

Crystal Structure of Conserved Domains 1 and 2 of the Human DEAD-box Helicase DDX3X in Complex with the Mononucleotide AMP

Martin Högbom¹†, Ruairi Collins¹†, Susanne van den Berg¹
Rose-Marie Jenvert², Tobias Karlberg¹, Tetyana Kotenyova¹
Alex Flores¹, Gunilla B. Karlsson Hedestam³
and Lovisa Holmberg Schiavone^{1*}

¹Structural Genomics Consortium, Department of Medical Biochemistry and Biophysics, Karolinska Institute SE-171 77 Stockholm, Sweden

²School of Life Sciences Södertörns högskola SE-141 04 Huddinge, Sweden

³Department of Microbiology Tumor and Cell Biology Karolinska Institute SE-171 77 Stockholm, Sweden

DExD-box helicases are involved in all aspects of cellular RNA metabolism. Conserved domains 1 and 2 contain nine signature motifs that are responsible for nucleotide binding, RNA binding and ATP hydrolysis. The human DEAD-box helicase DDX3X has been associated with several different cellular processes, such as cell-growth control, mRNA transport and translation, and is suggested to be essential for the export of unspliced/partially spliced HIV mRNAs from the nucleus to the cytoplasm. Here, the crystal structure of conserved domains 1 and 2 of DDX3X, including a DDX3-specific insertion that is not generally found in human DExD-box helicases, is presented. The N-terminal domain 1 and the C-terminal domain 2 both display RecA-like folds comprising a central β -sheet flanked by α -helices. Interestingly, the DDX3X-specific insertion forms a helical element that extends a highly positively charged sequence in a loop, thus increasing the RNA-binding surface of the protein. Surprisingly, although DDX3X was crystallized in the presence of a large excess of ADP or the slowly hydrolyzable ATP analogue ATP γ S the contaminant AMP was seen in the structure. A fluorescent-based stability assay showed that the thermal stability of DDX3X was increased by the mononucleotide AMP but not by ADP or ATP γ S, suggesting that DDX3X is stabilized by AMP and elucidating why AMP was found in the nucleotide-binding pocket.

© 2007 Elsevier Ltd. All rights reserved.

*Corresponding author

Keywords: DEAD-box; helicase; HIV; RNA; nucleotide

Introduction

RNA-binding DExD-box helicases are ATPases that function as RNA-unwinding enzymes, chaperones that promote the correct folding of RNAs or assembly/disassembly factors of protein–RNA complexes, during all stages of RNA-metabolism.¹

Present address: M. Högbom, Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden.

† M.H. and R.C. contributed equally to this work.

Abbreviations used: TLC, thin-layer chromatography; RRE, Rev-responsive element.

E-mail address of the corresponding author: lovisa.holmberg-schiavone@ki.se

DDX3 is a human DEAD-box helicase that is found in the two closely related forms DDX3X and DDX3Y (encoded on the X chromosome and Y chromosome respectively),² which are 92% identical in sequence. DDX3Y is produced specifically in spermatocytes, where it is essential for spermatogenesis.^{3,4} DDX3X has been associated with several different cellular processes, such as cell-growth control,⁵ transport of mRNA⁶ and translation.^{7,8} Interestingly, during HIV replication, the transport of unspliced/partially spliced HIV mRNAs that contain a Rev-responsive element (RRE) from the nucleus to the cytoplasm is dependent on the viral protein Rev, shuttling protein CRM1 (exportin 1) and DDX3X. Rev can be immunoprecipitated by DDX3X and CRM1, suggesting that the proteins form complex(es), and it has been proposed that the helicase activity of

DDX3X enables the threading of HIV mRNAs through the nuclear pore.⁹ The involvement of DDX3X in HIV infection makes it an interesting candidate for antiviral therapy,¹⁰ and it is therefore important to determine the three-dimensional structure of the human DDX3X DEAD-box helicase.

The RNA-binding DExD-box proteins belong to the SF2 superfamily of helicases based on their primary sequence. Most DExD-box helicases contain two subdomains (named 1 and 2) that are often flanked by divergent sequences.¹ Nine short consensus sequences that form the conserved helicase core are found in the two subdomains; motifs Q, I (Walker A containing the phosphate-binding P-loop), II (Walker B, DExD-box), Ia, Ib and III are found in domain 1, whereas motifs IV, V and VI are found in domain 2. Structural, biochemical and mutational analysis show that the different motifs are involved in nucleotide binding (Q, I and II), RNA binding (Ia, Ib, IV and V) and ATP hydrolysis (III and possibly VI).¹ Structure determination of individual domains shows that both domains 1 and 2 have a fold belonging to the RecA superfamily with five β -strands surrounded by five α -helices.¹ So far, structures containing both domains, connected *via* a short flexible linker have been solved of the human splicing factor UAP56 with/without ADP,¹¹ human eIF4A-III,^{12,13} yeast mRNA translation and degradation factor Dhh1p,¹⁴ *Methanococcus jannaschii* DEAD-box protein,¹⁵ the *Drosophila* VASA protein,¹⁶ and the eIF4A protein.¹⁷ There are additional structures of domain 1, including eIF4A from yeast without^{18,19} or with ADP/ATP bound,¹⁹ human UAP56 with a citrate ion bound,²⁰ *Dugesia japonica* VASA-like gene B protein,²¹ BstDEAD,²² and *Thermus thermophilus* Hera helicase in complex with AMP or orthophosphate.²³ Moreover, a separate structure of domain 2 of UAP56 has been solved.²⁰ The two-domain structures with nucleotide present show its location in a cleft between the two domains, interacting mainly with domain 1, with the adenine moiety forming interactions with amino acid residues in motif I, whereas the γ and β phosphate groups on the nucleotide interact with amino acid residues in motif II through a magnesium ion.¹ The relative orientation of the two domains varies extensively between the different structures as long as no RNA substrate is present.^{11,12,14,15,17} It is now becoming evident that DExD-box helicases acquire an active, well defined and closed conformation upon RNA binding as observed in the structures of *Drosophila* VASA protein and human eIF4AIII in complex with poly(U).^{12,13,16} It is seen from the structures with substrate present that the RNA is bent upon binding, hinting at possible mechanisms for helicase activity.^{12,13,16,24}

