oligonucleotide P1 (5′-CGGAG CUCGAAUUCGGAUCC-3′) was purchased from GENENET and labeled by T4 polynucleotide kinase using [γ-^{32}P] ATP. The oligonucleotide P2 (5′-TAATACGACTCACTATAGGGGGATCCGAATTCGAGCTCCG-3′) containing T7 RNA polymerase promoter sequence and the nucleotide sequence complementary to that of P1 was synthesized by in vitro transcription using T7

RNA polymerase; nucleotides complementary to those of P1 were underlined. The resulting RNA product P2 was purified on a 10% polyacrylamide gel in the presence of 8 M urea. The dsRNA used in gel shift assay was made by mixing P1 and P2 with a molar ratio of 1.5:1. The mixture was heated to 95°C for 5 min and then slow-cooled to 37°C over 90 min. Under this condition, ~100% of the labeled P1 was hybridized to P2. The synthesized dsRNA was stored at −30°C as described previously (Jankowsky et al. 2001). Gel shift assay was done with 1 μM of labeled RNA (ssRNA or dsRNA) and variable concentrations of the proteins, which were incubated together overnight on 25°C in a volume of 10 μL. After incubation, the reactions were stopped by adding 10 μL of a stop buffer containing 1% SDS, 50 mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue, and 20% glycerol. The products were analyzed by electrophoresis on 20% polyacrylamide gel at 4°C for 20 min at 400 V/cm. The gel was dried and then subjected to autoradiography as described previously (Cordin et al. 2004).

Crystallization

Crystallizations were done using screen kits, such as Crystal Screen I and II, Index, PEG/ION, Crystal Screen Lite (Hampton Research), and Wizard I and II (Emerald structures). The crystals suitable for crystallographic analysis were obtained as follows: The protein was concentrated to 14 mg/mL in 50 mM Tris-HCl (pH 8.0) containing 200 mM NaCl, and mixed with a half volume of 0.2 M Li_2SO_4, 0.1 M phosphate-citrate buffer (pH 4.2) containing 10% 2-propanol. The mixture was crystallized by hanging-drop vapor diffusion at 20°C. The crystals of the Se-Met derivative of Ph1280p were prepared under conditions identical to those described for native crystals.

Data collection and refinement

The crystals of Se-Met-derivatized Ph1280p were transferred to a cryoprotectant reservoir solution containing 20% (v/v) glycerol. The crystal mounted on a cryo-loop was flash-cooled in a nitrogen gas stream at 100 K. X-ray diffraction data were collected at the synchrotron radiation facility at BL41XU in SPring-8, Japan. MAD (multiwavelength anomalous dispersion) data were indexed, integrated, scaled, and merged by HKL2000 (Otwinowski and Minor 1996). The structure was, however, solved successfully by SAD using a peak data set that was collected at a wavelength of 0.97904 Å. Eight selenium sites of the 12 possible methionines in the molecule were located using SHELXD (Schneider and Sheldrick 2002) at 3.8 Å. The initial phases were calculated using SOLVE at 3.8 Å, and phases were then improved and extended to a resolution of 3.5 Å by RESOLVE (Terwilliger and Berendzen 1999). During phasing, the coordinates (x, y, z) and B-factor of Se atoms were refined while the occupancies estimated by SHELXD were fixed. Automated main-chain model building and density modifications were carried out by RESOLVE using the procedure supplied. The model was rebuilt using the program Coot (Emsley and Cowtan 2004). The model of the N-terminal helicase core domain of Ph1280p was built by superimposing a homology model, which was created from the model of the *E. coli* DNA helicase RecQ (PDB code: 1QYW) using the program MODELLER (Sali et al. 1995), on the α-carbon trace fragments and the electron density map using SSM (Krissinel and Henrick 2004) superposition implemented by Coot. The model of the helical C-terminal domain was built into the electron density map from semi-automated polyalanyl α-helix tracing. Several cycles of model building and phase improvement were carried out by RESOLVE. The model was refined with the CNS (Brunger et al. 1998) using positional and temperature factor refinement. The diffraction data statistics and the crystallographic refinement statistics are summarized in Table 1. The atomic coordinates have been deposited in the Protein Data Bank (accession code: 2Z41).

