Crystal structure of a DEAD box protein from the hyperthermophile *Methanococcus jannaschii*

Randall M. Story, Hong Li*, and John N. Abelson[†]

California Institute of Technology, Division of Biology 147-75, Pasadena, CA 91125

We have determined the structure of a DEAD box putative RNA helicase from the hyperthermophile *Methanococcus jannaschii*. Like other helicases, the protein contains two α/β domains, each with a recA-like topology. Unlike other helicases, the protein exists as a dimer in the crystal. Through an interaction that resembles the dimer interface of insulin, the amino-terminal domain's 7-strand β-sheet is extended to 14 strands across the two molecules. Motifs conserved in the DEAD box family cluster in the cleft between domains, and many of their functions can be deduced by mutational data and by comparison with other helicase structures. Several lines of evidence suggest that motif III Ser-Ala-Thr may be involved in binding RNA.

Essentially all processes involving RNA and DNA metabolism require, at some point, the unwinding and/or reorganization of nucleic acid duplexes, actions that are achieved by helicases. Helicases are among the most numerous proteins. For example in yeast, the seventh largest group of domains is from the DEAD/DEAH group of RNA-dependent ATPases and helicases (1). The DEAD and DEAH families are related closely, with their names derived from the sequence of the Walker B motif (motif II) in their ATPase active sites (2). A larger family includes DNA helicases as well and has seven characteristic "helicase motifs" distributed across a core domain of roughly 400 residues (3). Although they have generally been shown to be nucleic acid-dependent ATPases, many members of the "helicase" family have yet to display helicase activity *in vitro*, including such proteins as the Swi/Snf family of transcriptional activators as well as types I and III restriction enzymes (3, 4). In addition, some of the mRNA splicing factors may be involved in the proofreading and/or the rearrangement of short regions of base pairing (5). Such functions may seem quite different from a processive helicase, but they could ultimately reflect fundamental activities such as coupling ATP hydrolysis to conformational changes that affect binding of nucleic acid and/or drive translocation.

Despite the huge numbers of known helicases and their universal role in so many processes, high-resolution structural studies have lagged, with the determination of crystal structures only occurring over the last few years (6). Notwithstanding the considerable interest in the helicase family over the past several decades and the emerging structural data, the fundamental mechanism or mechanisms of helicase activity are still a subject of intense debate. We describe below an intact structure of a member of the very abundant DEAD/DEAH family, whose diverse functions have generated considerable interest, particularly over the last decade. The protein MjDEAD comes from the hyperthermophile *Methanococcus jannaschii*, consists of only 367 residues, and can serve as a model for understanding this family.

Materials and Methods

The ORF for MJ0669, a putative DEAD box protein from the hyperthermophile *M. jannaschii*, was isolated by PCR and ligated into the Pet16b T7 overexpression vector. The decahistidine-tagged protein was overexpressed in the BL21 cell line that contained the pRI952 plasmid for expression of the *argU* and *ileX* tRNA genes to accommodate codons that are rare in *Escherichia coli* (7). Cells were grown, and selenomethinione was incorporated by using the methionine biosynthesis suppression method of Van Duyne *et al.* (8). Protein was purified by sonication of a cell suspension in buffer containing 50 mM Tris (pH 7.5), 10% (vol/vol) glycerol, 1 mM EDTA, 5 mM DTT, 0.02% (wt/vol) NaN$_3$, and 200 mM ammonium sulfate. The MjDEAD protein remained in the cell pellet and was removed by extraction with the same buffer but with 500 mM, rather than 200 mM, ammonium sulfate. The protein was purified further by heating to 65°C to denature contaminating proteins, followed by Ni-nitrilotriacetic acid chromatography. The yield was 14 mg from a 2-liter culture. Amino acid analysis revealed loss of all methionine in the substituted protein. Protein was dialyzed and concentrated to 18 mg/ml in a solution of 20 mM Tris (pH 7.5), 0.02% (wt/vol) NaN$_3$, 500 mM ammonium sulfate, 100 mM NaCl, 1 mM EDTA, and 5 mM DTT. Crystals grew by mixing the protein solution 50:50 with a reservoir containing 57–62% saturated ammonium sulfate and 100 mM sodium acetate at a pH of 5.5–6.25 followed with equilibration by vapor diffusion. Crystals of roughly $0.4 \times 0.2 \times 0.2$ mm grew in several days at room temperature or 30°C. For data collection, crystals were transferred serially to a solution containing 65% saturated ammonium sulfate, 100 mM NaCl, 100 mM acetate (pH 5.5), 0.02% (wt/vol) NaN$_3$, 2 mM DTT, and ultimately 30% (wt/vol) xylitol.

Structure Solution and Refinement. The crystals are of space group C222 (1), diffract x-rays up to 3.0-Å resolution, and have a dimer in the asymmetric unit. Multiwavelength data were collected at the 5.0.2 beamline at the advanced light source (ALS) synchrotron (Lawrence Berkeley National Laboratory, Berkeley, CA) and processed by using the HKL program suite (9). Data statistics are summarized in Table 1. The structure was solved by multiple anomalous dispersion phasing (10). The heavy atom sites were located by using the automated Patterson search routine in the program SOLVE (www.solve.lanl.gov; ref. 11) and were later found to correspond to 16 of the 20 possible methionines in the two molecules. Phases calculated from these sites by SOLVE yielded an interpretable electron density map, which was further improved by noncrystallographic averaging between the two separate domains of the two protomers in the asymmetric unit with the program DM (12). Model building of the protein structure proceeded with the program O (13), followed by

