RNA

Crystallographic structure of the amino terminal domain of yeast initiation factor 4A, a representative DEAD-box RNA helicase

E. R. Johnson and D. B. McKay

RNA 1999 5: 1526-1534

References

Article cited in:

<http://www.rnajournal.org/cgi/content/abstract/5/12/1526#otherarticles>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

Notes

To subscribe to *RNA* go to:
<http://www.rnajournal.org/subscriptions/>

REPORT

Crystallographic structure of the amino terminal domain of yeast initiation factor 4A, a representative DEAD-box RNA helicase

ERIC R. JOHNSON and DAVID B. MCKAY

Beckman Laboratories for Structural Biology, Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305-5400, USA

ABSTRACT

The eukaryotic translation initiation factor 4A (eIF4A) is a representative of the DEAD-box RNA helicase protein family. We have solved the crystallographic structure of the amino-terminal domain (residues 1–223) of yeast eIF4A. The domain is built around a core scaffold, a parallel α – β motif with five β strands, that is found in other RNA and DNA helicases, as well as in the RecA protein. The amino acid sequence motifs that are conserved within the helicase family are localized to the β strand $\rightarrow$ α helix junctions within the core. The core of the amino terminal domain of eIF4A is amplified with additional structural elements that differ from those of other helicases. The phosphate binding loop (the Walker A motif) is in an unusual closed conformation. The crystallographic structure reveals specific interactions between amino acid residues of the phosphate binding loop, the DEAD motif, and the SAT motif, whose alteration is known to impair coupling between the ATPase cycle and the RNA unwinding activity of eIF4A.

Keywords: crystal structure; DEAD box; eIF4A; RNA helicase

INTRODUCTION

Eukaryotic initiation factor 4A (eIF4A) participates in the initiation of polypeptide translation from mRNA, presumably by exerting an RNA duplex unwinding activity (for review, see Pain, 1996). In contemporary models, the initial step in mRNA recognition is binding to the 7-methylguanosine cap, which is effected by the cap-binding protein eIF4E as part of a larger complex (often called eIF4F) that includes eIF4G and eIF4A. Then, working in conjunction with eIF4B, eIF4A is thought to “melt” the duplex secondary structure in 5′ untranslated regions of mRNA in vivo. This model is supported circumstantially by the demonstration that in vitro, eIF4A has an ATP-dependent RNA unwinding activity. In early work, it was observed that this unwinding activity required eIF4B or other accessory proteins (Jaramillo

et al., 1991; Methot et al., 1994); more recently, it has been shown that eIF4A working alone can dissociate RNA duplexes of sufficiently low thermodynamic stability (Rogers et al., 1999).

eIF4A is a member of an extended protein family referred to as the “DEX(D/H)-box” RNA helicases. This family of proteins has been defined at the level of amino acid sequence by the presence of a conserved, ~400-residue domain that carries the characteristic sequence fingerprints of helicases (Gorbalenya & Koonin, 1993). This family encompasses a spectrum of proteins involved in such diverse activities as eukaryotic RNA splicing (Staley & Guthrie, 1998), ribosome biogenesis (Venema & Tollervey, 1995), and RNA degradation (Py et al., 1996), to name a few; additionally, for many of the DEX(D/H)-box helicases identified at the level of sequence, the biological function remains obscure. The DEX(D/H)-box helicases belong to helicase superfamily 2 (SF2) in the classification of Gorbalenya and Koonin (1993), and despite sharing characteristic sequence fingerprints with other helicases, standard comparison searches (e.g., BLAST 2.0) reveal no obvious global sequence alignment with either the DNA helicases of superfamily 1 (SF1) or the viral RNA helicases that are also included in SF2.

Reprint requests to: David McKay, Department of Structural Biology, Sherman Fairchild Building, Stanford University School of Medicine, Stanford, California 94305-5126, USA; e-mail: Dave.McKay@Stanford.edu.

Abbreviations: eIF4A: eukaryotic initiation factor 4A, and similarly for other initiation factors; SeMet: seleno-L-methionine; CHES: 2-[N-cyclohexylamino]ethanesulfonic acid; PEG-MME₅₀₀₀: polyethylene glycol monomethyl ether of average molecular weight 5,000; Rmsd: root mean square deviation.

Crystallographic structures have been reported for two DNA helicases of the SF1 family. The structure of the PcrA helicase of *Bacillus stearothermophilus* (Subramanya et al., 1996) and complexes of PcrA with DNA that represent substrate and product states (Velankar et al., 1999) have been solved. Also, the structure of the Rep helicase of *Escherichia coli* complexed with single-strand DNA has been presented (Korolev et al., 1997). Structures for a representative of the SF2 family of helicase, the RNA helicase of hepatitis C virus (HCV), has also been reported, both in the absence of oligonucleotide (Yao et al., 1997; Cho et al., 1998) and complexed with single-strand DNA (Kim et al., 1998).

In this context, we have crystallized and pursued the structure determination of recombinant yeast eIF4A, a representative DEX(D/H)-box helicase. In crystals of full-length protein, we find that the amino terminal domain of the molecule is well-ordered, whereas the remaining carboxy-terminal region is not. We have, consequently, also expressed the crystallized amino terminal domain alone and confirmed its structure at higher resolution to avoid the ambiguities introduced by a large fraction of the molecule being disordered in the crystals.

