Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase

Jonathan M. Caruthers, Eric R. Johnson, and David B. McKay*

Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305

Communicated by Manuel F. Morales, University of the Pacific, San Francisco, CA, September 20, 2000 (received for review August 24, 2000)

The eukaryotic translation initiation factor 4A (eIF4A) is a member of the DEA(D/H)-box RNA helicase family, a diverse group of proteins that couples an ATPase activity to RNA binding and unwinding. Previous work has provided the structure of the amino-terminal, ATP-binding domain of eIF4A. Extending those results, we have solved the structure of the carboxyl-terminal domain of eIF4A with data to 1.75 Å resolution; it has a parallel α-β topology that superimposes, with minor variations, on the structures and conserved motifs of the equivalent domain in other, distantly related helicases. Using data to 2.8 Å resolution and molecular replacement with the refined model of the carboxyl-terminal domain, we have completed the structure of full-length eIF4A; it is a "dumbbell" structure consisting of two compact domains connected by an extended linker. By using the structures of other helicases as a template, compact structures can be modeled for eIF4A that suggest (*i*) helicase motif IV binds RNA; (*ii*) Arg-298, which is conserved in the DEA(D/H)-box RNA helicase family but is absent from many other helicases, also binds RNA; and (*iii*) motifs V and VI "link" the carboxyl-terminal domain to the amino-terminal domain through interactions with ATP and the DEA(D/H) motif, providing a mechanism for coupling ATP binding and hydrolysis with conformational changes that modulate RNA binding.

Eukaryotic initiation factor 4A (eIF4A) is a member of a protein family referred to as DEA(D/H)-box RNA helicases. [The "DEA(D/H)" motif by which this family is named is the Walker "B" motif that participates in binding MgATP]. Their identification as helicases historically is derived from conserved signature sequences first described by Gorbalenya and Koonin that are characteristic of RNA and DNA helicase proteins in general (1). Sequences for more than 500 representatives of this protein family, ranging in size from ≈400 aa to >1,200 aa, are available [see, for example, Pfam family 00270 (2)]; what they have in common is a conserved, ≈400-residue polypeptide segment that includes the helicase signature sequences. These proteins participate in diverse activities that involve interactions with RNA; representative activities of these proteins include RNA splicing (3), ribosome biogenesis (4), and RNA degradation (5). Moreover, the activities of many proteins that are identified by sequence similarity as members of this family remain obscure. Our present subject, the 394-residue yeast eIF4A protein, is a prototype, minimal DEA(D/H)-box helicase.

eIF4A participates in the initiation of polypeptide synthesis. Its apparent role is to facilitate the "melting" of secondary structure in mRNAs that might otherwise impede translation initiation [for reviews, see Pain (6) and Linder *et al.* (7)]. To accomplish this, eIF4A works in concert with other initiation factors; specifically, it is targeted to the 5′ cap region of mRNA by eIF4E (the cap-binding protein) and eIF4G, and it effects its helicase activity in cooperation with eIF4B (8); eIF4A has only a very weak helicase activity on its own (9). Thus, the intrinsic biochemical activities of eIF4A, which include RNA binding and an ATPase activity, are a necessary, but not sufficient component for the complex biological activities in which it participates. An analogous situation arises with many other proteins of this family, where a minimal DEA(D/H)-box helicase domain often works in conjunction with other proteins or, alternatively, with other domains within a large (≫400-residue), single polypeptide to effect a particular activity on a specific RNA substrate. By implication, the minimal DEA(D/H)-box helicase domain is a necessary, but often not sufficient, requirement for the activities in which these proteins participate.

In the classification of Gorbalenya and Koonin, helicases can be divided into several different "superfamilies" (1). Many DNA helicases fall into superfamily 1 (SF-1); among this group, crystallographic structures are available for the *Escherichia coli* Rep helicase (10) and the *Bacillus stearothermophilus* PcrA helicase (11–13). Crystallographic structures also are available for the UvrB DNA helicase (14–16) and hepatitis C virus (HCV) RNA helicase (17–20), which are members of superfamily 2 (SF-2). Although the DNA helicases and the HCV RNA helicase have substantial differences in structure—for example, the Rep and PcrA have four distinct structural domains, whereas HCV helicase has only three—it was found that the two domains that harbor the conserved helicase motifs have similar folding topologies and tertiary structures and that the conserved motifs, despite differing in sequence between the SF-1 and SF-2 families, superimpose at the same spatial positions within the structures (21).

The DEA(D/H)-box helicases fall into the SF-2 family in the classification of Gorbalenya and Koonin (1). The structure of the amino-terminal domain (residues 1–223) of eIF4A has been reported (22, 23); it is a parallel α-β domain, a subfragment of which is identical in folding topology and tertiary structure with the other helicases [and also with the recA protein (24)]. In this work, we report the structure of the carboxyl-terminal domain of eIF4A and discuss its relationship to other helicases. We also report the structure of full-length eIF4A. Using this information, we model hypothetical oligonucleotide-binding structures for eIF4A on the templates of other helicase structures. These models suggest involvement of the conserved helicase motifs of eIF4A in its functional activities.