Recombinant full-length DDX3X produced in bacteria displays ATPase and RNA-unwinding activity.⁹ The full-length protein (662 amino acid residues) consists of domains 1 and 2 flanked by N and C-terminal divergent sequences. Interestingly, there is also a DDX3X (and Y)-specific insertion between motifs I and Ia (residues 250–259) which is

not generally found in other DExD-box helicases. Here, we present the structure of domains 1 and 2 in DDX3X comprising the helicase core (residues V168–G582) in complex with AMP at 2.2 Å resolution, and reach conclusions about the specific insertion.

Results

A sequence alignment of six different DDX3 orthologues (not shown) shows that human DDX3X displays a high level of identity with the mouse PL10 (95%), human DDX3Y (92%) and *Xenopus* An3 (85%) proteins, whereas the level of identity with *Drosophila* Belle (59%) and yeast Ded1p (51%) proteins is lower. This alignment shows that the domain 1 and domain 2 conserved sequence comprising the helicase core, including the nine consensus motifs, starts around residue 150 (DDX3X numbering) and ends around residue 550 (DDX3X numbering).

A multiple construct approach was applied to identify truncated versions of DDX3X, spanning domains 1 and 2, that are expressed well and produce soluble protein in a small-scale format (results not shown). For this purpose, 11 different constructs of DDX3X were cloned into the pNIC-Bsa4 vector that encodes a His-tag followed by a TEV-protease cleavage site immediately upstream of the inserted gene or gene fragment. Constructs spanning both domains, starting at V168 or A170 and ending at Y576 or G582, produced soluble protein in a small-scale format and purified well in a large-scale format (results not shown). However, expressed and purified protein starting at A170 in the DDX3X sequence precipitated during concentration. In the structure of DDX3X (see below) the preceding V168 and A169 take part in the formation of a large β -sheet, explaining why the construct starting at A170 was unstable after purification. Instead, two other constructs were pursued, V168–Y576 and V168–G582. In the end, the best-diffracting crystals were obtained with the construct spanning V168–G582. Crystallization trials were performed without nucleotide and with ADP, ATP γ S, ANP and AMPPCP but crystals were produced only with ADP and ATP γ S. Curiously, the monophosphate AMP was seen in both of the resulting structures (see below).

Overall structure

The crystallized construct DDX3X(V168–G582) could be traced from V168 (a preceding methionine residue, introduced in the cloning, is also visible in the density) to Y580. The linker between domain 1 and domain 2 (residues 407–410) and residues 535–536 and 581–582 of domain 2 are disordered. As expected, the two-domain structure of DDX3X (V168–G582) displays the two RecA-like domains found in other DExD-box RNA helicases.^{11,12,14,15,17} The N-terminal domain 1 contains the RecA-like

fold comprising a central five-stranded β -sheet flanked by 5 α -helices. An additional β -strand and two α -helices form a cap-like structure (residues 167–215) that includes the Q-motif important for nucleotide binding. Similarly, domain 2 comprises a RecA-like fold, including a central β -sheet flanked by α -helices.

Figure 1 shows the structure with domain 1 on the right and domain 2 on the left, with the consensus motifs indicated. Compared to the two available DEAD-box helicase structures with RNA bound (*Drosophila* VASA¹⁶ and the human eIF4AIII^{12,13}) DDX3X displays a less compact “open” structure with regards to the relative orientation of the domains.

The RNA-bound “closed” form of the proteins seem to be well defined and dictated by the RNA binding. Indeed, 359 of the 420 residues from both domains in the VASA structure can be superimposed on the eIF4AIII structure with a C α rmsd of below 1.5 Å. The open forms, without RNA bound, on the other hand display large differences in the relative orientations between the two domains and the different proteins for which structures are available display diverse conformations.^{11,12,14,15,17}

It thus seems that the sequence linking the two domains provides a very large degree of flexibility. We find it quite likely that the observed open structures are defined by monomer–monomer contacts that favor crystallization rather than a single biologically relevant defined open conformation.