While this work was being prepared for publication, the work of Benz et al. (1999) on the structure of a similar fragment of yeast eIF4A was published. The overall tertiary structures of the domain solved by Benz and colleagues and the amino-terminal domain reported here are essentially the same. There are differences in the amino acid residue boundaries of the recombinant proteins: this work uses residues 1–223, while Benz and colleagues use residues 9–232 plus an amino-terminal histidine tag. The domain used by Benz and colleagues was defined by deletion constructs, whereas the one reported here was determined from a partial model of the full-length protein. A region of the molecule that is disordered in the Benz structure (residues 125–135, which spans a Gly-Gly motif conserved throughout this helicase family) can be traced in this

work, as its conformation is stabilized by crystal packing. Finally, in all of the Benz structures, the P-loop of the Walker A motif is in the open conformation observed in most P-loop proteins, whereas in the structure reported here, it is in an unusual closed conformation.

RESULTS AND DISCUSSION

Crystallographic structure determination

Crystals of full-length yeast eIF4A grow as thin plates, the diffraction from which is anisotropically mosaic. Using a single crystal in which seleno-*L*-methionine (SeMet) was substituted for methionine, we collected diffraction data at three wavelengths, located 18 of a possible 22 selenium atom sites for the two molecules in the triclinic unit cell, and calculated phases using the multiwavelength anomalous dispersion methodology (Table 1; see Materials and Methods for details). This produced an electron density map in which a significant fraction of the polypeptide backbone could be traced (Fig. 1A). The selenium sites facilitated sequence placement by unambiguously identifying nine methionines in each molecule, all within the first 220 amino acids. A model that eventually included residues 11–223¹ of both protomers in the unit cell was built and refined. Notably, within plus or minus a residue, these are the same boundaries of the well-ordered part of the amino-terminal domain in the work of Benz et al. (1999). It became apparent that the readily interpretable part of the electron density map encompassed only one structural domain of each eIF4A molecule. The amino terminal domains pack in continuous sheets in the crystal; between these layers are regions of weak electron density whose interpretation has been problematical.

¹The sequence numbering corresponds to entry P10087 of the SwissProt Database. Because the SwissProt entry does not have an amino-terminal methionine, the numbering differs by –1 from numbering in some of the literature.

TABLE 1. Crystallographic data collection and phasing statistics for SeMet-substituted full-length yeast eIF4 (40–2.8 Å).

Wavelength	Observations (total/unique)	Completeness	Rmerge ^a	$\langle I \rangle / \langle \sigma I \rangle$	Phasing power (iso/ano)	f'	f''		
$\lambda 1 = 0.9250 \text{ \AA}$	47,206/16,526	0.887 (0.734)	0.079 (0.115)	16.2 (6.8)	3.86/2.32	−2.80	3.64		
$\lambda 2 = 0.9793 \text{ \AA}$	50,454/16,722	0.913 (0.786)	0.085 (0.109)	17.8 (10.0)	1.98/2.69	−7.06	4.45		
$\lambda 3 = 0.9799 \text{ \AA}$	53,377/17,192	0.920 (0.804)	0.071 (0.097)	18.3 (10.3)	—/2.06	−8.43	3.24		
Resolution (Å)									
	7.8	5.5	4.5	3.9	3.5	3.2	3.0	2.8	Total
Reflections	806	1,485	1,957	2,241	2,523	2,692	2,527	2,229	16,460
Figure of merit	0.786	0.823	0.734	0.681	0.652	0.585	0.535	0.497	0.638
Number of selenium sites used for phasing: 18									

Space group: P1. Unit cell (Å): $a = 38.8$, $b = 71.3$, $c = 73.2$, $\alpha = 94.0^\circ$, $\beta = 89.6^\circ$, $\gamma = 101.0^\circ$. The dataset collected at $\lambda 3$ was used as the reference dataset. The values for f' and f'' were determined from an EXAFS scan.

^aRmerge = $\sum |I_i - \langle I \rangle| / \sum \langle I \rangle$ where I = diffraction intensity and $\langle I \rangle$ = mean measured intensity.

FIGURE 1. Representative electron density. **A:** Experimental maps from crystals of full-length eIF4A. Cyan: native Fourier computed with coefficients $F_{obs} \exp(\alpha_{exp})$ to 2.8-Å resolution, contoured at 1.0σ using all reflections from the edge dataset ($\lambda = 0.9799$ Å), where α_{exp} are MAD phases determined as described in Materials and Methods; red: anomalous difference Fourier map computed with coefficients $(F_{obs}^+ - F_{obs}^-) \exp(\alpha_{exp} - 90^\circ)$ contoured at 8.0σ using all Bijvoet differences from the Se absorption peak dataset ($\lambda = 0.9793$ Å). The initial model prior to refinement is also shown. **B:** Map from crystals of amino-terminal domain (amino acids 1–223), calculated with coefficients $(2F_{obs} - F_{calc}) \exp(\alpha_{calc})$ to 2.5 Å, contoured at 1.2σ , where F_{calc} and α_{calc} are computed from refined model. The refined model is also shown. Oxygen atoms are red; nitrogen, blue; carbon, black; sulfur, green; selenium, magenta. Figures were prepared with Bobscript (Kraulis, 1991; Esnouf, 1997, 1999) and POV-RAY. **C:** Phase-combined maps from crystals of full-length eIF4A, contoured at 1.5σ . C_α trace of amino terminal domain is cyan; partial trace of second domain is yellow. The figure was made with O (Jones et al., 1991). **D:** Schematic drawing of molecular packing in crystals of full-length eIF4A, in same orientation as **C**. Ribbon diagram of amino terminal domains is yellow; partial trace of second domain is cyan; dot surface of molecular envelope is red. The figure was made with RasMol (Sayle & Milner-White, 1995).