Materials and Methods

Protein Expression and Purification. A plasmid for recombinant expression of a carboxyl-terminal domain of yeast eIF4A (amino acid residues 230–394 with a L230 M mutation at the "start" codon) was derived from an expression plasmid for full-length eIF4A (23) essentially by deletion of codons 2–230 as follows: the original plasmid has an *Nde*I restriction site (CATATG) at the "start" codon (underlined); a second *Nde*I site was introduced at codons 229–230 by using the Stratagene "Quik-change" mutagenesis method; the resulting plasmid was digested with *Nde*I and religated to yield the desired deletion. Protein was expressed at 37°C in BL21(DE3) cells; expression was induced by the addition of 0.4 mM isopropyl-β-D-thio-galactoside (IPTG) when

Abbreviations: eIF4A, eukaryotic initiation factor 4A; HCV, hepatitis C virus; SF-1, helicase superfamily 1; SF-2, helicase superfamily 2.

Data Deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.rcsb.org (PDB ID codes 1FUU and 1FUK for full-length eIF4a and its carboxyl-terminal domain, respectively).

*To whom reprint requests should be addressed. E-mail: Dave.McKay@Stanford.edu.

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

Table 1. Crystallographic data collection and multiwavelength phasing statistics

Wavelength, Å	Observations (total/unique)	Completeness	R_{sym}*	$\langle I\rangle/\langle\sigma(I)\rangle$	Phasing power	f′	f″
λ1 = 1.77120	13729/5570	0.950 (0.945)	0.040 (0.056)	15.15 (15.02)	2.48	−12.1	10.9
λ2 = 1.84616	14174/5667	0.971 (0.980)	0.041 (0.062)	15.34 (13.93)	3.04	−36.1	15.2
λ3 = 1.84533	14176/5666	0.967 (0.954)	0.047 (0.069)	15.32 (14.03)	3.77	−21.5	35.8

	Resolution (Å)								
	5.96	4.74	4.14	3.77	3.5	3.29	3.13	2.99	Total
Reflections	641	700	672	701	708	741	716	659	5538
Figure of merit	0.810	0.833	0.775	0.798	0.794	0.776	0.795	0.764	0.799

Sm^{3+} derivative of the carboxyl-terminal domain of yeast eIF4A. Resolution range (last shell), 26.6–3.0 (3.11–3.00) Å.
Space group: $P2_12_12_1$; unit cell (Å) a = 34.7, b = 52.1, c = 82.3. Values of f′ and f″ were initially estimated from an EXAFS scan and refined in CNS.
*$R_{sym} = \Sigma|I_{hkl} - \langle I_{hkl}\rangle|/\Sigma\langle I_{hkl}\rangle$, where I_{hkl} = single value of measured intensity of hkl reflection, and $\langle I\rangle$ = mean of all measured value intensity of hkl reflection.

cells reached an OD_{600} of 0.5 and was allowed to continue for 5 h (25). Cells then were harvested by centrifugation, resuspended in 100 mM KCl/25 mM Bis-Tris, pH 7.4 (buffer A), and lysed by sonication. Lysate was clarified by centrifugation, and crude supernatant was applied to a Q-Sepharose column preequilibrated with buffer A. The column was eluted with a gradient of 0.1–0.6 M KCl. Fractions containing eIF4A carboxyl-terminal domain were combined and purified further by gel-filtration chromatography on a Superdex-75 column (Pharmacia). Eluant containing carboxyl-terminal protein was combined and concentrated to ≈22 mg/ml for crystallization.

Crystallization and Data Collection. Initial crystallization trials with Hampton Screens I and II in the absence of reducing agents proved unsuccessful. Analysis by PAGE under native conditions revealed the presence of multimers of the protein; formation of multimers was suppressed by the addition of 50 mM 2-mercaptoethanol (BME). In the presence of reducing agent, native crystals grew as elongated rods from 20% polyethylene glycol monomethyl ether of average molecular weight of 550 (PEG-MME_{550}), 1 mM $ZnSO_4$, and 100 mM Mes, pH 6.5; the presence of Zn^{2+}, or a similar divalent metal ion, was mandatory for crystallization. For a heavy atom derivative, 2 mM $SmCl_3$ was used instead of $ZnSO_4$ in the crystallization.

For data collection, the concentration of PEG-MME_{550} in the mother liquor was increased to 40% for cryoprotection and crystals were flash-cooled in a stream of nitrogen gas at ≈100 K. Native crystals are orthorhombic, space group $P2_12_12_1$, a = 34.7 Å, b = 52.1 Å, c = 82.3 Å, with one molecule per asymmetric unit.

Native diffraction data were collected to a resolution limit of 1.75 Å on beamline 9–2 of the Stanford Synchrotron Radiation Laboratory (SSRL) at a wavelength of 0.980 Å. Multiwavelength anomalous dispersion (MAD) data for a Sm^{3+} derivative were collected to a resolution limit of 3.0 Å at the Lawrence Berkeley Laboratory Advanced Light Source (ALS) synchrotron at wavelengths 1.84533 Å, 1.84616 Å, and 1.77120 Å for the Sm^{3+} absorption peak, dispersive edge, and remote energies, respectively. In both cases, data were recorded on Quantum IV charge-coupled device detectors and processed with the programs HKLVIEW, DENZO, and SCALEPACK (26).