Nucleotide-binding pocket

DDX3X(V168-G582) was co-crystallized with ATP γ S and ADP. In both cases AMP (a contaminant,

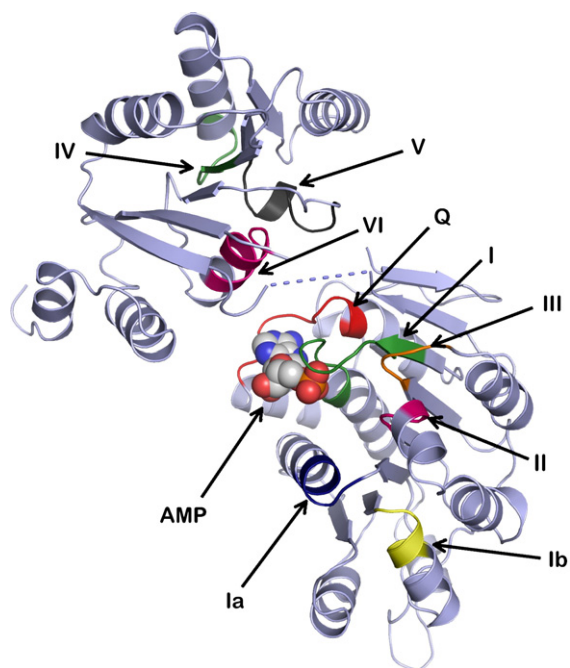


Figure 1. Overall structure. The consensus motifs are indicated, the disordered part of the sequence connecting the two domains is indicated by a broken line.

see below) was seen in the electron density map (Figure 2(a)). The interaction of AMP with amino acid residues in the nucleotide-binding pocket of DDX3X(V168-G582) is highly similar to the interaction of AMP with the N-terminal domain of the *T. thermophilus* DEAD-box helicase Hera.²³ The adenine moiety of the nucleoside interacts with amino acids in the Q motif, whereas residues in the P-loop in motif I interact with the phosphate group. In DDX3X(V168-G582), N6 on the purine interacts with the backbone carbonyl group of arginine 202 and with O ϵ of glutamine 207 in the Q-motif. This interaction scheme gives selectivity for adenine nucleotides compared to guanine nucleotides. N ϵ of glutamine 207 also interacts with the N7 position of the adenine. A conserved aromatic ring (tyrosine 200 in DDX3X(V168-G582), otherwise phenylalanine or a tryptophan) stacks on the base. The phenolic oxygen atom also hydrogen bonds to the 2'OH on the ribose. The phosphate group makes polar interactions with the hydroxyl group on threonine 231 and to backbone nitrogen atoms of residues 227–231 in the P-loop in motif I as well as to three water molecules.

Figure 2(b) shows superpositions of the nucleotide-binding region and DExD-box motif for DDX3X(V168-G582) (blue), UAP56 with ADP (green) (PDB code 1XTJ,¹¹ without ligand (magenta) (PDB code 1XTI)¹¹), VASA (red) with the non-hydrolyzable ATP analogue ANP (PDB code 2DB3),¹⁶ yeast Dhhp (yellow) (PDB code 1S2M)¹⁴ and yeast eIF4A (black) (PDB code 1FUU).¹⁷ The AMP molecule from DDX3X(V168-G582) and ANP from VASA are shown for reference. It seems that the P-loop does not adopt distinct open or closed conformations but has a flexibility that allows it to adapt to several conformations and sizes of ligands.²³ The most restricted (closed) conformation is found in yeast eIF4A (black), which leaves no room for a ligand. The DDX3X(V168-G582) structure with AMP bound adopts a more open conformation (blue) that is similar to the UAP56 without any ligand bound (magenta). UAP56 in complex with ADP (green) adopts an open conformation that is very similar to VASA with a triphosphate ligand (red) and to yeast Dhhp (yellow) with no ligand bound. The side-chains of the DExD motif adopt very much the same conformations, independent of the ligand, the only conformations that stand out are the two UAP56 structures that display different rotamers of the glutamate residue in the motif.

RNA-binding surface

The DDX3X(V168-G582) domain 2 has to rotate approximately 180° relative to domain 1 to obtain the closed conformation of the protein required for RNA binding. This rearrangement would bring positively charged patches on the solvent-exposed surface of domain 2 into closer proximity with positively charged surfaces on domain 1, thus forming the RNA-binding site of the protein. A comparison of the closed structure of the VASA-