BIOCHEMISTRY

Table 1. Statistics for three-wavelength multiple anomalous dispersion data collection and phase determination*

Data sets	Wavelength, Å	Resolution, Å	Measured reflections	Unique reflections	R_{sym}[†]	$\langle I \rangle / \langle \sigma(I) \rangle$	Completeness, %	Diffraction ratios[‡] λ_1	λ_2	λ_3
λ_1 (edge)	0.9795	25–2.95	319,748	20,143	0.095 (0.367)	8.9 (8.4)	100.0 (100.0)	0.066	0.028	0.035
λ_2 (peak)	0.9793	25–2.95	322,686	20,100	0.096 (0.379)	8.9 (8.6)	100.0 (100.0)		0.050	0.048
λ_3 (remote 1)	0.9611	25–2.95	322,650	20,138	0.097 (0.419)	8.7 (8.4)	100.0 (100.0)			0.051

Resolution, Å	Phasing power (actr. refl.)	MFOM	Corr. coef. of NCS operations (CT/NT)	R value	R_{free} value	Bond rms deviations, Å[§]	Angle rms deviations, °[§]	No. of nonhydrogen atoms
50–3.0	1.8 (1.17)	0.66 (0.46)	0.677/0.688	0.270	0.316	0.0091	1.4	5,850

actr. refl., acentric reflections; MFOM, mean figure of merit; corr. coef., correlation coefficient; NCS, noncrystallographic symmetry; CT/NT, carboxyl-terminal/amino-terminal.

*Numbers in parentheses correspond to those in the last resolution shell.

[†]R_{sym} indicates agreement of individual reflection measurements over the set of unique averaged reflections. $R_{sys} = \Sigma_{h,i} |I(h)_i - \langle I(h) \rangle| / \Sigma_{h,i} I(h)_i$ where h is the Miller index, i indicates individually observed reflections, and $\langle I(h) \rangle$ is the mean of all reflections of the Miller index h. Bijvoet mates are treated as independent reflections when computing R_{sym}.

[‡]Diffraction ratios are defined as $\langle (\Delta|F|)^2 \rangle^{1/2} / \langle |F|^2 \rangle^{1/2}$ where $\Delta|F|$ are taken between each two wavelengths for dispersive ratios and between matched Bijvoet pairs for anomalous ratios.

[§]Geometry values of the final protein model were compared with ideal values. The cutoff used in data processing, $I/\sigma I$, was −3, whereas the cutoff for refinement, $F/\sigma F$, was 0. Data (10%) were excluded from refinement and used as a test data set to calculate R_{free}.

extensive rebuilding and refinement including noncrystallographic symmetry restraints with the program CNS (14). The current model for both molecules consists of residues 3–365, lacking two residues at each terminus. Refinement statistics for the current model are in Table 1. Of the nonproline, nonglycine residues, 82.7% lie in the most favored region of a Ramachandran plot, and only 0.5% are in disallowed regions as defined by PROCHECK (15).

Results and Discussion

Monomer Structure. The MjDEAD monomer consists of two α/β domains, each with a fold of RecA-like topology (ref. 16; Fig. 1 *A* and *B*). This architecture is found in all known monomeric helicase structures; hexameric helicases contain one RecA-like domain per polypeptide (6). The two domains are quite similar–a superposition of 111 atoms gives a root mean square deviation (rmsd) of 1.87 Å for the amino- and carboxyl-terminal domains of MjDEAD, whereas each domain superimposes individually on the ATPase domain of RecA (94 atoms, 1.79 Å rmsd and 67 atoms, 1.93 Å rmsd for the amino- and carboxyl-terminal domains, respectively). The amino-terminal domain contains sequences that are associated with ATP binding and hydrolysis ("Walker motifs"; ref. 2) and that are found throughout a large family of NTPases. These motifs are absent in the carboxyl-terminal domain. There is no apparent sequence similarity between the domains, suggesting that the arrangement arose by gene duplication followed by substantial divergence.

All previously determined helicase structures have extra domains inserted into the two-domain core, whereas the structure reported here represents purely the common structural core. The core domains of all known structures superimpose on their counterparts in other proteins, suggesting a common ancestor for this large and diverse family (17). Generally, about 100–130 α-carbons in the amino-terminal domain superimpose with an rmsd of 2 Å on that of MjDEAD; PcrA, Rep, and UvrB are the most similar, followed by the hepatitis C virus NS3 (HCV) and T7 gene 4 helicases. These similarities do not correspond strictly to those expected for the superfamilies 1 and 2 (SF1 and SF2) of helicases (3); PcrA, for example, is an SF1 protein, HCV and DEAD box proteins are SF2. The core helicase domain can be thought of as a covalently linked "dimer" of two recA-like subdomains, where only one has an ATPase activity.

Of particular relevance is yeast eIF-4A, a DEAD box protein with 36% sequence identity to the *Methanococcus* protein. eIF4A is the prototype of this family and has served as the paradigm for understanding DEAD box proteins. Two essentially identical structures of the amino-terminal domain have been reported recently (18, 19). The amino-terminal domains of the *Methanococcus* and eIF4A DEAD box proteins superimpose very well (204 α-carbons with an rmsd of 1.15 Å), with the only significant differences in loops that connect the α/β framework, and are distant from the conserved motifs and presumed ATPase active site. Because the sequence conservation is spread throughout the primary sequence, we expect that the overall structures of these two DEAD box proteins will be correspondingly similar and that the conserved features thought to be most important for function among DEAD box proteins will be found in both domains of both structures. As even the two-domain eIF4A possesses helicase activity *in vitro* (20), the structure described here can serve as a model for a minimal helicase.