Structure Determination of the Carboxyl-Terminal Domain of eIF4A. The structure of the carboxyl-terminal domain of eIF4A was solved by using experimental phases from the Sm^{3+} derivative. The location of a single Sm^{3+} ion was derived from anomalous difference Patterson peaks. MAD phases were computed by using the program package CNS (27); the overall figure of merit to 3.0 Å resolution was 0.80; phasing statistics are summarized in Table 1. After solvent flattening, much of the polypeptide chain could be traced in the experimental electron density map. Subsequent rounds of model refinement against the native data, phase combination between model and experimental phases, and phase extension with model phases yielded a model that includes residues 233–351 and 357–394, 122 water molecules, and 1 Zn^{2+} ion. The loop from Arg-352 through Lys-356, as well as the first three amino-terminal residues, apparently is disordered and could not be traced. Residues 321–327 and 349–351 near the Zn^{2+} site have relatively high average B factors, suggesting that zinc binding may distort the structure locally. Only one residue (Glu-241) is in the disallowed region of the Ramachandran plot; well defined electron density and a distinct hydrogen bond network support the unorthodox conformation of this residue. Refinement statistics are summarized in Table 2.

Structure Determination of Full-Length eIF4A. The structure of one protomer of the full-length eIF4A molecule was completed by using molecular replacement with the model of the carboxyl-terminal domain and data from triclinic crystals described previously (23). These crystals have two eIF4A molecules per asymmetric unit, related by an approximate noncrystallographic 2-fold screw axis. Earlier work described the complete structures of both amino-terminal domains and, additionally, a partial polypeptide backbone model for one carboxyl-terminal domain; however, the carboxyl-

Table 2. Crystallographic data collection and refinement statistics

	Carboxyl-terminal domain of eIF4A*	Full-length eIF4A†
Wavelength, Å	0.980	0.980
Resolution range (last shell), Å	30–1.75 (1.81–1.75)	40–2.8 (2.9–2.8)
Observations (total/unique)	41481/15307	53337/17192
Completeness, %	96.5 (93.2)	92.0 (80.4)
Completeness, $I>3\sigma$	87.3 (64.4)	NA
R_{sym}‡	0.029 (0.151)	0.071 (0.097)
$\langle I\rangle/\langle\sigma(I)\rangle$	17.4	18.3
R_{cryst}§	0.221	0.244
R_{free}	0.252	0.273
rmsd bond length, Å	0.010	0.010
rmsd angles, °	1.51	1.69
No. of residues modeled	157	593
No. of waters modeled	122	146

*Space Group: $P2_12_12_1$, unit cell (Å) a = 34.7, b = 52.1, c = 82.3 (Å). One protomer per asymmetric unit.
†Space group P1, unit cell a = 38.8, b = 71.3, c = 73.2 (Å), α = 94.0°, β = 89.6°, γ = 101.0°. Two protomers per asymmetric unit.
‡R_{sym} is defined in Table 1.
§$R_{cryst} = \Sigma|F_{obs} - F_{calc}|/\Sigma F_{obs}$, where F_{obs} = observed structure factor amplitude and F_{calc} = structure factor calculated from model. R_{free} is computed in the same manner as R_{cryst}, using a test set composed of 15% (carboxyl-terminal domain) or 10% (full-length eIF4A) of the observed reflections, which were randomly selected and omitted from all model-phase calculations.