An amino-terminal fragment constituted of residues 1–223 of eIF4a was then expressed, purified, and crystallized. Crystallographic data were collected at a single wavelength to 2.5 Å and the structure was solved by molecular replacement and refined (Table 2). The current model includes residues 11–223 and 55 solvent molecules. The amino-terminal ten residues are disordered in the crystal. The method of Luzzati (1952) gives an estimate of error in atomic positions of 0.31 Å.

The refined model for the amino terminal domain was subsequently placed in crystals of the full-length protein. Minor refinement and manual adjustment of the model, followed by phase combination with exper-

imental and model phases, yielded an electron density map in which much of the polypeptide backbone of a second domain could be traced for one of the two molecules in the triclinic unit cell (Fig. 1C–D). However, it has not yet been possible to unambiguously place the amino acid sequence in this segment of polypeptide. Despite this, the overall envelope of the second domain reveals that it is connected to the first by a distended polypeptide linker. The crystal packing is such that the second domains form a layer that shows significant disorder on the molecular level, correlating with the macroscopic disorder of the crystals of the full-length protein.

TABLE 2. Crystallographic data collection and refinement statistics for amino-terminal domain of eIF4A.

Wavelength (Å)	0.9999
Resolution range (Å)	40–2.5 (2.66–2.50)
Observations (total/unique)	29,694/8,026
Completeness	0.938 (0.727)
Completeness ($I > 3\sigma$)	0.859 (0.570)
R_{sym}^a	0.062 (0.153)
$\langle I \rangle / \langle \sigma(I) \rangle$	21.1 (4.8)
R_{cryst}^b	0.223 (0.267)
R_{free}^b	0.274 (0.317)
Rmsd bond length (Å)	0.007
Rmsd angles (°)	1.26
Number of residues modeled	213
Number of waters modeled	55

Space group: P2₁2₁2₁; unit cell (Å): $a = 38.8$, $b = 74.3$, $c = 81.0$.

^a $R_{sym} = \sum |I_i - \langle I \rangle| / \sum \langle I \rangle$ where I = diffraction intensity and $\langle I \rangle$ = mean measured intensity.

^b $R_{cryst} = \sum |F_{obs} - F_{calc}| / \sum F_{obs}$ where F_{obs} = observed structure factor amplitude in working set, F_{calc} = structure factor calculated from model, and R_{free} is the same using F_{obs} from the test set. A test set comprised of 10% of the observed reflections (830 reflections) was randomly selected and omitted from all model phase calculations.

Tertiary fold and its relationship to other helicase structures

The amino terminal domain of eIF4A has a parallel α - β topology with eight β strands (Fig. 2A). Seven of the β strands run parallel, whereas the first, short amino-

proximal strand at the edge of the β sheet runs anti-parallel to the others. The strand on the opposite, amino-distal edge of the β sheet leads into a distended segment of peptide that is relatively disordered, with average B-factors of $\sim 55 \text{ Å}^2$ as compared to 30.7 Å^2 for the molecule as a whole (Fig. 2B). The conformation of this loop is stabilized by the side chain of Phe129 packing with a neighboring molecule; in the crystal form reported by Benz et al. (1999), this region is too disordered to trace. This segment begins with a Gly125-Gly126 motif that is conserved in sequence throughout the DEX(D/H)-box helicase family. [Notably, the mutation G125D is lethal whereas a G126D mutation results in a severe growth defect in yeast (Schmid & Linder, 1991).] Sites of proteolytic sensitivity have been determined for the mouse eIF4A, which is 66% identical in sequence to the yeast protein (Lorsch & Herschlag, 1998). Major sites are located in the distended segment of peptide following the Gly-Gly motif, and also immediately after the structural boundary of the amino-terminal domain, consistent with the expectation that the sites of proteolysis will occur at domain boundaries or in relatively accessible peptide segments on the surface.

It was noted when the PcrA DNA helicase structure was solved (Subramanya et al., 1996) that the ATP binding domain had a structural core that was identical in topology to a subdomain of the RecA protein (Story

FIGURE 2. Schematic drawings of the structure of the amino-terminal domain of eIF-4A. **A:** The conserved helicase sequence motifs are highlighted in color: motif I, Walker A motif AQSGTGKT, amino acid residues 65–72, blue; motif Ia, PTRELA, residues 97–102, yellow; GG (unique to DEX(D/H)-box RNA helicases), residues 125–126, orange; TPGR [identified by comparison of HCV and Rep structures (Korolev et al., 1998)], residues 145–148, pink; motif II, Walker B, DEAD, residues 169–172, red; and motif III, SAT, residues 200–202, green. The amino and carboxy termini correspond to amino acid residues 11 and 223 respectively. **B:** Colored according to average B-factor of main chain atoms: Blue: 15 Å^2 ; white: 35 Å^2 ; red: 45 Å^2 ; orange: $>55 \text{ Å}^2$. Figures was prepared with BOBSCRIPT (Kraulis, 1991; Esnouf, 1997, 1999) and POVray version 3.4.