A departure of the MjDEAD structure from the other helicase structures is that the relative orientation of the amino- and carboxyl-terminal domains is significantly different (Fig. 1*C*). In other proteins there are relatively small differences in the proximity of these two domains, including a closing up of the cleft between the two domains upon ATP/DNA binding, with large motions occurring only in the extra domains (21–23). The MjDEAD structure exhibits a huge "opening up" of the domains relative to other helicases. A large motion would be necessary, for example, to reposition one of the domains to achieve the conformation of the ATP/DNA-bound PcrA. We assume that the "closed form" represents a state common to all of the two-domain helicases and will model MjDEAD in this conformation (Fig. 1*C*) for a number of analyses throughout this paper.

The unusual orientation of the domains could merely be caused by the selection by crystal packing of one conformation of a flexible protein. However, because the approximate 2-fold symmetry axis is noncrystallographic, the packing environment of each monomer of the dimer is, in fact, entirely different. If indeed the protein is flexible, the surprising selection of the same open conformation of both dimer subunits, despite unique crystal contacts, would be more reasonable if the structure of both subunits of the dimer were already similar in solution. Such similarity would make sense if either conformational changes of the individual subunits within the dimer are cooperative or if the open structure itself is reasonably stable. The linker between the domains is far from the dimer interface (Fig. 2*A*), suggesting that the conformations of the subunits might be relatively uncoupled, arguing for a relatively stable conformation. The open structure could simply be caused by the lack of interacting proteins or
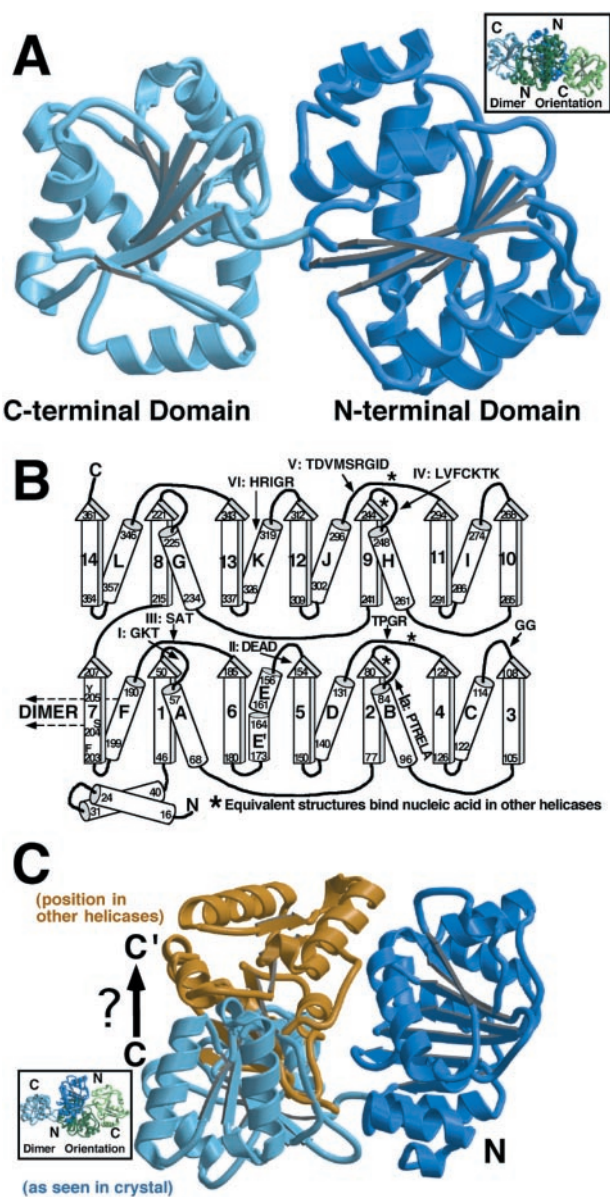
Fig. 1. Monomer structure. (*A*) The MjDEAD monomer showing the amino- and carboxyl-terminal domains (labeled N and C in subsequent figures). The linker between the domains can be seen in the middle of the figure. The orientation of the dimer in this view (in this and subsequent figures) is depicted in an *Inset*, with the equivalent domains colored blue, as in the main figure. Fig. 1*A*, as well as Fig. 1*C*, Fig. 2 *A* and *B*, and Fig. 3 *A* and *C* were made with MOLSCRIPT (41) and RASTER 3D (42). (*B*) The topological organization of the MjDEAD monomer, illustrating the similarities of the two domains. The "RecA-like core" stretches from β-strands 1 and 2 and 4–7 as numbered for the amino-terminal domain and their connecting α-helices. Sequence numbers at the edges of secondary structure elements are indicated, as are those loop regions observed to bind the nucleic acid backbone in some or all of known helicase complexes with nucleic acid. The region of polypeptide equivalent to the GG motif (motif 1*B*) also contacts nucleic acid in the HCV NS3 helicase. α-Helix F and β-strand no. 7 (that pack against their symmetry-related counterparts to form a dimer) are indicated. (*C*) Difference in the amino- and carboxyl-terminal domain orientation relative to other proteins. Superposition of only the amino-terminal domain with that of other proteins reveals a structure "opened up" relative to the others (blue domains). Independent superposition of the carboxyl-terminal domain on a "closed" structure (in this case the PcrA DNA and AMPPNP structure) leads to a closing of the MjDEAD structure to a conformation more like that observed for other helicases (blue amino-terminal domain and gold carboxyl-terminal domain). Single-stranded (ss)DNA binds at the top of the two domains in other helicases in this orientation. We assume that this closed structure for MjDEAD will likewise resemble the structure of the DNA/ATP bound form of this enzyme.
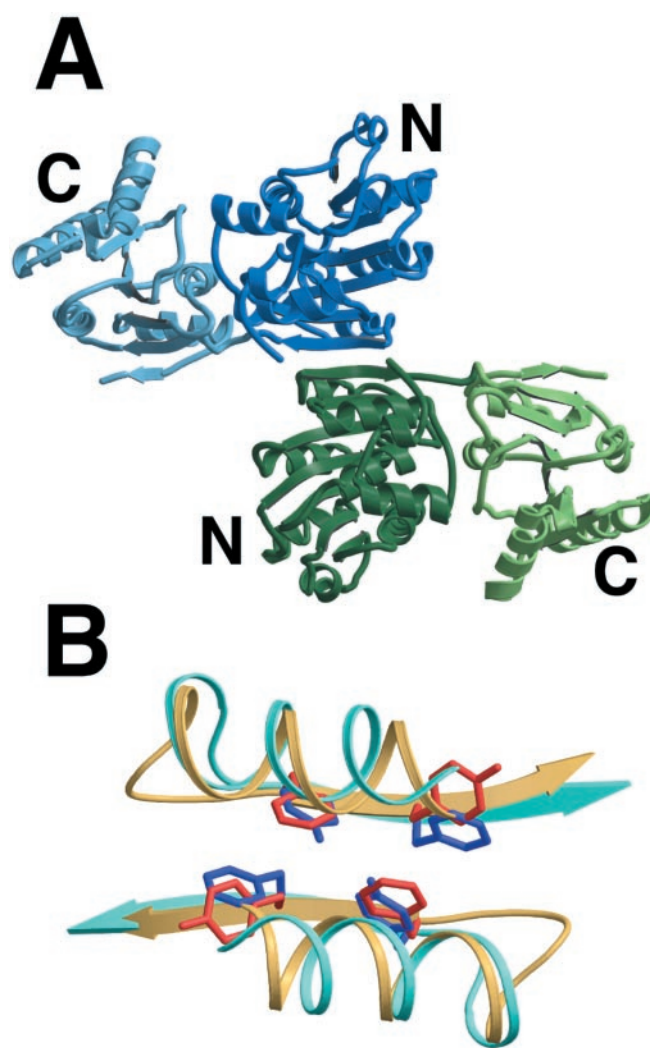