BIOPHYSICS

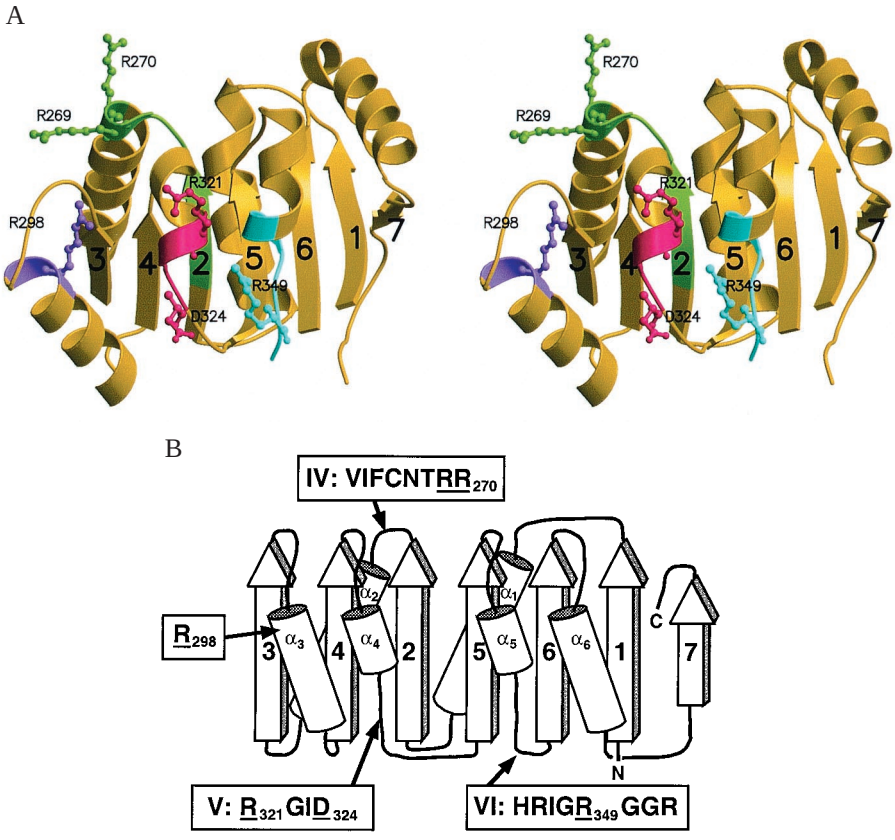

Fig. 1. Structure of the carboxyl-terminal domain of eIF4A. (*A*) Stereoview, ribbon drawing of the structure. Conserved motifs are colored as follows: motif IV, VIFCNTRR, residues 263–270, green; ''conserved R'' motif, residue Arg-298, purple; motif V, RGID, residues 321–324, magenta; motif VI, HRIGRGGR, residues 345–352, cyan. The strands of the β-sheet are labeled sequentially. This and subsequent ribbon drawings were prepared with MOLSCRIPT (34) and rendered with RASTER3D (35). (*B*) Topology diagram of the structure. β-Strands are shown as arrows; α-helices, as cylinders. β-Strands and α-helices are labeled sequentially as 1–7 and α_1–α_6, respectively. Sequences of the conserved motifs are shown in boxes; residues whose side chains are illustrated in *A* and subsequent figures are underlined.

terminal domains are significantly disordered in this crystal form, defeating a *de novo* determination of their structure. One copy of the refined molecular model of the carboxyl-terminal domain of eIF4A was placed initially by using the partial backbone trace as a guide; subsequent rigid body refinement, model building, and model refinement resulted in a model for this copy of eIF4A that includes residues 11–350 and 356–394. Placing the carboxyl-terminal domain for the second eIF4A protomer in these crystals yielded a molecular replacement solution that improved the overall refinement statistics; however, the resulting $2F_o - F_c$ and $F_o - F_c$ electron density maps in this region could not be interpreted reliably, and the average B factor for this domain was >100 Å^2, confirming that this domain is substantially more disordered than its counterpart in the first protomer. Hence, the carboxyl-terminal domain of this protomer was deemed too uncertain to be included in the final model. Refinement statistics for the crystals of full-length eIF4A are summarized in Table 2.

Results and Discussion

We first solved the crystallographic structure of the carboxyl-terminal domain of eIF4A with high-resolution data. The resulting model was used in molecular replacement to complete the structure of full-length eIF4A at lower resolution. Finally, we have superimposed the model of the carboxyl-terminal domain on the helicase ''template'' that is provided by structures of the Rep, PcrA, and HCV helicases to develop a working hypothesis for the involvement of specific amino acid residues in the biochemical activities of eIF4A.

Structure of the Carboxyl-Terminal Domain of eIF4A. The crystallographic structure of the carboxyl-terminal domain of eIF4A has been solved to a resolution of 1.75 Å. The average B factor for the structure is 35.7 Å^2. The anticipated coordinate error, estimated by the method of Luzzati (28), is 0.28 Å. A single Zn^{2+} ion is coordinated by the side chains of residues Glu-248 and Asp-252 of one molecule and Glu-341 and His-345 of a neighboring molecule.

The carboxyl-terminal domain has a parallel α-β structure (Fig. 1) with the same topology as the equivalent domain of other helicases [domain 2A of the Rep (10) and PcrA (11) DNA helicases; domain 2 of the HDV RNA helicase (17)]. The α-carbon backbone of this domain of eIF4A superimposes on the backbones of the Rep, PcrA, HCV, and UvrB helicases with root mean square differences of 1.8 Å, 1.9 Å, 1.8 Å, and 2.0 Å for 86, 84, 85, and 112 equivalent C_α atoms, respectively. The conserved topological ''core'' shared by these helicases is composed of the six parallel β-strands plus helices 1, 2, 4, and 5. Helix 3, which follows β-strand 3 of eIF4A carboxyl-terminal domain, is not found in DNA helicases of SF-1, for which the polypeptide emigrates to another structural domain (domain 2B) after β-strand 3, and reenters the domain at β-strand 4. α-Helix 6 is not present in HCV. Additionally, eIF4A has a short segment of $\approx$16 residues after helix 6, which adds a seventh β-strand and a well defined coil at the carboxyl terminus.

The conserved helicase sequence motifs IV, V, and VI, as originally described by Gorbalenya and Koonin (1), are highlighted in Fig. 1. Their positions in the eIF4A domain are topologically equivalent to the positions of these motifs in other helicases (note that motif IV would be equivalent to the recently described motif IVa of DNA helicases, as well as to motif IV of HCV helicase, as discussed in ref. 21). In addition, we observe an arginine residue, Arg-298 in helix 3, whose side chain is well ordered in the structure and whose spatial position suggests an involvement in RNA binding. Inspection of sequence alignments of DEA(H/D) RNA helicases [e.g., Pfam version 5.4, family 00271 (2)] reveal that this arginine is strictly conserved and, further, that it occurs as part of a QXXR motif, where the X amino acids often are charged. Notably, it is not present in the DNA helicases of SF-1, which lack this helix (and, hence, would not have been identified as a motif in sequence

alignments looking for correlations across all helicases). The suggestion that Arg-298 may participate in RNA binding is discussed in more detail below.

After this structure was completed, we searched for similar structures in the Protein Data Bank through the Dali server (www2.ebi.ac.uk/dali). The search identified a domain of the UvrB helicase as the most closely related structure. In addition, we submitted the amino acid sequence to several "fold recognition" servers. The PDB-BLAST method (bioinformatics.burnham-inst.org/pdb_blast) identified UvrB as the closest homolog for which a structure is known; this method correctly aligned eIF4A and UvrB over a stretch of 80 residues for which the structures superimpose without insertions or deletions; however, the method failed to accurately identify insertions and deletions that would have correctly superimposed shorter segments of the polypeptide backbone.