& Steitz, 1992). This scheme has generalized to the closely related Rep helicase (Korolev et al., 1997), the HCV RNA helicase (Yao et al., 1997; Cho et al., 1998; Kim et al., 1998), and most recently to eIF4A (Benz et al., 1999) (Fig. 3). The domains are elaborated around a structural core composed of five parallel β strands connected by α helices in a parallel α - β motif. The β strands within the core have the topological order 1-5-4-2-3, with the Walker A and B motifs that are required for ATP binding occurring in strands 1 and 4, respectively. This provides a scaffold for segments of polypeptide that are involved in ATP and oligonucleotide binding that cluster along the top of the domain and which have characteristic fingerprint sequences. In the classification of Gorbalenya and Koonin (1993), different families of helicases are distinguished by conservation of fingerprint sequences within a family and variations in sequence between families. Superposition of the RecA-like core of the eIF4A amino terminal

domain presented here on that of the HCV, Rep and PcrA helicases yields root mean square differences in C_{α} position of 1.76 Å, 1.77 Å, and 1.63 Å for 73, 88, and 103 equivalent atoms, respectively.

Conformation and interactions of the nucleotide binding motifs

The β -strand $\rightarrow$ loop $\rightarrow$ α helix (strand 1 $\rightarrow$ helix A in Fig. 3) that encompasses the Walker A sequence motif (consensus sequence (G/A)XXXXGK(T/S), often referred to as the P-loop motif) has been shown to bind the terminal phosphates of ATP/GTP or ADP/GDP in many systems. The loop is usually found in an open conformation that sterically complements the phosphates, with the ϵ -amino group of the lysine of the motif forming a salt bridge with the β phosphate of NDPs or the β and γ phosphates of NTPs. In the absence of nucleotide, the loop generally remains in the open conformation (for review, see Smith & Rayment, 1996). In the structure of the ATPase domain of eIF4A of Benz et al. (1999), the loop is also in the open conformation in the presence of ATP, ADP, and without nucleotide (but with a sulfate ion bound).

In our eIF4A amino terminal domain structure, the phosphate binding loop is in an unusual closed conformation that would sterically conflict with nucleotide binding, and as a consequence, displays different intramolecular interactions than those observed by Benz et al. (1999) (Fig. 4). This altered conformation is due largely to the entire loop swiveling about Gly70 and Ala65. Within the loop, the backbone carbonyl of Gly68 hydrogen bonds weakly (i.e., distance = 3.0 Å) with the amide nitrogen of Thr72, thereby presumably stabilizing the closed conformation. To accomplish this, Gly68 adopts Ramachandran angle values (ϕ, ψ) = (142°, 173°), which lie well within the forbidden region for nonglycine residues. Whether this is an obligatory interaction is uncertain; mutation of this glycine to isoleucine in mouse eIF4a did not preclude dATP binding at 3 mM concentration (Rozen et al., 1989), but more extensive analysis of the effects of mutating this residue on the ATPase cycle has not been reported. Another interaction that appears to stabilize the closed conformation of the loop is a hydrogen bond from the hydroxyl of Ser67 to the carboxyl of Glu170 of the DEAD-box motif. Additionally, the side chain of Lys71 adopts an alternative conformation to that which is observed when nucleotide is bound, such that the ϵ amino group is near the position that Mg^{2+} would bind if MgATP were present. The amino group of Lys71 makes a hydrogen bond to the hydroxyl of Ser67, and is in proximity to the carboxyl of Asp169 of the DEAD-box motif, suggesting a strong electrostatic interaction. It is notable that mutating the equivalent of Asp169 to asparagine or Glu170 to glutamine in mouse eIF4A abolishes ATPase and duplex unwind-

FIGURE 3. Topology of the amino terminal domains of eIF4A (top) HCV (middle), and Rep/PcrA (bottom). Arrows represent β strands; cylinders represent α helices. The RecA-like core is drawn with bold lines; β strands and α helices of the core are numbered to illustrate identical topology. Inclusion of helix E is arbitrary. Sequences of the corresponding helicase motifs are shown in boxes.

ing activities, but does not impair RNA binding, whereas mutating the equivalent of Lys71 in the P-loop to asparagine abolishes all three activities (Pause & Sonenberg, 1992).