Acknowledgments

We thank Drs. M. Kawamoto and H. Sakai of the RIKEN Harima Institute for help with data collection using synchrotron radiation at beamlines. The synchrotron radiation experiments were done at BL41XU in SPring-8 with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) (proposal no. 2007A1205). This work was supported in part by a grant from the National Project on Protein Structural and Functional Analyses from the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

References

Anderson, J.S.J. and Parker, R. 1998. The 3′ to 5′ degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3′ to 5′ exonucleases of the exosome complex. *EMBO J.* **17:** 1497–1506.

Araki, Y., Takahashi, S., Kobayashi, T., Kajiho, H., Hoshino, S., and Katada, T. 2001. Ski7p G protein interacts with the exosome and the Ski complex for 3′ to 5′ mRNA decay in yeast. *EMBO J.* **20:** 4684–4693.

Benard, L., Carroll, K., Vallw, R.C., Masison, D.C., and Wickner, R.B. 1999. The Ski7p antiviral protein is an EF1-α homolog that blocks expression of non-Poly (A) mRNA in *Saccharomyces cerevisiae*. *J. Virol.* **73:** 2893–2900.

Bernstein, D.A., Zittel, M.C., and Keck, J.L. 2003. High-resolution structure of the *E. coli* RecQ helicase catalytic core. *EMBO J.* **22:** 4910–4921.

Brown, J.T., Bai, X., and Johnson, A.W. 2000. The yeast antiviral proteins Ski2p, Ski3p, and Ski8p exist as a complex in vivo. *RNA* **6:** 449–457.

Brunger, 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. Crystallography & NMR system (CNS): A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54:** 905–921.

Carmel, A.B. and Mathews, B.W. 2004. Crystal structure of the BstDEAD N-terminal domain: A novel DEAD protein from *Bacillus stearothermophilus*. *RNA* **10:** 66–74.

Caruthers, J.M. and McKay, D.B. 2002. Helicase structure and mechanism. *Curr. Opin. Struct. Biol.* **12:** 123–133.

Caruthers, J.M., Johnson, E.R., and McKay, D.B. 2000. Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc. Natl. Acad. Sci.* **97:** 13080–13085.

Chan, K.M., Delfert, D., and Junger, K.D. 1986. A direct colorimetric assay for Ca^{2+}-stimulated ATPase activity. *Anal. Biochem.* **157:** 375–380.

Cheng, Z., Coller, J., Parker, R., and Song, H. 2005. Crystal structure and functional analysis of DEAD-box protein Dhh1p. *RNA* **11:** 1258–1270.

Cordin, O., Tanner, N.K., Doere, M., Linder, P., and Banroques, J. 2004. The newly discovered Q motif of DEAD box RNA helicases regulates RNA-binding and helicase activity. *EMBO J.* **23:** 2478–2487.

Cordin, O., Banroques, J., Tanner, N.K., and Linder, P. 2006. The DEAD-box protein family of RNA helicases. *Gene* **367:** 17–27.

de la Cruz, J., Kressler, D., and Linder, P. 1999. Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. *Trends Biochem. Sci.* **24:** 192–198.

Doma, M.K. and Parker, R. 2006. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* **440:** 561–564.

Emsley, P. and Cowtan, K. 2004. Coot: Model-building tools for molecular graphics. *Acta Crystallogr.* **D60:** 2126–2132.

Gajiwala, K.S. and Burley, S.K. 2000. Winged helix proteins. *Curr. Opin. Struct. Biol.* **10:** 110–116.

Gorbalenya, A.E. and Koonin, E.V. 1993. Helicases: Amino acid sequence comparisons and structure–function relationships. *Curr. Opin. Struct. Biol.* **3:** 419–429.

Jankowsky, E., Gross, C.H., Shuman, S., and Pyle, A.M. 2001. Active disruption of an RNA-protein interaction by a DExH/D RNA helicase. *Science* **291:** 121–125.

Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., et al. 2001. Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7. *DNA Res.* **8:** 123–140.

Koonin, E., Wolf, Y., and Aravind, L. 2001. Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res.* **11:** 240–252.

Kraulis, P.J. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24:** 946–950.