Structure of Full-length eIF4A. In previous work on crystals of full-length eIF4A protein (23), the amino-terminal domains of the two crystallographically independent protomers in the asymmetric unit were modeled successfully, but it proved impossible to complete a polypeptide trace of the carboxyl-terminal domains because of the high noise level of the electron density maps in those regions, indicative of significant disorder. In this work, we have used molecular replacement with the refined, high-resolution model of the carboxyl-terminal domain to complete the structure of the more well ordered of the two full-length molecules in these crystals. The polypeptide linker between the domains was traced, and minor adjustments were made within the carboxyl-terminal domain by interpreting $2F_o - F_c$ and $F_o - F_c$ maps during the course of refinement. The average B factor for this domain is 78.6 Å^2, as compared with 29.9 Å^2 for the amino-terminal domains, illustrating the substantial disorder. Segments of polypeptide whose continuity could be ascertained reliably, but in which conformational details were clouded by noise, were residues 256–259, 284–285, 317–323, and 365–368. The resulting model of full-length eIF4A protein reliably shows the overall tertiary structure of the molecule, despite the relatively low precision in prescribing some of the conformational details of the carboxyl-terminal domain.

In this crystal form, eIF4A has a "dumbbell" structure in which the amino- and carboxyl-terminal domains are connected by an extended, 11-residue linker (Fig. 2). The end-to-end length of the molecule is ≈80 Å; the linker is ≈18 Å long. The conserved helicase motifs on the amino- and carboxyl-terminal domains have no spatial relationship that suggest a mechanism of coupling ATP binding/hydrolysis with RNA binding/release in this structure. This structure and the intrinsic disorder of the carboxyl-terminal domain within it suggest that, in solution, eIF4A can be a distended molecule and that the linker between domains is relatively flexible (although it does not have an unusually high fraction of glycine, alanine, or serine, which is characteristic of many flexible polypeptides). (Implicitly, we assume that crystallization has selected and stabilized, through packing contacts, one representative conformation from the many that the molecule can access in solution.)

Modeling a Compact RNA-Binding Structure on a Helicase Template. The extended dumbbell structure of eIF4A is consistent with some published data on eIF4A activity, but inconsistent with other data. Under some conditions, the size of the single-stranded RNA-binding site is approximately 15 nt per eIF4A monomer (29), which is compatible with the extended structure. However, mutations in motif VI of the carboxyl-terminal domain (cyan in Fig. 2) reduce the hydrolysis rate for ATP bound to the amino-terminal domain (specifically, bound to motif I, blue in Fig. 2) (30); the extended structure is incompatible with this observation. To account for this observation, it is necessary to suggest that when ATP binds eIF4A, the protein undergoes a conformational change that results in a compact structure in which the two domains interact directly with each other. Changes in proteolytic digestion patterns have demonstrated that eIF4A undergoes a series of nucleotide- and RNA-dependent conformational changes (31, 32), which are consistent with this suggestion, although the proteolysis patterns do not, by themselves, demonstrate juxtaposition of the two domains.

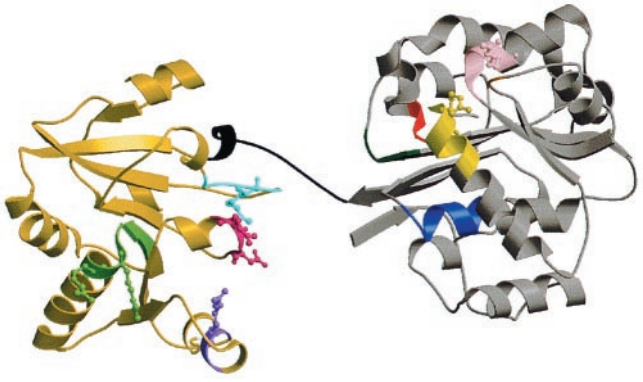

Fig. 2. Ribbon drawing of the structure of full-length eIF4A. The amino- and carboxyl-terminal domains are colored silver and gold, respectively; the 11-residue linker connecting them is colored black. The conserved amino-terminal motifs are colored as follows: motif I, Walker A motif ASQSGTGKT, residues 65–72, blue; motif Ia, PTRELA, residues 97–102, yellow; GG, residues 125–126, orange; TPGR, residues 145–148, pink; motif II, Walker B motif DEAD, residues 169–172, red; motif III, SAT, residues 200–202, green. The conserved carboxyl-terminal motifs are colored as described in Fig. 1*A*.

The two domains of eIF4A share a structural similarity with their counterparts in other helicases (this work and refs. 22 and 23), and the conserved helicase motifs occur at similar positions in the structures. It is likely that the similarity in domain structure is mandated by similarity in the mechanisms of interaction with ATP and oligonucleotides. In this context, we suggest that eIF4A forms a compact structure similar to that of other helicases at some point in its ATP-dependent helicase cycle. Although this suggestion may seem inconsistent with the extended structure we observe in our crystals, it should be borne in mind that the eIF4A molecule by itself has an intrinsically weak helicase activity (9), which can be enhanced substantially by the collaboration of eIF4B (8) or eIF4H (33). By implication, the collaborating proteins may facilitate formation and stabilization of a compact eIF4A structure. We may then ask hypothetically, "What would such a model suggest about interactions of eIF4A with RNA and ATP, and are there data that would justify the model?"