In addition to the interactions of the carboxyls of the first two residues of the DEAD-box motif with residues of the closed P-loop, the fourth residue of the DEAD box, Asp172, interacts with residues of the conserved SAT motif in the neighboring β strand (Fig. 4A). Specifically, the side chain hydroxyls of both Ser200 and Thr202 of this motif make hydrogen bonds to a single oxygen of the carboxyl of Asp172. By comparison, in the structure of Benz et al. (1999), each of the two hydroxyls makes a hydrogen bond to a different oxygen of the Asp172 carboxyl. It is notable in this context that mutating the SAT motif of mouse eIF4a to AAA abolished RNA unwinding, while at the same time enhancing ATPase and RNA binding activity (Pause & Sonenberg, 1992). In a similar vein, mutating the DEAD motif in mouse eIF4A to DEAH also enhanced ATPase and RNA binding activities, while reducing RNA unwinding activity to $\sim 10\%$ of the value seen with wild-type protein (Pause & Sonenberg, 1992). (To what extent the aspartate-to-histidine mutation would affect the specific interactions with the SAT motif is unclear. In the structure of the HCV helicase (Yao et al., 1997; Cho et al., 1998; Kim et al., 1998), the side chain hydroxyls of the analogous TAT motif are directed toward the imidazole of the histidine side chain, suggesting that

similar hydrogen bonds could form in eIF4A with a D172H mutation, albeit with somewhat altered geometry.) In a more distant system, a SAT $\rightarrow$ LAT mutation in the yeast splicing factor PRP2 results in a dominant negative phenotype that can be partially overcome by simultaneous expression of wild-type PRP2 protein; in vitro, the mutant protein has 40% of the wild-type ATPase activity, but splicing activity is abolished (Plumpton et al., 1994). These mutational data suggest that the specific interactions we observe between the SAT and DEAD-box motifs may be crucial for coupling the ATPase cycle with RNA duplex unwinding.

Implications

eIF4A is a representative member of the DEX(D/H)-box RNA helicase protein family. In view of the relatively high level of amino acid sequence conservation of the minimal ~ 400 residue helicase domain of these proteins, it is probable that many of the observations that can be made about eIF4A will generalize to other members of the family. The structure of the amino-terminal domain presented here agrees with that presented by Benz et al. (1999) in overall tertiary structure and boundaries of the well-ordered part of the domain, despite both the lengths of the recombinant proteins and the crystallization conditions being different. The fact that the boundaries of our domain were derived from a partial model of full-length eIF4A removes any question as

to whether the construct used by Benz et al. (1999) deleted any portion of the amino-terminal domain.

The structure reveals the presence of the conserved, RecA-like core that has been observed in other helicases, and reinforces the suggestion that this core forms a structural scaffold about which the ATP binding domains of helicases are built (Subramanya et al., 1996). The additional structural elements that augment the core in this domain of eIF4A differ from those in both the DNA helicases and the HCV RNA helicase. Whether these additional elements participate in specific biochemical functions is not known at this time; it is provocative to consider whether these structural elements may be sites of specific protein-protein interaction with eIF4G or other proteins involved in translation initiation.

Although it has not been possible to unambiguously trace the complete structure of the second domain of eIF4A at this time, it is clear that it is connected to the first domain by an extended linker of ~10–12 amino acid residues (Fig. 1C–D). In our triclinic crystal form, it appears to be held in place by relatively weak intermolecular crystal contacts, which presumably stabilize a particular conformation of what may be a relatively flexible molecule in solution. If eIF4A were to bind single-stranded oligonucleotides in a manner similar to that observed for HCV (Kim et al., 1998) and the DNA helicases (Korolev et al., 1997; Velankar et al., 1999), where two domains clamp onto the oligonucleotide, it would need to undergo a substantial change from the extended conformation observed in our crystals to bring the two domains together. When RNA or ATP/ADP bind to mouse eIF4A, it alters the digestion pattern of proteolytically sensitive sites, supporting a proposal for ATP- and RNA-induced conformational changes in the molecule (Lorsch & Herschlag, 1998).

The alternative scenario is that eIF4A would bind oligonucleotides in a significantly different manner than HCV and the DNA helicases. It is notable in this context that in the HCV helicase, the two domains that are responsible for RNA binding are positioned in close proximity to each other primarily through interactions with the third domain; there is little direct contact between them (Yao et al., 1997). We see no apparent third domain in the eIF4A polypeptide that could serve a similar function (although it is conceivable that such a function could be effected by the other proteins with which eIF4A interacts, eIF4B and eIF4G).

Data in the literature on the effects of site-specific mutations in eIF4A proteins suggest that the interactions we observe between amino acid residues of the P-loop, the DEAD box, and the SAT motif may relate to the (currently unknown) molecular mechanism of coupling the ATPase activity with RNA unwinding and intramolecular conformational change(s). Consensus sequences that accommodate most of the members of the DEX(D/H)-box protein family are (G/A)x(S/T)G(S/C/T)GK(S/T), DEX(D/H), and SAT respec-

tively for these three motifs, where the residues involved in the interactions displayed in Figure 4A are underlined. It is apparent that the variations in amino acid sequences within the DEX(D/H)-box RNA helicase family would still conserve most, and in many cases all of the specific intramolecular interactions observed in this region of the molecule. The precise functional significance of these specific interactions and of the closed conformation of the P-loop that we observe in this crystal form of yeast eIF4A remains to be determined.

MATERIALS AND METHODS

In the experimental work, we first crystallized full-length eIF4A protein and proceeded with its structure determination. When it became apparent that only the amino-terminal domain of the molecule was well-ordered in these crystals, we expressed this domain, crystallized it, and solved its structure to complete a model whose overall reliability would not suffer the uncertainties inherent to a crystal in which a significant fraction of the molecule is disordered. Below, we describe experimental work on both the full-length protein and the amino-terminal domain in sequence.