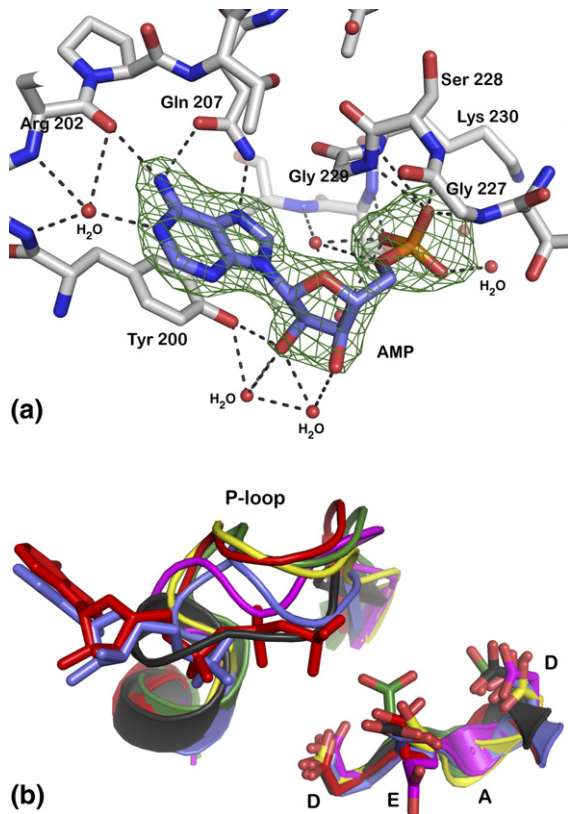


Figure 2. (a) AMP binding in DDX3X(V168-G582), omit $F_o - F_c$ electron density for the ligand, contoured at $0.14 \text{ e}/\text{\AA}^3$ (4σ). Polar interactions and hydrogen bonds are indicated by broken lines and water molecules are represented by red spheres. (b) Superposition of the nucleotide-binding region with the P-loop and DEAD motif for DDX3X(V168-G582) (blue), UAP56 with ADP (green), without nucleotide (magenta), VASA (red), yeast Dhhp (yellow) and yeast eIF4A (black). The AMP molecule from DDX3X(V168-G582) (blue) and ANP from VASA (red) are shown for reference.

protein in complex with poly(U) with a modeled closed structure of the DDX3X(V168-G582) protein show that all amino acid residues that are involved in interaction with the RNA in the VASA-structure are present in the DDX3X(V168-G582) structure at corresponding positions. This indicates, as expected, that these residues are involved in RNA binding in both proteins, and that the RNA binding mode in this area should be very similar.

DDX3X has an insertion of ten residues (250–259, EALRAMKENG) between the P-loop and the RNA-binding motif Ia. A sequence alignment of 32 different human DEXD-box helicases shows that this striking insertion is seen only in DDX3X and DDX3Y (Figure 3). The inserted sequence precedes a positively charged sequence, maintained in the different DDX3-orthologues.

The insertion forms a helix that packs against the protein and extends the following highly positively charged (RYGRRK) loop sequence, also unique to DDX3X/Y (Figures 3 and 4). Interestingly, this loop, together with other surface residues on DDX3X (V168-G582), forms a continuous positively charged cavity. A model of the closed RNA-bound form of the DDX3X(V168-G582) structure, based on the VASA structure and its short RNA ligand shows that this cavity is positioned close to the 3' end of the RNA. It seems very likely that this extended patch, which is found only in the DDX3X/Y proteins, is involved in RNA binding and positioning in these proteins.

ATP hydrolysis and nucleotide-binding properties of DDX3X(V168-G582)

Full-length DDX3X hydrolyzes ATP in an RNA-dependent fashion.⁹ Surprisingly, DDX3X(V168-G582) displayed no ATPase activity in the presence of four different RNA substrates and at several different conditions tested (results not shown, see Materials and Methods).

DDX25	GKTAAFVLA---MLSRVNAL-----ELFPQCLCLAPTYELA
DDX19	GKTAAFVLA---MLSQVEPA-----NKYEQCLCLSPTYELA
UAP56	GKTAAFVLA---TQQIEPV-----NGQVTVLVMCHRELA
eIF4AIII	GKTATFSIS---VLQCLDIQ-----VRETQALILAPTRELA
DDX20	GKTCVFSSTI---ALDSVLE-----NLSTQILILAPTRELA
eIF4A	GKTATFAIS---ILQQIELD-----LKATQALVLAAPTRELA
DDX42	GKTAAFIWF---MLIHIMDQ-----KELEPG-DGPIAVIVCPRELC
DDX46	GKTI AFLLE---MFRHIMDQ-----RSLEEG-EGPIAVIMPTRELA
DDX4	GKTA AFLLE---ILAHMMHDGITA-----SRFKEL-QEPECIIVAPTRELV
DDX59	AFLLE---VIM-----RALFES-KTFSALIIPTRELA
DDX23	AFLLE---LVVWITLP-KI-----DRIESDQGYAII LAPTRELA
DDX17	AYLLE---AIVHINHQ-PY-----LE-R-GDGPICLV LAPTRELA
DDX5	SYLLE---AIVHINHQ-PF-----LE-R-GDGPICLVLA
DDX53	GKTL SYLMP---GFIH LDSQ-PI-----SREQ-RNGP GMLVLT
DDX43	GKTL CYLMP---GFIH VLQ-PS-----LKGQ-RNRPGMLVLT
DDX28	GKTL SYLLE---LQRL LGQ-PS-----LDSLPIPARGLVLT
DDX21	GKTF SFAIF---LIEK LHGE-LQ-----DRKR-GRAEQVLV LAPTRELA
DDX49	GKTA AFVLP---ILQK LSED-P-----YG-----IFCLVLTPTRELA
DDX47	GKTG AFALP---ILNA LET-P-----QR-----LFALVLTPTRELA
DDX3X	GKTA AFLLP---ILSQIYSDG EALRAMKENG YGRRKQYFISLV LAPTRELA
DDX3Y	GKTA AFLLP---ILSQIYTDG EALKAVKENG YGRRKQYFISLV LAPTRELA
DDX27	GKTA AFALP---VLER IYK-I-----AQAP---VTRV LVLVPTRELG
DDX6	KGSGAYLLE---LLERL---D-----LK--KDNIQAMVIVPTRELA
DDX55	GKTLAEVIF---ILEI LLRREEK-----LK--KSQVGAIIITPTRELA
DDX56	GKTAAYATP---MQLL LHRKAT-----GPVVEQAVRGLV LVPKELA
DDX51	GKTLAEVIF---VVQA LLSR-----VVCHIRALVVLPTKELA
DDX31	GKTLAYCIP---VVQS LQAMESK-----IQ-RSDG EYALV LVPPTRELA
DDX18	GKTLAEFLIPAVELIVKLR-FMPR-----NGTGVL IISPTRELA
DDX10	GKTLAEFLVLEALYRQ-WTST-----DGLGVLIISPTRELA
DDX52	GKTLAFSIE---ILMQ LK-QPAN-----KGFRAI IISPTRELA
DDX54	GKTLACFLE---LMFER LKTHSAQ-----TGARAL IISPTRELA
DDX1	GYIVKSQHS---GNAQVTQT-----KFLPNA EKALIVE SRELA