We have superimposed both the amino- and carboxyl-terminal domains of eIF4A onto their counterparts in the crystallographic structures of PcrA (PDB id 3PJR), Rep (PDB id 1UAA), and HCV (PDB id 1A1V) helicases complexed with oligonucleotides, as well as UvrB (PDB id 1D9X) without oligonucleotide. Representative models are shown in Fig. 3, with the orientation of the amino-terminal domain of eIF4A approximately the same as in Fig. 2. In these models, the distance from eIF4A residue 223 to residue 233 is 18–20 Å, which can be accommodated readily by the linker polypeptide. Such modeling is necessarily imprecise, because (*i*) the positioning of the carboxyl-terminal domain relative to the amino-terminal domain varies by several Angstroms from one helicase template to another and (*ii*) for both PcrA and HCV helicases, several structures have been solved in different crystal forms and with different ligands (DNA, AMPPNP, ADP) bound to the proteins; comparison of these structures reveals substantial intramolecular conformational changes that are thought to be related to the helicase mechanism (12, 20). Nonetheless, the models share several features: (*i*) an extended single-stranded oligonucleotide-binding region across the "top" of domains 1 and 2 of other

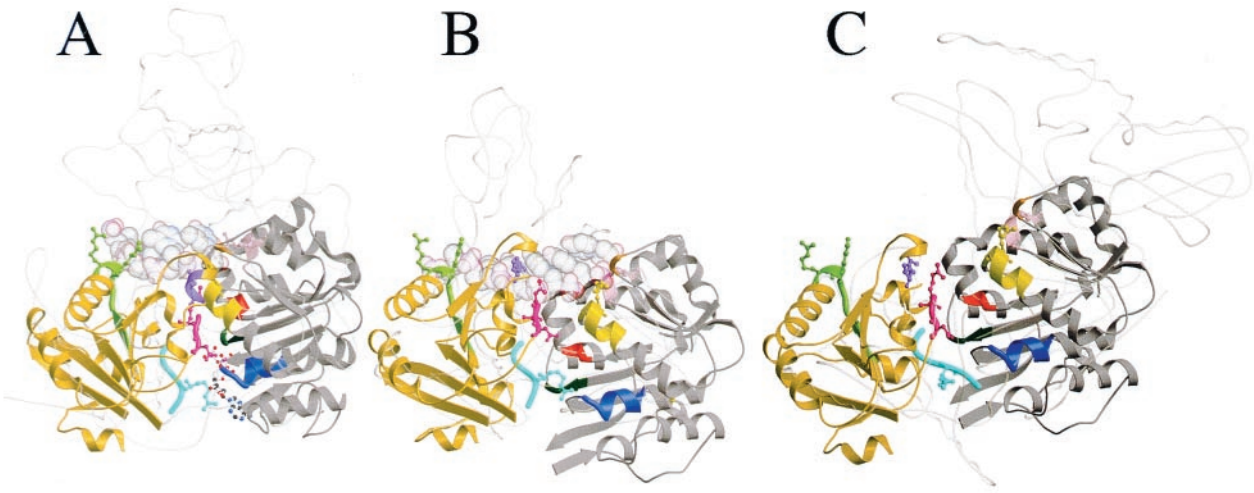

Fig. 3. Overlay of amino- and carboxyl-terminal domains of eIF4A onto the equivalent domains of other helicases. The amino-terminal domain of eIF4A is in approximately the same orientation as in Fig. 2. The color convention for the domains and conserved motifs of eIF4A is the same as in Fig. 2. In all cases, the ''target'' helicase is shown as a transparent tube drawing against the solid eIF4A model. Oligonucleotides are shown as CPK space-filling models. (*A*) PcrA DNA helicase (PDB ID code 3PJR) with ATP and the single-stranded portion of the DNA. ATP is shown as a ball-and-stick representation. (*B*) HCV RNA helicase with ssDNA (PDB ID code 1A1V). (*C*) UvrD DNA helicase (PDB ID code 1D9X).

helicases (in the orientation shown in Fig. 3) and (*ii*) juxtaposition of motifs I and II of domain 1 (the classical Walker A and B nucleotide-binding motifs, colored blue and red, respectively, in Fig. 3) with motifs V and VI of domain 2 (colored magenta and cyan, respectively). For eIF4A, these features translate into (*i*) a putative RNA-binding region with an extended network of arginines, viz., Arg-98 (of motif IA) and 148 (of the TPGR motif) of the amino-terminal domain, and Arg-269, -270, and -298 (plus, perhaps, 321) from the carboxyl-terminal domain, along the interface of the single-stranded oligonucleotides, and (*ii*) juxtaposition of side chains of several amino acid residues of motifs V and VI with the ATP-binding site and the DEAD motif of the amino-terminal