Protein expression and purification

An expression plasmid in which the eIF4A gene had been cloned into a pET21c vector (Novagen) was provided by Drs. John Lorsch and Dan Herschlag. Recombinant full-length yeast eIF4A (amino acids 1–394) containing SeMet was expressed in the methionine auxotroph *E. coli* strain B834(DE3) (Novagen). Freshly transformed cells were grown at 37 °C to an optical density of 0.5 in M9 medium supplemented with 10% Luria broth (LB) and 25 mg/mL SeMet (Sigma). Cells were then centrifuged, resuspended in M9 medium containing 25 mg/mL SeMet and 0.4 mM isopropyl- β -D-thiogalactoside (IPTG), and grown for 3–4 h at 37 °C. Cells were harvested by centrifugation, resuspended in 100 mM KCl, 25 mM Bis-tris, pH 7.4, (Buffer A), and broken open with a French press. Protein in the cell lysate was precipitated at 4 °C by addition of ammonium sulfate to a concentration of 2.8 M. The precipitate was resuspended and dialyzed into Buffer A, applied to a Q-sepharose fast flow column (Pharmacia) preequilibrated with the same buffer, and eluted with a gradient of 0.1–0.6 M KCl. Fractions containing eIF4A protein were combined and chromatographed by gel filtration on a Superdex-75 column (Pharmacia). eIF4A eluted at the volume expected for a monomeric protein.

To express the amino terminal domain of eIF4A (residues 1–223), a stop codon was introduced after codon 223 by site-directed mutagenesis (Stratagene Quik-change procedure) in the plasmid used for expression of full-length protein. Protein was expressed in *E. coli* strain BL21(DE3) (Novagen) in LB medium at 37 °C and purified using the same protocol described above for the full-length protein.

Crystallization and data collection

Full-length SeMet-substituted eIF4A protein (~10 mg/mL) was crystallized at 20 °C in 50 mM CHES, 100 mM KOAc, 0.2 M

guanidinium HCl, 20% polyethylene glycol monomethyl ether 5,000 (PEG-MME₅₀₀₀), pH 9.0 in hanging drops. (In systematic trials, it was found that a minimum of 50 mM guanidinium HCl was required for crystallization; this could not be substituted by other compounds with a guanidino group, such as arginine. The basis for this requirement is unclear; it may improve the solubility of the protein). A crystal with dimensions $\sim 0.4 \text{ mm} \times \sim 0.15 \text{ mm} \times \sim 0.04 \text{ mm}$ was soaked for $\sim 1 \text{ min}$ in crystallization buffer with the addition of 20% ethylene glycol as cryoprotectant and flash-frozen in a liquid N₂ stream at 100 K. The crystal was triclinic, space group P1, with cell dimensions $a = 38.8 \text{ \AA}$, $b = 71.2 \text{ \AA}$, $c = 73.2 \text{ \AA}$, $\alpha = 94.0^\circ$, $\beta = 89.6^\circ$, $\gamma = 101.0^\circ$, and with two eIF4A protomers per unit cell.

Diffraction data were collected on beamline 1–5 of the Stanford Synchrotron Radiation Laboratory (SSRL). Three wavelengths near the selenium K edge ($\lambda = 0.9250$, 0.9793 , and $0.9799 \text{ \AA}$, for remote, peak, and edge datasets, respectively) were chosen from the fluorescence spectrum of the crystal. Data frames of 1.5° oscillation were collected with a Quantum IV CCD detector through a phi rotation range of 360° . The images were converted to the Mar 30-cm plate format using the program, *smv_to_bigmar* (courtesy of Dr. Frank Whitby), and processed with the HKL program suite, using the “no merge original index” option in SCALEPACK (Otwinowski & Minor, 1997).

Purified amino terminal domain protein was crystallized at 20°C in hanging drop with 30% PEG-MME₅₀₀₀, 100 mM KOAc, 50 mM CHES, pH 9.0 as precipitant. A crystal with dimensions $0.15 \text{ mm} \times 0.15 \text{ mm} \times 0.03 \text{ mm}$ was soaked approximately 1 min in a cryoprotectant stabilization solution constituted of 30% PEG-MME₅₀₀₀, 20% ethylene glycol, 100 mM KOAc, 5 mM Mg(OAc)₂, 5 mM ADP, 50 mM HEPES, pH 7.0, and then flash-frozen in a stream of N₂ gas at 100 K. The crystal was orthorhombic, space group P2₁2₁2₁ with unit cell dimensions $a = 38.8 \text{ \AA}$, $b = 74.3 \text{ \AA}$, $c = 81.0 \text{ \AA}$. Data were collected on SSRL beamline 1–5 using the same instrumentation as described above; the wavelength was $0.9999 \text{ \AA}$; and frames of 1° oscillation were collected through a cumulative oscillation range of 120° . Data frames were processed directly with the HKL program suite (Otwinowski & Minor, 1997).