Figure 3. Sequence alignment of 32 different human DEXD-box helicases covering the conserved region between motifs I and Ia (amino acid residues 229–279 in DDX3X) and showing the DDX3X/Y-specific insertion between residues 250 and 259 (red box). Totally conserved/highly conserved residues are shown in green.

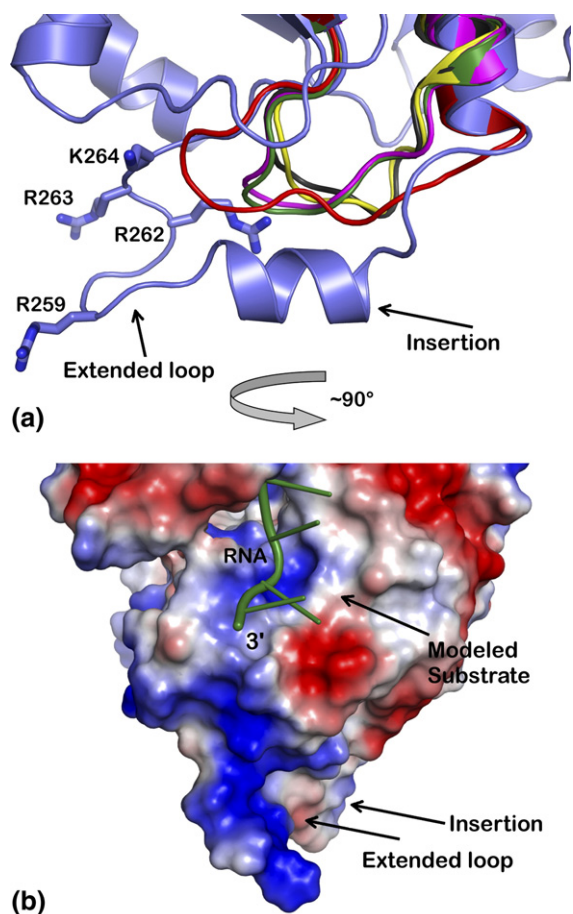


Figure 4. (a) Superposition of the loop region in the vicinity of the DDX3-specific insertion, coloring as in Figure 2(b) and mapped on the DDX3X(V168-G582) structure (blue). Positively charged residues in the extended loop in DDX3X(V168-G582) are shown as sticks. (b) Surface contact potential representation of an RNA-bound model of DDX3X(V168-G582) (based on the *Drosophila* VASA structure). The extended loop creates an elongated, positively charged cavity positioned close to one end of the short RNA ligand.

Although crystallization screens were set up with 20 mM ADP or ATP γ S, AMP was found in the nucleotide-binding pocket of DDX3X(V168-G582) in the crystallized protein (Figure 2(a)) making DDX3X the second DEAD-box helicase solved with AMP, the first being the recently solved domain 1 of Hera helicase.²³ AMP could have been formed by a phosphotransferase activity, spontaneous degradation of ADP and ATP γ S to AMP or by contaminating AMP in the nucleotide stocks. To exclude the possibility that AMP was formed enzymatically by DDX3X(V168-G582) or contaminating *E. coli* proteins the DDX3X(V168-G582) preparation was incubated with [α -³²P]ATP for 1 h at 37 °C and the formation of [α -³²P]AMP was monitored by running the reaction on thin-layer chromatography (TLC) plates with [α -³²P]AMP as a marker. As can be seen in Figure 5(a), no formation of AMP could be detected in the reaction, making it unlikely that a

phosphotransferase activity was present in the DDX3X(V168-G582) preparation. Analysis of the purity of the ADP by HPLC showed that the ADP was contaminated with 5% AMP (results not shown). As a result, approximately 1 mM AMP was present in the crystallization reaction where crystals appeared within 24 h, together with <500 μ M protein and a large excess of ADP. Further HPLC analysis of ADP that had been incubated with the reservoir solution used for crystallization for 24 h at 6 °C showed that a small amount of ADP was hydrolyzed to AMP within this time-span (results not shown).