Krissinel, E. and Henrick, K. 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr.* **D60:** 2256–2268.

Kyrpides, N.C. and Woese, C.R. 1998a. Archaeal translation initiation revisited: The initiation factor 2 and eukaryotic initiation factor 2B $\alpha-\beta-\delta$ subunit families. *Proc. Natl. Acad. Sci.* **95:** 3726–3730.

Kyrpides, N.C. and Woese, C.R. 1998b. Universally conserved translation initiation factors. *Proc. Natl. Acad. Sci.* **95:** 224–228.

Maquat, L.E. 2004. Nonsense-mediated mRNA decay: Splicing, translation and nRNP dynamics. *Nat. Rev. Mol. Cell Biol.* **5:** 89–99.

Matsumoto, Y., Sarkar, G., Sommer, S.S., and Wickner, R.B. 1993. A yeast antiviral protein, SKI8, shares a repeated amino acid sequence pattern with β subunits of G proteins and several other proteins. *Yeast* **9:** 43–51.

Merritt, E.A. and Bacon, D.J. 1997. Raster3D: Photorealistic molecular graphics. *Methods Enzymol.* **277:** 505–524.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. 1997. The exosome: A conserved eukaryotic RNA processing complex containing multiple 3′–5′ exoribonucleases. *Cell* **91:** 457–466.

Orban, T.I. and Izaurralde, E. 2005. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA* **11:** 459–469.

Otwinowski, Z. and Minor, W. 1996. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276:** 307–326.

Parker, R. and Song, H. 2004. The enzymes and control of eukaryotic mRNA turnover. *Nat. Struct. Mol. Biol.* **11:** 121–127.

Pugh, G.E., Nicol, S.M., and Fuller-Pace, F.V. 1999. Interaction of the *Escherichia coli* DEAD box protein DbpA with 23 S ribosomal RNA. *J. Mol. Biol.* **292:** 771–778.

Rhee, S.K., Icho, T., and Wickner, R.B. 1989. Structure and nuclear localization signal of the SKI3 antiviral protein of *Saccharomyces cerevisiae*. *Yeast* **5:** 149–158.

Rocak, S.A. and Linder, P. 2004. DEAD-box proteins: The driving forces behind RNA metabolism. *Nat. Rev. Mol. Cell Biol.* **5:** 232–241.

Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. 1995. Evaluation of comparative protein modeling by MODELLER. *Proteins* **23:** 318–326.

Schneider, T.R. and Sheldrick, G.M. 2002. Substructure solution with SHELXD. *Acta Crystallogr.* **D58:** 1772–1779.

Schwartz, T., Rould, M.A., Lowenhaupt, K., Herbert, A., and Rich, A. 1999. Crystal structure of the Zα domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science* **284:** 1841–1845.

Story, R.M. and Steitz, T.A. 1992. Structure of the recA protein-ADP complex. *Nature* **355:** 374–376.

Story, R.M., Li, H., and Abelson, J.N. 2001. Crystal structure of a DEAD box protein from the hyperthermophile *Methanococcus jannaschii*. *Proc. Natl. Acad. Sci.* **98:** 1465–1470.

Tanner, N.K., Cordin, O., Banroques, J., Doere, M., and Linder, P. 2003. The Q motif: A newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol. Cell* **11:** 127–138.

Terwilliger, T.C. and Berendzen, J. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr.* **D55:** 849–861.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680. doi: 10.1093/nar/22.22.4673.

van Hoof, A., Lennertz, P., and Parker, R. 2000. Yeast exosome mutants accumulate 3′-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Mol. Cell. Biol.* **20:** 441–452.

van Hoof, A., Frischmeyer, P.A., Dietz, H.C., and Parker, R. 2002. Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* **295:** 2262–2264.

Wang, L., Lewis, M.S., and Johnson, A.W. 2005. Domain interactions within the Ski2/3/8 complex and between the Ski complex and Ski7p. *RNA* **11:** 1291–1302.

Widner, W.R. and Wickner, R.B. 1993. Evidence that the *SKI* antiviral system of *Saccharomyces cerevisiae* acts by blocking expression of viral mRNA. *Mol. Cell. Biol.* **13:** 4331–4341.