Structure determination and refinement

Full-length eIF4A

The unmerged reflections ($40\text{--}2.8 \text{ \AA}$) output by the program SCALEPACK were scaled with SOLVE [version 1.04, beta release (Terwilliger & Berendzen, 1999)] using the “SCALE_MAD” option. The scaled, merged reflections and anomalous differences were transferred to the CCP4 mtz file format and subsequent Patterson and Fourier calculations were carried out with the CCP4 program suite (Bailey, 1994). Both anomalous (f'' peak) and dispersive (remote vs. f' edge) differences were used for Patterson map calculations. Previous work had shown that the two eIF4A molecules in the unit cell were related by an approximate noncrystallographic twofold screw axis nearly parallel to the crystallographic c axis (data not shown), which facilitated Patterson interpretation. The locations of 18 of a possible 22 selenium atoms were

found by selecting pairs of sites consistent with strong Patterson peaks and the noncrystallographic symmetry, followed by phase calculations in MLPHARE and subsequent anomalous and dispersive difference Fouriers to locate other peaks. The dataset collected at the Se K edge ($0.9799 \text{ \AA}$) was treated as a native dataset and remote ($0.9250 \text{ \AA}$) and absorption (f'') peak ($0.9793 \text{ \AA}$) datasets were treated as derivatives with anomalous scattering (Ramakrishnan & Biou, 1997; Terwilliger & Berendzen, 1999). After 16 of the possible 22 selenium sites were found, a Fourier map calculated to $2.8 \text{ \AA}$ using phases computed in MLPHARE of the CCP4 package gave electron density with clear solvent boundaries and recognizable secondary structures. Two additional selenium sites were placed and heavy atom refinement and experimental phase calculation was done using SHARP (de La Fortelle & Bricogne, 1997), which gave a significantly improved map. Two clusters of eight selenium atoms related by the noncrystallographic twofold screw axis were used to determine the noncrystallographic symmetry (NCS) operator. The NCS operator between the two sets of sites was used for twofold averaging and solvent flattening of the experimental map with SOLOMON (Abrahams & Leslie, 1996), which gave some improvement in the electron density for both molecules. At this stage, much of the backbone of one domain of the molecule could be traced, and sequence was placed with the aid of the selenium sites to identify the positions of SeMet residues.

An initial model consisting of 191 residues, 58 of which were alanine, was built into the experimental map for both molecules in the unit cell and refined with CNS (version 0.3c; Brünger et al., 1998) against all data from $40\text{--}2.8 \text{ \AA}$, using the experimental Hendrickson–Lattman coefficients calculated by SHARP as the maximum likelihood refinement target (mlhl) throughout. The NCS operator was improved by superposition of the refined partial models of the two molecule. Positional minimization and simulated annealing at 2500 K followed by slow cooling in 25 K steps gave an R_{cryst} of 0.328, R_{free} of 0.363. Model and experimental phases were combined to produce $2F_{\text{obs}} - F_{\text{calc}}$ and $F_{\text{obs}} - F_{\text{calc}}$ maps. After several cycles of model building and refinement, a continuous model for residues 11–223 was built, but virtually no electron density was observed for residues 1–10, and density beyond residue 223 was relatively weak and difficult to interpret.

Amino-terminal domain

A model of the amino-terminal domain was placed in the orthorhombic crystal cell using the rotation and translation search options of CNS (version 0.5; Brünger et al., 1998). The top translation function solution obtained with CNS gave $R_{\text{cryst}} = 0.329$ and $R_{\text{free}} = 0.333$ for data $28\text{--}2.8 \text{ \AA}$. One cycle of positional refinement and simulated annealing (2500 K) using the maximum likelihood (mlf) target in CNS was used to improve the fit between calculated and observed structure factors. Subsequent rounds of refinement and model building were done using all data collected, which was reasonably complete to $2.5 \text{ \AA}$ and had partial data to $2.2 \text{ \AA}$. The final model includes residues 11–223 of the protein and 55 solvent molecules, and has $R_{\text{cryst}} = 0.223$ and $R_{\text{free}} = 0.274$ for data in the range $28\text{--}2.5 \text{ \AA}$ (Table 2). MALDI-TOF mass spectroscopy on protein from crystallization gave a molecular weight of 25,050, which is consistent with an ~ 220 amino acid pro-

tein; the amino terminal 10 residues are presumably disordered in the crystal rather than proteolytically clipped from the protein. Coordinates have been deposited in the RCSB PDB, accession identifier 1QVA.

ACKNOWLEDGMENTS

We thank Drs. Dan Herschlag and Jon Lorsch for plasmids and discussions, Henry Bellamy, Sigurd Wilbanks, and Christine Trame for assistance with data collection, and Joe Wedekind for helpful discussions. This work was supported by grants GM39928 from the National Institutes of Health and MCB-9874528 from the National Science Foundation to D.B.M. This work is based upon research conducted at the Stanford Synchrotron Radiation Laboratory, which is funded by the Department of Energy, Office of Basic Energy Sciences. The Biotechnology Program is supported by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program and the Department of Energy, Office of Biological and Environmental Research.