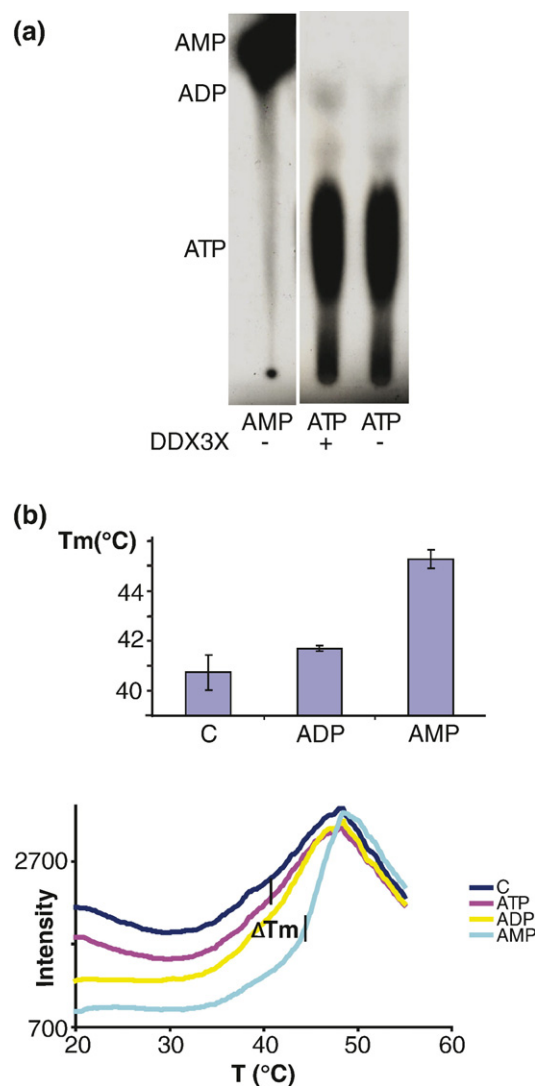


Figure 5. Nucleotide-binding properties of DDX3X (V168-G582). (a) No detectable AMP was formed when DDX3X(V168-G582) was incubated with [α -³²P]ATP and nucleotides separated by TLC and detected by autoradiography/phosphoimager: [α -³²P]AMP was run as a comparison. (b) Thermofluor stability assays suggest DDX3X is stabilized by the mononucleotide AMP. The control shows the thermal shift of DDX3X without added nucleotide, whereas 8 mM nucleotide (final concentration) has been added to the samples labeled ATP, ADP or AMP, error bars are based on three to five different experiments.

A fluorescence-based stability assay was used to elucidate why AMP was preferably bound to DDX3X(V168-G582).²⁵ The thermostability of DDX3X(V168-G582) was monitored in the presence of different nucleotides/nucleotide analogues (8 mM ATP, ADP or AMP, and 4 mM ANP, AMPPCP or ATP γ S). This stability assay showed that the only nucleotide that gave a small increase in the thermal stability of DDX3X(V168-G582), by $\sim 2\text{--}4^\circ$ deg.C was AMP (Figure 5(b) and results not shown).

Altogether, the data suggest that DDX3X(V168-G582) is stabilized by AMP and that the protein crystallized with contaminating AMP that was present in the nucleotide stock and by the spontaneous hydrolysis of small amounts of ADP to AMP.

Discussion

The family of RNA-binding DExD-box helicases displays a relatively high level of sequence similarity and contains nine signature motifs that are present in all known sequences.¹ The general consensus is that specificity of DExD-box helicases is obtained by amino acid sequences that flank conserved domains 1 and 2. However, some sequence divergence is present within the conserved helicase core. One such divergent region is found between nucleotide-binding motif I and RNA-binding motif Ia in domain 1; for example, human eIF4A, its isoform eIF4AIII and UAP56 all contain 25 amino acid residues between conserved amino acids GKT and TRELA in motifs I and Ia, respectively, whereas DDX3X and its male germ cell-specific copy DDX3Y contain 43 residues (Figure 3). The structure presented here shows that the insertion forms a helix that positions a positively charged loop in close proximity to a putative RNA ligand (Figure 4) and we speculate that this unique feature is an important determinant in RNA-substrate interaction by DDX3X/Y. This unique structural feature, together with the particular physiochemical properties of this segment, should be considered in drug design efforts; especially since it most likely is directly involved in substrate binding.

It is somewhat surprising that AMP was found in the nucleotide-binding pocket of DDX3X(V168-G582) but, clearly, the structure presented here shows that AMP-binding to a DEAD-box helicase containing the conserved helicase core is possible and not a peculiarity that is specific to constructs comprising only conserved domain 1, i.e. domain 1 of the *T. thermophilus* Hera helicase,²³ where structural determinants for correct formation of the nucleotide-binding pocket may be lacking. The results presented here suggest that the monophosphate form of the nucleotide is found in the nucleotide-binding pocket, since it has a stabilizing effect on DDX3X(V168-G582). The physiological relevance of this observation awaits further elucidation, since AMP should not be formed during the ATP hydrolysis cycle and cellular concentrations of ATP and ADP should be in large excess of AMP.

A recent characterization of the ATPase and helicase activity of full-length DDX3X indicates that DDX3X displays relaxed substrate specificity with regards to the base and sugar moiety, and the authors suggest that this relaxed specificity could be used in the development of specific inhibitors.²⁶ However, the structure presented here shows that the interaction of DDX3X(V168-G582) with the base and ribose is similar to that of other DExD-box helicases and it does not give any clues to the suggested relaxed substrate specificity described in the other study.²⁶

The conformation of the P-loop differs slightly in DDX3X(V168-G582) compared to the other published structures. It has been postulated that this loop is a molecular sensor of bound nucleotide, since structures that contain nucleotide generally have a more open conformation of the P-loop than those lacking nucleotide.¹ However, the conformations of the P-loops in DDX3X(V168-G582) and in Hera helicase are neither open or fully closed, suggesting that there are more than two possible states.²³ Therefore, we favor a model where the energetics of none of the P-loop conformations is very much favored, apart from the requirement to provide a binding pocket large enough for the ligand.