Fig. 2. Dimer Structure. (*A*) Structure of the MjDEAD dimer found in the asymmetric unit in the crystal with the molecules related by an approximate 2-fold symmetry axis. The individual monomers are shown in blue and green. Two equivalent β-strands (no. 7) are hydrogen bonded, effectively extending the β-sheet to 14 strands. (*B*) Closeup of the dimer interface of MjDEAD and its superposition with the B:D interface of insulin (PDB code 1trz). Each interface is created by a similar interaction across a roughly 2-fold symmetry axis of two α-helices and two hydrogen-bonded β-strands, depicted as coils and arrows. Insulin is colored yellow with red side chains, and MjDEAD is colored cyan with blue side chains. The arrangement of equivalent aromatic residues on the β-strands (YsF for MjDEAD, FfY for insulin) is shown.

substrates. Alternatively, large relative motions of two domains might substitute for the large motions of the extra domains observed in other structures (22, 23). An open conformation and/or flexibility in the linker region may be related to the observation of ATP/ADP- and RNA-dependent sites of cleavage in eIF4A (24), which helped define the amino- and carboxyl-terminal domains before a structure. Surprisingly, although the carboxyl-terminal domain in crystals of intact eIF4A (19) was ordered too poorly to allow interpretation, the relative orientation of the two domains was clear, and it also differed radically from that observed in other helicase structures.

Dimer Structure. The basic mechanism of helicase action is still a matter of controversy. The rolling (25) and the inchworm (26, 27) models have emerged as the primary candidates for a general mechanism. Whereas the rolling model requires a dimer or

higher oligomer, the inchworm model could function with either a monomer or oligomer. Published structural data seem to favor the inchworm model heavily, partly because no helicases have crystallized as dimers, even in the presence of nucleic acid.

The MjDEAD protein, however, exists as a dimer in the crystal. The two molecules in the asymmetric unit interact, forming a dimer related by approximate 2-fold symmetry (Fig. 2A). Hydrogen bonding between the final β-strand (no. 7) of each amino-terminal domain creates a 14-strand β-sheet that spans the two molecules. Additional interactions include a number of tightly packed hydrophobic interactions, such as an interdigitation between symmetry related Phe-203 and Tyr-205 residues on the hydrogen-bonded β-strands and packing of helix F with the symmetry related helix F. Solvent-exposed surface (1,400 Å^2) would be buried by formation of the dimer. This number is reasonable but is at the lower end of what is observed typically for dimer interfaces (28). A comparison of functional interfaces with crystal packing interactions (29) suggests a 98% probability that a surface area of this size reflects a true dimer interface.