Received July 8, 1999; returned for revision August 5, 1999; revised manuscript received August 31, 1999

REFERENCES

- Abrahams JP, Leslie GW. 1996. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D* 52:30–42.
- Bailey S. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D* 50:760–763.
- Benz J, Trachsel H, Baumann U. 1999. Crystal structure of the ATPase domain of translation initiation factor 4A from *Saccharomyces cerevisiae*—the prototype of the DEAD box protein family. *Structure* 7:671–679.
- Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D* 54:905–921.
- Cho HS, Ha NC, Kang LW, Chung KM, Back SH, Jang SK, Oh BH. 1998. Crystal structure of RNA helicase from genotype 1b hepatitis C virus. A feasible mechanism of unwinding duplex RNA. *J Biol Chem* 273:15045–15052.
- de La Fortelle E, Bricogne G. 1997. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength diffraction methods. *Methods Enzymol* 276:472–494.
- Esnouf RM. 1997. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J Mol Graph Model* 15:132–134.
- Esnouf RM. 1999. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr D* 55:938–940.
- Gorbalenya AE, Koonin EV. 1993. Helicases: Amino acid sequence comparisons and structure-function relationships. *Curr Opin Struct Biol* 3:419–429.
- Jaramillo M, Dever TE, Merrick WC, Sonenberg N. 1991. RNA unwinding in translation: Assembly of helicase complex intermediates comprising eukaryotic initiation factors eIF-4F and eIF-4B. *Mol Cell Biol* 11:5992–5997.
- Jones TA, Zhou JY, Cowan SW, Kjeldgaard M. 1991. Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 47:110–119.
- Kim JL, Morgenstern KA, Griffith JP, Dwyer MD, Thomson JA, Murcko MA, Lin C, Caron PR. 1998. Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: The crystal structure provides insights into the mode of unwinding. *Structure* 6:89–100.
- Korolev S, Hsieh J, Gauss GH, Lohman TM, Waksman G. 1997. Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. *Cell* 90:635–647.
- Korolev S, Yao N, Lohman TM, Weber PC, Waksman G. 1998. Comparisons between the structures of HCV and Rep helicases reveal structural similarities between SF1 and SF2 super-families of helicases. *Protein Sci* 7:605–610.
- Kraulis P. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* 24:946–950.
- Lorsch JR, Herschlag D. 1998. The DEAD box protein eIF4A. 2. A cycle of nucleotide and RNA-dependent conformational changes. *Biochemistry* 37:2194–2206.
- Luzzati PV. 1952. Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Crystallogr* 5:802–810.
- Method N, Pause A, Hershey JW, Sonenberg N. 1994. The translation initiation factor eIF-4B contains an RNA-binding region that is distinct and independent from its ribonucleoprotein consensus sequence. *Mol Cell Biol* 14:2307–2316.
- Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
- Pain VM. 1996. Initiation of protein synthesis in eukaryotic cells. *Eur J Biochem* 236:747–771.
- Pause A, Sonenberg N. 1992. Mutational analysis of a DEAD box RNA helicase: The mammalian translation initiation factor eIF-4A. *EMBO J* 11:2643–2654.
- Plumpton M, McGarvey M, Beggs JD. 1994. A dominant negative mutation in the conserved RNA helicase motif "SAT" causes splicing factor PRP2 to stall in spliceosomes. *EMBO J* 13:879–887.
- Py B, Higgins CF, Krisch HM, Carpousis AJ. 1996. A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. *Nature* 381:169–172.
- Ramakrishnan V, Biou V. 1997. Treatment of multiwavelength anomalous diffraction data as a special case of multiple isomorphous replacement. *Methods Enzymol* 276:538–557.
- Rogers GW Jr, Richter NJ, Merrick WC. 1999. Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. *J Biol Chem* 274:12236–12244.
- Rozen F, Pelletier J, Trachsel H, Sonenberg N. 1989. A lysine substitution in the ATP-binding site of eucaryotic initiation factor 4A abrogates nucleotide-binding activity. *Mol Cell Biol* 9:4061–4063.
- Sayle RA, Milner-White EJ. 1995. RASMO: Biomolecular graphics for all. *Trends Biochem Sci* 20:374.
- Schmid SR, Linder P. 1991. Translation initiation factor 4A from *Saccharomyces cerevisiae*: Analysis of residues conserved in the D-E-A-D family of RNA helicases. *Mol Cell Biol* 11:3463–3471.
- Smith CA, Rayment I. 1996. Active site comparisons highlight structural similarities between myosin and other P-loop proteins. *Biochem J* 315:1590–1602.
- Staley JP, Guthrie C. 1998. Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell* 92:315–326.
- Story RM, Steitz TA. 1992. Structure of the recA protein-ADP complex. *Nature* 355:374–376.
- Subramanya HS, Bird LE, Brannigan JA, Wigley DB. 1996. Crystal structure of a DExx box DNA helicase. *Nature* 384:379–383.
- Terwilliger TC, Berendzen J. 1999. Automated structure solution for MIR and MAD. *Acta Crystallogr D* 55:849–861.
- Velankar SS, Soutanas P, Dillingham MS, Subramanya HS, Wigley DB. 1999. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* 97:75–84.
- Venema J, Tollervey D. 1995. Processing of pre-ribosomal RNA in *Saccharomyces cerevisiae*. *Yeast* 11:1629–1650.
- Yao N, Hesson T, Cable M, Hong Z, Kwong AD, Le HV, Weber PC. 1997. Structure of the hepatitis C virus RNA helicase domain. *Nat Struct Biol* 4:463–467.