In vivo and *in vitro* mutagenesis studies have been performed by others to investigate the role of specific amino acids in the nucleotide-binding pocket. Mutation of lysine 230 to glutamate and serine 382 to leucine reduces the ATPase and RNA-unwinding activity *in vitro* and abolishes expression from RRE-containing mRNAs *in vivo*.⁹ In DDX3X(V168-G582), the nitrogen atom of lysine 230 is approximately 6 Å from the oxygen atom on the α -phosphate group of AMP, suggesting that the β/γ phosphate group on a putative ADP/ATP molecule could interact with lysine 230. Serine 382 in the conserved SAT/motif III is further away from the α -phosphate group (13 Å) and it is unlikely that there would be an interaction with a putative β/γ phosphate group. Instead, it has been suggested that a small reduction in ATP and RNA-binding affinity results in a loss of helicase activity.¹ Mutation of threonine 204 to glutamate, which is located in the Q motif, inactivates hamster-DDX3X, which is more than 95% identical with the human protein. It was shown recently that threonine 204 is phosphorylated by cyclin B *in vitro*, and it was suggested that phosphorylation would lead to an inactive protein.²⁷ The structure rationalizes the observed mutation data, since a glutamate or phosphorylation at residue 204 would result in steric clashes in the near vicinity of the adenine base, which would interfere with nucleotide binding.

Surprisingly, DDX3X(V168-G582) displays no ATPase activity in the presence of different single-stranded or partially double-stranded RNA substrates. In contrast, the full-length protein displays ATPase activity in the presence of RNA substrates.^{9,26} Similarly, several other recombinant DExD/H-box helicases show reduced ATPase/helicase activity when flanking sequences are removed. For example,

the ATPase activity of UAP56 is reduced by half by removing 46 residues from the N terminus that are specific to UAP56,¹¹ and the helicase activity of DP103 (DDX20) is dependent on C-terminal flanking sequences that are not part of the conserved helicase core.²⁸ In contrast, constructs of other DExD/H-box helicases comprising only the conserved helicase core display ATPase activity *in vitro*; e.g. the yeast splicing factor Prp22p.²⁹ It is not clear to us why some helicase cores display ATPase/helicase activity and others do not. What is becoming evident is that, although conserved domains 1 and 2 show high levels of sequence conservation and similar structural folds, much remains to be discovered about how the helicase core structure is linked to activity.

We find it most likely that the lack of ATPase activity in DDX3X(V168-G582) is due to missing N-terminal and/or C-terminal sequences. Since, when doing a search against the Pfam family database,³⁰ the flanking sequences contain no detectable domain architecture, it is more difficult to speculate how these additional sequences give specificity and/or activity. However, it is not unlikely that these represent natively disordered regions, since the first 130 and last 90 amino acid residues of DDX3X are predicted to be disordered by the protein dynamic disorder server DISOPRED.³¹ These so-called intrinsically unstructured regions could fold only upon interaction with other protein domains or RNA substrates.^{32,33}

As mentioned above, ATPases in general, including RNA and DNA helicases, have been proposed as an interesting protein family in terms of potential novel drug targets.³⁴ In particular, DDX3X has been identified as a possible antiviral drug target, since it is involved in the transport of HIV mRNAs from the nucleus to the cytoplasm.^{10,35} The three-dimensional structure presented here provides important information to strengthen a drug discovery program directed towards DDX3X. Putative drug-binding sites include the nucleotide-binding pocket and the substrate recognition site. As shown here, the nucleotide-binding pocket of DDX3X is highly similar to those of other DExD-box helicases and it is probably not trivial to obtain specific inhibition of a particular DExD-box helicase by targeting the nucleotide-binding pocket.³⁵ However, specific inhibitors have been developed that are directed towards the nucleotide-binding pocket of other ATPases, such as the kinases, and it may prove a successful way ahead for the DExD-box helicases. It should be investigated whether the propensity to bind mononucleotides by certain DExD-box helicases such as DDX3X(V168-G582) and domain 1 of the Hera helicase can be utilized in the design of nucleotide inhibitors. Another approach to inhibit a particular helicase such as DDX3X may be to target its unique sequences of potential importance for substrate recognition/activation, such as the DDX3X/Y-specific insertion described here. It is thus necessary to elucidate the structures and functions of these unique features to obtain more information on substrate interaction and helicase

activation, and to identify potential sites for drug intervention, for a given DExD-box helicase.

Materials and Methods

Cloning of the *ddx3x* gene

Eleven different constructs were amplified from the *ddx3x* gene (gi:33873845) and moved into the pET-based pNIC-Bsa4 vector by ligation-independent cloning. All constructs contained a removable His tag at the N terminus.