Table 3. Summary of helicase domain 2 motifs and interactions

Motif	Sequence		Interactions
IV	PcrA (355–365):	AVLY**R**T**N**AQS**R**	R359 binds DNA and forms salt bridge to E600; N361 interacts with ssDNA; R365 near dsDNA.
	Rep (346–356):	AILY**R**G**NH**QS**R**	R350, N352* H353*, R356* interact with ssDNA.
	HCV (365–372):	LIFCHS**K**K	K371 interacts with ssDNA.
	UvrB (444–451):	LVTVKTV***R***	R451 may interact with DNA.
	IF4A (263–270):	VIFCNT***RR***	R269 and/or R270 may interact with ssRNA.
V	PcrA (565–571):	**H**AA**K**GL**E**	H565 interacts with ssDNA; K568 forms salt bridge with E224 and D227 of $DEYQD_{227}$ motif†; E571 interacts with ribose of AMPPNP.
	Rep (558–564):	**H**AS**K**GLE	H558 interacts with DNA; K561 forms salt bridge with D218 of $DEYQD_{218}$ motif†.
	HCV (413–419):	ALMTGFT	No specific interactions for these residues.
	UvrB (501–505):	***R***EGL**D**	R501 may interact with DNA; D505 interacts with ATP.
	IF4A (321–324):	***R***GI***D***	R321 may interact with $DEAD_{172}$ motif† or with ssRNA; D324 may interact with ATP.
VI	PcrA (606–613):	VGIT**R**AEE	R610 interacts with γ phosphate of AMPPNP.
	Rep (598–605):	VGIT**R**AQK	R602 is near the P-loop and may interact with ATP.
	HCV (460–467):	**Q**RRG***R***TG***R***	Q460 interacts with H290 of $DECH_{290}$ motif; a conformational change could allow R464 and R467 to interact with phosphates of ATP.
	UvrB (531–538):	**Q**TIG**R**AA**R**	Q531 interacts with H341 of $DESH_{341}$ motif; R535 and R538 interact with phosphates of ATP.
	IF4A (345–352):	***H***RIG***R***GG***R***	H345 may interact with D172 of $DEAD_{172}$ motif; R349 and R352 could interact with phosphates of ATP.

Amino acid residues in PcrA, Rep, HCV and UvrB helicases showing specific interactions that are likely to be involved in biochemical activity are boldface underlined; residues for which one can propose hypothetical interactions in eIF4A and other helicases are shown in boldface italics. Motif sequences are aligned vertically. For UvrB, *Thermus thermophilus* sequence and numbering are used. ssDNA, single-stranded DNA; dsDNA, double-stranded DNA.

*In at least one of the two independent copies of Rep in the crystallographic asymmetric unit.

†The classical DExx (Walker ''B'') motif is four residues in length; interactions are seen with the residue following the DExx motif.

domain, suggesting a role in coupling interactions with ATP or ADP to conformational changes in the protein. Representative specific interactions observed in other helicase structures, and the extrapolated interactions that they suggest for the carboxyl-terminal motifs of eIF4A, are summarized in Table 3. [The probable role of the motifs in the amino-terminal domain if eIF4A has been discussed previously (22, 23) and will not be repeated here.]

Motif IV participates in oligonucleotide binding in all of the helicase–DNA complexes. Arg-269 and Arg-270 of eIF4A (colored green in Figs. 1–3) may interact with RNA.

The side chain of Arg-298 (colored purple in Figs. 1–3) is well ordered in the eIF4A structure, and sequence alignments suggest that this residue belongs to a "QXXR" motif that is specific to the DEA(D/H)-box helicases. In the models shown in Fig. 3, it would interact with oligonucleotides.

In the DNA helicases, residues from motif V "bridge" the ATP-binding site and the oligonucleotide-binding site. A histidine at the beginning of the motif (see Table 3) interacts with DNA; a lysine in the middle interacts with the DEYQD motif (the equivalent of the DEAD motif of eIF4A, plus the aspartic acid that follows it), and in PcrA with ATP bound, the carboxyl of Glu-571 interacts with the ribose of ATP.

In motif V of the DEA(D/H)-box helicases, only the four-residue RGID is conserved. Arg-321 of eIF4A could interact with either the DEAD motif or with the oligonucleotide. Asp-324 may interact with ATP similar to Glu-571 of PcrA. This motif therefore could play a role in coupling ATP binding and hydrolysis to RNA binding and/or a conformational change.

In mouse eIF4A, mutation of any of the three arginines in motif VI to glutamine reduces the ATPase activity to less than 20% of its wild-type value, whereas mutation of either the second or third arginine to lysine yielded a protein that retained >60% of its wild-type activity. Each of these mutations abolishes helicase activity in an assay employing both eIF4A and eIF4B. These data support an involvement of residues in this motif in both ATP hydrolysis and coupling of ATP binding or hydrolysis to RNA binding and helicase activity (30).

In motif VI of PcrA, the Arg-610 binds the γ-phosphate of ATP. In UvrB, Arg-535 and Arg-538 both interact with the phosphates of ATP. These interactions suggest that in eIF4A, Arg-349 and/or Arg-352 may bind the phosphates of ATP, although modeling the interactions is hampered by the fact that we have been unable to trace residues 352–356 in the structures.

His-345 of eIF4A is in a position to form a salt bridge to Asp-172, the final aspartic acid of the DEAD motif, resulting in a histidine–aspartate interaction. In both the HCV and UvrB helicases, we see a similar interaction, but with a "switch" of residues; to wit, H → Q in motif VI and D → H in the DExx motif, resulting in a glutamine–histidine interaction. It is interesting to note that in the sequences of most DEAH-box helicases, a glutamine is conserved at the sequence position of His-345 in eIF4A, which also would result in the same glutamine–histidine interaction; this "bridge" between motif VI and the DExx motif is found in many helicases of SF-2 in one or the other of these two alternative versions.