Surprisingly, the interface bears a striking similarity to the interface of the insulin dimer (Fig. 2B; ref. 30). Insulin dimerizes by an approximate 2-fold interaction of nearly identical α/β structures from the B and D chains. Superposition of the two helices and two β-strands of the MjDEAD interface with their insulin counterparts yields an rmsd of 2.0 Å for 42 atoms. Remarkably, there is an interaction analogous to the zipper-like arrangement of four aromatic side chains contributed by Y203 and F205 of each MjDEAD monomer; the four corresponding insulin residues are F24 and Y26 from both the B and D chains (Fig. 2B). A database that groups protein interfaces into 351 families does not indicate any proteins unrelated to insulin that have such an interface (31), suggesting that this is a rare interaction and a remarkable example of convergent evolution.

The related yeast eIF4A does not show formation of a dimer in any of its crystal forms. β-Strand no. 7 contains the sequence PVR instead of the YSF found in MjDEAD (FFY in insulin) that contributes aromatics to the interface. Moreover, in the eIF4A structures, residues 12–14 (not found in MjDEAD) add a short antiparallel β-strand instead. We did database searches to see whether similar sequences are found in other helicases, looking for a sequence Aromatic-X-Aromatic, placed roughly 16 residues from the Ser-Ala-Thr (SAT) motif. We found several proteins with this potential motif, including Ded1 of yeast and a number of related proteins such as DEAD box protein 3 (human and mouse) and An3 (Xenopus), as well as several members of the Vasa family. We do not know whether these sequence similarities are meaningful, and we are not aware that any of these proteins have been shown to be dimers.

We also found that members of the Brr2 family (a yeast protein involved in disruption of U4/U6; ref. 32) have such a sequence two or three additional residues downstream relative to the SAT motif, consistent with a small insertion in the loops between secondary structure elements as in insulin. Intriguingly, this family of large proteins has two helicase domains, with the aromatic motif found in both domains of all representatives (FYF and YNF in Brr2), raising the possibility that these domains have an analogous dimer-like interaction within the protein.

This dimer does not necessarily lend support to the rolling model, because dimerization is compatible with an inchworm mechanism. Moreover, although the character and insulin-like structure of the dimer interface strongly suggest that it is real and exists at some point, we do not know whether it is present under conditions that would allow a potential helicase activity. However, it does provide a model for how such a functional dimer could be formed. Although any helicase activity of the MjDEAD protein is uncharacterized, the closely related eIF4A is a well characterized helicase (20), and the structure of the partially ordered intact form (19) of the protein does not indicate similar dimerization. It is difficult to imagine that two closely related proteins could function by such dissimilar mechanisms.

One possible function for the dimer could be to provide extra domains comparable to those that are found in larger monomeric helicase structures or that are contributed by additional proteins such as eIF4B to eIF4A. However, the noncrystallographic 2-fold puts the second subunit on the opposite side of the first, relative to the extra domains of other helicases and to the expected sites of 3′-ssRNA binding (Fig. 3C). Nonetheless, a plausible model for RNA binding in which an RNA duplex lies between the two domains suggests that the duplex or an unwound 5′ strand could approach the other subunit of the dimer.

Conserved Motifs. Motif I (GxxGKT/S, also "A" motif) is a well characterized sequence that was first identified by Walker et al. (2) by comparing the amino acid sequences of F1-ATPase, myosin, adenylate kinase, and RecA and that has since been characterized structurally in numerous NTPases (33). This phosphate binding loop or so-called "P-loop" forms a pocket that binds the β- and γ-phosphates of the NTP by contacts with the polypeptide backbone as well as with the invariant lysine and threonine/serine. As expected, mutations in this region of DEAD box proteins reduce or eliminate ATPase activity yet sometimes preserve ATP binding (34, 35). The first D and the E of DEAD (motif II, Walker "B") and their analogs have also been characterized structurally and by mutagenesis in a number of proteins and have shown to be involved in water-mediated Mg binding and catalysis, respectively (16, 35, 38). The final D interacts with SAT (motif III) as described below.

A region of strong electron density was present in the original experimental electron density map at the center of the P-loop. This was also the site of one of the two strongest peaks (see also motif III below) in the F_o–F_c difference electron density map after fitting the initial protein model. We interpret this peak as a sulfate ion occupying the site for binding the β-phosphate of ATP. A bound sulfate at this site is frequently observed with NTPases crystallized in the presence of high concentrations of ammonium sulfate, including helicases such as HCV, PcrA, and one of the yeast eIF-4A structures (18, 21, 22). The loop is not in the collapsed form seen in the other eIF-4A structure (19).

Motif III has been suggested, based on mutagenic data, to be involved in the "helicase" activity, because mutations in this motif generally knock out this activity while preserving both ATP hydrolysis and RNA binding (34, 35). In mammalian eIF4A the SAT to AAA mutation actually increases K_{cat}/K_m of ATP hydrolysis by 240% (K_m 75% and V_{max} 173% of wild type), whereas RNA crosslinking decreased only slightly to 67% of wild type. The mutagenesis findings are consistent with the idea that this motif is involved in the coupling of ATP hydrolysis to RNA binding, such that mutations do not destroy these activities themselves but rather their coupling—and thus helicase activity. Additionally, mutations of this motif in Prp2 and Prp43 are dominant negative (36, 37), leading to accumulation of inactive splicing complexes for Prp2. Together, these data are consistent with the idea that mutation of this motif destroys "helicase" activity, leaving the enzyme "stuck" to the nucleic acid, which could improve ATPase while preserving RNA-binding activity.

The related RecA protein was proposed to couple DNA binding to ATP hydrolysis by way of a conserved glutamine residue in a loop that is structurally analogous to motif III and is thought to bind DNA (16). The DNA helicase PcrA also has a glutamine at this location (Q254), and subsequent residues in this loop directly contact bases of the displaced 3′ ssDNA. Based on this similarity, as well as structural and mutagenic data, Wigley and coworkers (38) have proposed a coupling mechanism common to the RecA family. MjDEAD, however, has a much
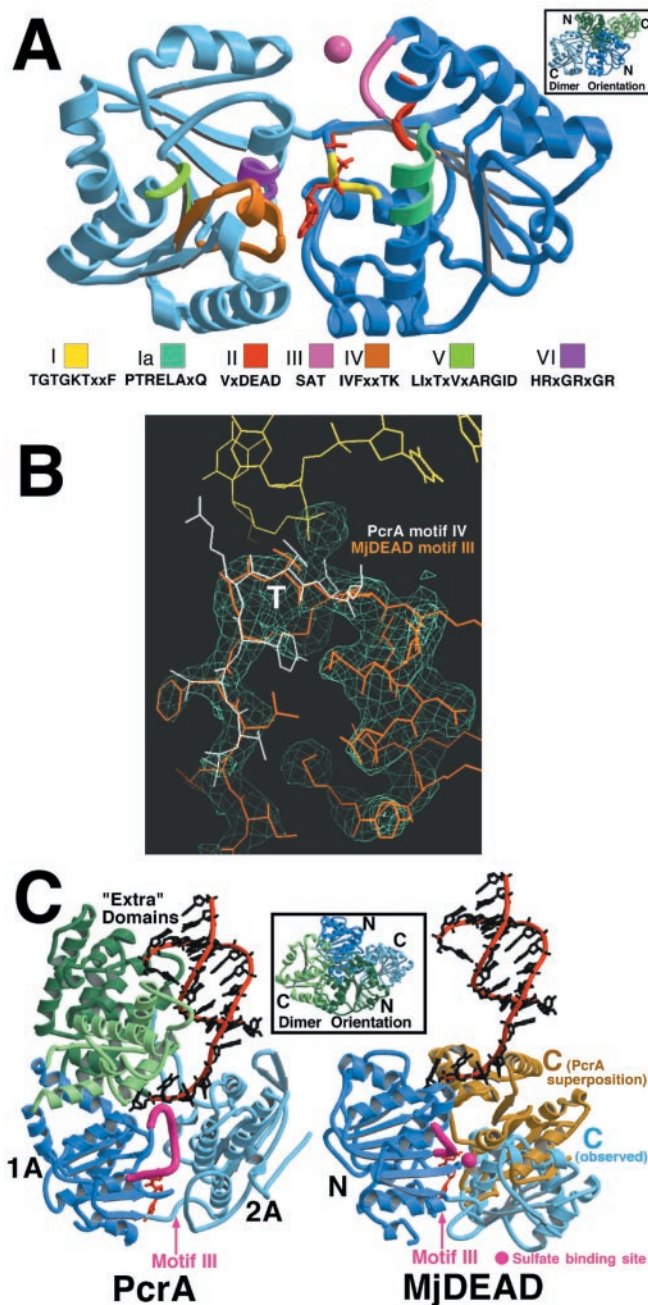
Fig. 3. Conserved motifs and their functions. (*A*) Conserved motifs of the MjDEAD monomer. The conserved motifs I–VI of the DEAD box family are indicated. This monomer is the opened form of the protein—the ''closed'' form would bring the motifs into closer proximity. A pink sphere represents a likely sulfate ion bound to motif III. The ATP is model-built, based on its position when PcrA is superimposed. (*B*) Alignment of motif III (SAT) with motif IV of PcrA. The experimental electron density (averaged multiple anomalous dispersion phases) is shown for MjDEAD at 1.4 σ. Alignments with motifs Ia and IV of MjDEAD or other helicases are similar and include the equivalent labeled threonine. A 2.3-Å structure of eIF4A shows an identical conformation (19). The subsequent α-helix of PcrA is out of register with that of MjDEAD and has been omitted for clarity. The depicted region of PcrA binds to a phosphate of the DNA backbone, shown in yellow. The MjDEAD motif III has a sulfate ion strongly bound in the same position relative to the aligned structure. (*C*) The complex of PcrA with tailed duplex DNA and AMPPNP (ref. 22; 2pjr) and its comparison with MjDEAD. The amino-terminal domain of MjDEAD has been superimposed on domain 1A of PcrA; the carboxyl-terminal domain does not align and must be superimposed separately on domain 2A as depicted by the gold domain and in Fig. 1*C*. The position of DNA that results from the 1A/N superposition is shown modeled relative to MjDEAD. The observed carboxyl-

shorter motif III loop than PcrA and seems unable to contact the 3′ single-stranded (ss) nucleic acid analogously (Fig. 3*C*). Moreover, there is not a glutamine or apparently any residue in this loop that could provide an obvious contact with the γ-phosphate.