Although the interactions we propose are tentative, they reconcile results of previous mutagenesis studies with the structural information and provide a working model for examining the role of residues whose functional significance heretofore has not been tested.

Conclusions

The high-resolution structure of the carboxyl-terminal domain of eIF4A reveals the structural similarity of this domain in DEA(D/H)-box helicases to the equivalent domain in other helicases. In our crystals, full-length eIF4A protein is a "dumbbell" molecule with two globular domains connected by an extended (and in all probability, flexible) linker. Modeling a hypothetical compact, oligonucleotide-binding structure for eIF4A on the templates provided by other helicase structures gives us a testable, working model for intermolecular interactions with RNA and for intramolecular interactions that may couple the ATPase cycle to conformational changes and helicase activity in eIF4A.

We thank Ram Samudrala for assistance with fold recognition programs and beamline staff members of SSRL and ALS for assistance with synchrotron data collection. This work was supported by award MCB-9874582 from the National Science Foundation to D.B.M. and a Stanford Dean's Postdoctoral Fellowship to J.M.C. This work is based on research conducted at SSRL, which is funded by the Department of Energy (BES, BER) and National Institutes of Health (NCRR, National Institute of General Medical Sciences), and at Advanced Light Source, which is funded by the Department of Energy.

1. Gorbalenya, A. E. & Koonin, E. V. (1993) *Curr. Opin. Struct. Biol.* **3,** 419–429.
2. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. & Sonnhammer, E. L. (2000) *Nucleic Acids Res.* **28,** 263–266.
3. Staley, J. P. & Guthrie, C. (1998) *Cell* **92,** 315–326.
4. Venema, J. & Tollervey, D. (1995) *Yeast* **11,** 1629–1650.
5. Py, B., Higgins, C. F., Krisch, H. M. & Carpousis, A. J. (1996) *Nature (London)* **381,** 169–172.
6. Pain, V. M. (1996) *Eur. J. Biochem.* **236,** 747–771.
7. Linder, P., Gasteiger, E. & Bairoch, A. (2000) *Yeast* **16,** 507–509.
8. Jaramillo, M., Dever, T. E., Merrick, W. C. & Sonenberg, N. (1991) *Mol. Cell. Biol.* **11,** 5992–5997.
9. Rogers, G. W., Jr., Richter, N. J. & Merrick, W. C. (1999) *J. Biol. Chem.* **274,** 12236–12244.
10. Korolev, S., Hsieh, J., Gauss, G. H., Lohman, T. M. & Waksman, G. (1997) *Cell* **90,** 635–647.
11. Subramanya, H. S., Bird, L. E., Brannigan, J. A. & Wigley, D. B. (1996) *Nature (London)* **384,** 379–383.
12. Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S. & Wigley, D. B. (1999) *Cell* **97,** 75–84.
13. Soultanas, P., Dillingham, M. S., Velankar, S. S. & Wigley, D. B. (1999) *J. Mol. Biol.* **290,** 137–148.
14. Machius, M., Henry, L., Palnitkar, M. & Deisenhofer, J. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 11717–11722.
15. Nakagawa, N., Sugahara, M., Masui, R., Kato, R., Fukuyama, K. & Kuramitsu, S. (1999) *J. Biochem. (Tokyo)* **126,** 986–990.
16. Theis, K., Chen, P. J., Skorvaga, M., Van Houten, B. & Kisker, C. (1999) *EMBO J.* **18,** 6899–6907.
17. Yao, N., Hesson, T., Cable, M., Hong, Z., Kwong, A. D., Le, H. V. & Weber, P. C. (1997) *Nat. Struct. Biol.* **4,** 463–467.
18. 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.
19. 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* **6,** 89–100.
20. Yao, N., Reichert, P., Taremi, S. S., Prosise, W. W. & Weber, P. C. (1999) *Struct. Fold. Des.* **7,** 1353–1363.
21. Korolev, S., Yao, N., Lohman, T. M., Weber, P. C. & Waksman, G. (1998) *Protein Sci.* **7,** 605–610.
22. Benz, J., Trachsel, H. & Baumann, U. (1999) *Structure* **7,** 671–679.
23. Johnson, E. R. & McKay, D. B. (1999) *RNA* **5,** 1526–1534.
24. Story, R. M., Weber, I. T. & Steitz, T. A. (1992) *Nature (London)* **355,** 318–325.
25. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) *Methods Enzymol.* **185,** 60–89.
26. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276,** 307–326.
27. 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.
28. Luzzati, P. V. (1952) *Acta Crystallogr.* **5,** 802–810.
29. Peck, M. L. & Herschlag, D. (1999) *RNA* **5,** 1210–2121.
30. Pause, A., Methot, N. & Sonenberg, N. (1993) *Mol. Cell. Biol.* **13,** 6789–6798.
31. Lorsch, J. R. & Herschlag, D. (1998) *Biochemistry* **37,** 2194–2206.
32. Lorsch, J. R. & Herschlag, D. (1998) *Biochemistry* **37,** 2180–2193.
33. Richter-Cook, N. J., Dever, T. E., Hensold, J. O. & Merrick, W. C. (1998) *J. Biol. Chem.* **273,** 7579–7587.
34. Kraulis, P. (1991) *J. Appl. Crystallogr.* **24,** 946–950.
35. Merritt, E. A. & Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524.