However, we made the surprising observation that the SAT motif (motif III) is quite similar in structure to motifs Ia and IV that bind DNA in other helicases (Fig. 3*B*). Eight α-carbons of MjDEAD superimpose on comparable structures in motifs Ia and IV of MjDEAD and motifs Ia and IV of PcrA with rmsds of 0.581, 0.826, 0.730, and 0.592 Å, respectively. The region of superposition includes four residues in a β-strand plus four residues in the following loop. Strikingly, the threonine residue of the SAT motif also superimposes on the equivalent threonine of the motif Ia and IV structures, where it is involved in making contacts with the DNA phosphates. Intriguingly, the MjDEAD structure showed a strong peak in this region in the experimental electron density maps that was, along with the peak at the P-loop, one of the two largest peaks in the F_O–F_C difference map (and, presumably, also a bound sulfate ion). This peak also coincides that with the relative location of the DNA phosphate bound to motifs Ia and IV of structures of helicase/DNA complexes (Fig. 3*B*), suggesting that it may be a genuine site for binding nucleic acid. Remarkably, the equivalent TAT region of hPrp28 was found to crosslink to the 5′ splice site RNA (39).

Binding of nucleic acid to motif III (SAT) would be consistent with mutagenic data, suggesting a role in coupling ATP hydrolysis to RNA binding, as well as models of a general mechanism for coupling ATP hydrolysis to nucleic acid binding in RecA-like proteins. Although it seems unlikely that motif III of MjDEAD could make a direct contact with the γ-phosphate of ATP, it does interact with the final aspartate of the DEAD motif (motif II), as in eIF4A (19), by potential contacts with both the backbone amino hydrogens and side chain hydroxyls of Ser-185 and Thr-187 of SAT (distances of 3.15–3.90 Å for five donor–acceptor pairs on each subunit). Additionally, the first aspartate of DEAD is expected to make a water-mediated contact with the active site magnesium, whereas the glutamate in the second position is thought to provide the general base for ATP hydrolysis (38). Thus, motif III could be indirectly coupled to ATP hydrolysis by way of motif II (and motif VI as described below).

Because motif III is not in the path of the 3′ ss nucleic acid observed to bind across the two domains in other helicases (although a longer motif III does bind ssDNA in PcrA and Rep), it may have a different function. For example, this region might be a binding site for the 5′ strand displaced from a duplex, a binding site for duplex RNA, or a site for ''holding on'' to nucleic acid released transiently during translocation. A model for RNA binding in which the 3′ strand extends directly into an RNA duplex that lies in the groove between domains places the duplex near motif III. Although MjDEAD apparently lacks a glutamine or any other residue in motif III to interact with the γ-phosphate and is not capable of contacting nucleic acid like PcrA, the proposal of a general model whereby ATP hydrolysis is coupled to nucleic acid binding by this loop is still reasonable.

At least some of the function or functions of motifs Ia, IV, and V can be deduced by inspection of other helicase structures, particularly those complexed with DNA. In the cocrystal structures of HCV, PcrA, and Rep protein, these motifs as well as the equivalent of the TPG sequence (motif Ic) are observed to bind

terminal domain is not positioned to bind nucleic acid analogously, and we assume that a conformational change must occur, as in Fig. 1*C*. Motif III in both proteins is shown in pink—that of MjDEAD is too short to bind RNA analogous to binding of the 3′ ssDNA tail by PcrA. However, it does bind a sulfate ion in the crystal, depicted as a pink ball.

the phosphate backbone of DNA. We fully expect analogous binding of RNA to the MjDEAD protein, because the loops that interact with the DNA backbone in the other helicases have structural counterparts in MjDEAD, including, in most cases, an identically positioned conserved threonine [TPG and motif V also superimpose on motifs IV and Ia; although in this case, the alignment consists of the threonine-containing loop and the following α-helix (not shown)]. In HCV, the polypeptide equivalent to the GG motif contacts DNA as well. As noted in the HCV/DNA complex (21), because of the two recA-like domains, there is symmetry in these binding sites so that motifs Ia and IV are essentially equivalent, as are the TPG loop and motif V. There are no obvious aromatic residues (possibly residue F164) of the sort observed to bind DNA by PcrA and thought to be important for helicase activity (22). Nor is it obvious why the DEAD box proteins have specificity for RNA in both nucleic acid stimulated ATPase and helicase activity.

It was noticed early on that the DEA**D** box proteins have **H**RIGR in motif VI, whereas DEA**H** proteins have **Q**RIGR. The finding that the equivalent amino acids highlighted in motifs II and VI can be paired as either D:H or H:Q led to the suggestion that an analogous interaction can be achieved by either pairing and was supported by mutagenesis of these motifs in eIF4A (34). The HCV/DNA structure indicated that the proposed H:Q interaction (in HCV the sequences are DEC**H** and **Q**RRGR) is indeed reasonable (21). Closure of the MjDEAD protein to an ATP-like state (Fig. 1*C*) moves the analogous D and H residues from roughly 13 to 5 Å apart, supporting the proposal that these residues in motifs II and VI could interact in the closed state.

Motif VI was also proposed to couple ATP hydrolysis by way of a mechanism essentially identical to that described for motif III (34, 40). At the time, the H/QRIGR in DEAD/DEAH

proteins was thought to be an RNA-binding motif, leading to the suggestion that the His/Gln is positioned in the equivalent location as on RecA, such that an analogous interaction with the γ-phosphate is coupled to RNA binding. Unexpectedly, although about the same size as RecA, the DEAD "helicase domain" consists of two RecA-like domains, with the HRIGR motif on the second (non-ATPase) domain. Remarkably, the histidine is about 4–5 Å from the γ-phosphate of the model-built ATP in either the observed or the "closed" forms of the MjDEAD protein suggesting that elements of the model may be correct. Additionally, this histidine is in roughly the same region of space as PcrA's Gln-254, thought to be involved in the coupling reaction (as well as Arg-287 and Arg-610), and might replace the functions of one or more of these residues.

Even if the HRIGR motif is not involved in RNA binding, its location in the interdomain cleft suggests that it may in fact lead to a conformational change or changes that affect binding and/or ATP hydrolysis. Its proximity to the DEA**D** residue that interacts with a potential RNA-binding motif (SAT motif) offers another mechanism for mediating a structural change. Thus, conserved residues of motifs II, III, and VI (DEAD, SAT, and HRIGR) may be involved in a network that couples ATP binding to conformational changes that affect RNA binding.

1. Rubin, G. M., Yandell, M. D., & Wortman, J. R., Miklos, G. L. G., Nelson, C. R., Hariharan, I. K., Fortini, M. E., Li, P. I., Apweiler, R., Fleischmann, W., *et al.* (2000) *Science* **287,** 2204–2215.
2. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982) *EMBO J.* **1,** 945–951.
3. Gorbalenya, A. E. & Koonin, E. V. (1993) *Curr. Opin. Struct. Biol.* **3,** 419–429.
4. Gorbalenya, A. E. & Koonin, E. V. (1991) *FEBS Lett.* **291,** 277–281.
5. Staley, J. P. & Guthrie, C. (1998) *Cell* **92,** 315–326.
6. Soultanas, P. & Wigley, D. B. (2000) *Curr. Opin. Struct. Biol.* **10,** 124–128.
7. Del Tito, B. J. D., Ward, J. M., Hodgson, J., Gershater, C. J. L., Edwards, H., Wysocki, L. A., Watson, F. A., Sathe, G. & Kane, J. F. (1995) *J. Bacteriol.* **177,** 7086–7091.
8. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993) *J. Mol. Biol.* **229,** 105–124.
9. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276,** 307–326.
10. Ramakrishnan, V. & Biou, V. (1997) *Methods Enzymol.* **276,** 538–557.
11. Terwilliger, T. C. & J. Berendzen. (1999) *Acta Crystallogr. D* **55,** 849–861.
12. Cowtan, K. (1994) *Joint CCP4 ESF-EACBM Newsl. Protein Crystallogr.* **31,** 34–38.
13. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47,** 110–119.
14. 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) *Acta Crystallogr. D* **54,** 905–921.
15. Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291.
16. Story, R. M. & Steitz, T. A. (1992) *Nature (London)* **355,** 374–376.
17. Korolev, S., Yao, N. H., Lohman, T. M., Weber, P. C. & Waksman, G. (1998) *Protein Sci.* **7,** 605–610.
18. Benz, J., Trachsel, H. & Baumann, U. (1999) *Structure (London)* **7,** 671–679.
19. Johnson, E. R. & McKay, D. B. (1999) *RNA* **5,** 1526–1534.
20. Rogers, G. W., Richter, N. J. & Merrick, W. C. (1999) *J. Biol. Chem.* **274,** 12236–12244.
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. Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S. & Wigley, D. B. (1999) *Cell* **97,** 75–84.
23. Korolev, S., Hsieh, J., Gauss, G. H., Lohman, T. M. & Waksman, G. (1997) *Cell* **90,** 635–647.
24. Lorsch, J. R. & Herschlag, D. (1998) *Biochemistry* **37,** 2194–2206.
25. Lohman, T. M. & Bjornson, K. P. (1996). *Annu. Rev. Biochem.* **65,** 169–214.
26. Bird, L. E., Brannigan, J. A., Subramanya, H. S. & Wigley, D. B. (1998) *Nucleic Acids Res.* **26,** 2686–2693.
27. Yarranton, G. T. & Gefter, M. L. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 1658–1662.
28. Miller, S., Lesk, A. M., Janin, J. & Chothia, C. (1987) *Nature (London)* **328,** 834–836.
29. Janin, J. (1997) *Nat. Struct. Biol.* **4,** 973–974.
30. Adams, M. J., Blundell, T. L., Dodson, E. J., Dodson, G. G., Vijayan, M., Baker, E. N., Harding, M. M., Hodgkin, D. C., Rimmer, B. & Sheat, S. (1969) *Nature (London)* **224,** 491–495.
31. Tsai, C. J., Lin, S. L., Wolfson, H. J. & Nussinov, R. (1996) *J. Mol. Biol.* **260,** 604–620.
32. Raghunathan, P. L. & Guthrie, C. (1998) *Curr. Biol.* **8,** 847–855.
33. Saraste, M., Sibbald, P. R. & Wittinghofer, A. (1990) *Trends Biochem. Sci.* **15,** 430–434.
34. Pause, A. & Sonenberg, N. (1992). *EMBO J.* **11,** 2643–2654.
35. Linder, P., Gasteiger, E. & Bairoch, A. (2000) *Yeast* **16,** 507–509.
36. Plumpton, M., McGarvey, M. & Beggs, J. D. (1994) *EMBO J.* **13,** 879–887.
37. Arenas, J. E. & Abelson, J. N. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 11798–11802.
38. Dillingham, M. S., Soultanas, P. & Wigley, D. B. (1999) *Nucleic Acids Res.* **27,** 3310–3317.
39. Ismaili, N., Sha, M., Gustafson, E. H. & Konarska, M. M. (2000) *RNA*, in press.
40. Pause, A., Methot, N. & Sonenberg, N. (1993) *Mol. Cell. Biol.* **13,** 6789–6798.
41. Kraulis P. (1991) *J. Appl. Crystallogr.* **24,** 946–950.
42. Merritt, E. A. & Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524.
43. Caruthers, J. M., Johnson, E. R. & McKay, D. B. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 13080–13085.