Expression of the *ddx3x* gene and purification of DDX3X protein

The plasmid constructs were transformed into the strain BL21(DE3) and tested on a small scale for expression before attempting large-scale expression and purification of protein. The large-scale culture, in Terrific Broth supplemented with 8 g/l of glycerol, grew at 37 °C. At an absorbance at 600 nm between 1 and 2 the temperature was lowered to 18 °C and recombinant protein production was induced by addition of isopropyl- β -D-thiogalactopyranoside (0.5 mM final concentration) and continued for 18 h at 18 °C.

Cells were harvested by centrifugation and the cell pellet was dissolved in buffer 1 (50 mM sodium phosphate (pH 7.5), 500 mM NaCl, 10% (v/v) glycerol, 0.5 mM Tris(2-carboxyethyl)-phosphine (TCEP)) with 10 mM imidazole and Complete EDTA-free protease inhibitor according to the manufacturer's instructions (Roche Biosciences). Before lysis, 20 units/ml of Benzonase (Novagen) was added and lysis was achieved by sonication. Cell debris was removed by centrifugation and the soluble fraction was filtered through a syringe filter (0.45 μ m pore size).

DDX3X protein was purified on an Äkta purifier system essentially as described,³⁶ with the following minor modifications: the lysate was loaded onto a 1 ml HisTrap FF crude column (Amersham Biosciences), non-bound protein was removed by washing with buffer 1 containing 25 mM imidazole and affinity-bound protein was eluted with buffer 1 containing 500 mM imidazole. Eluted protein was loaded onto a Superdex 200 16/60 gel-filtration column (Amersham Biosciences) in buffer 2 (20 mM Hepes (pH 7.5), 500 mM NaCl, 10% glycerol, 0.5 mM TCEP); DDX3X migrated as a monomer in this column.

The N-terminal His tag was removed by incubating 400 μ M DDX3X with a His-tagged TEV protease at a molar ratio of 1:50 for 18 h at 4 °C. After removal of TEV protease and the His tag, DDX3X was concentrated to 480 μ M, samples were flash-frozen in liquid nitrogen and stored at -80 °C. SDS-PAGE analysis showed that the protein was more than 90% pure (results not shown), and the correct mass of the different DDX3X constructs was verified by mass spectrometry (results not shown).

Thermal stability-shift assay

Changes in thermal stability due to ligand binding were measured by monitoring the intensity from the environmentally sensitive dye, SyproOrange (Molecular Probes, Eugene, OR) while heating the samples from 20–90 °C in a Bio-Rad iCycler RT PCR.²⁵ DDX3X(V168-G582) protein (4 μ M final concentration) was diluted in 50 mM Tris-HCl

(pH 8.0 at room temperature), 100 mM NaCl, 5 mM MgCl₂, 2 mM TCEP, before addition of nucleotides ATP, ADP, and AMP to a final concentration of 8 mM or ATP analogues ANP, AMPPCP, and ATPγS to a final concentration of 4 mM, and 1000-fold diluted SyproOrange. Fluorescence was measured from 20–89.6 °C and the thermal shift was determined as described.³⁷

ATPase assay

The ATPase assay was performed essentially as described,³⁸ with the following modifications: 0.1–1.4 μM DDX3X(V168-G582) was incubated with 1–10 μg of poly(U), poly(A), poly(C) or *in vitro* transcribed RNA mimicking stem-loop IIB in RRE (see below)³⁹ in 50 mM Tris-HCl (pH 7.5 at 37 °C) or 50 mM Hepes-KOH (pH 6.2 at 37 °C), 50 mM NaCl, 2 mM TCEP, 2 or 5 mM MgCl₂ and 0.5 mM ATP (Sigma Aldrich). In an attempt to obtain some activity, chloride was omitted from the reactions, the concentration of salt was varied between zero and 200 mM, and the assay was run also at pH 5 and pH 9. The reaction mixtures, total volume of 100 μl, were incubated at 37 °C for 1 h. A colorimetric assay based on the malachite green/molybdate reaction was used to detect the release of free phosphate.¹¹ A 200 μl portion of Biomol green reagent (Biomol International) was added to the reaction mixtures, and absorbance was measured at 620 nm after incubation for 25 min at room temperature. Alternatively, formation of (α-³²P)-labeled ADP/AMP was tested by incubating 0.8 μM DDX3X(V168-G582) in 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 5 mM MgCl₂, 1 mM DTT, with 1 μCi of [α-³²P]ATP (Amersham BioSciences) and 0.1 mM ATP (Sigma Aldrich). The reaction mixture, total volume of 20 μl, was incubated at 37 °C for 1 h and 3 μl of 0.5 M EDTA was added to stop the reaction. Possible formation of α-labeled AMP was monitored by loading 2 μl of the reaction mixtures onto PEI TLC plates (Macherey & Nagel), and chromatography was carried out in 1 M formic acid, 0.5 M lithium chloride for 3 h at room temperature. Radioactive nucleotides were visualized by autoradiography using a phosphoimager or Hyperfilm^{MP} (Amersham Bioscience). [α-³²P]AMP (Amersham Biosciences) was run as a control on the TLC plates.

In vitro transcription

The short RNA mimicking stem-loop IIB in RRE was transcribed with a Mega-Short transcription kit (Ambion) following the manufacturer's instructions. The two complementary oligonucleotides used as templates for *in vitro* transcription had the following sequence (transcribed region underlined): 5'TAATACGACTCACTATA GGGAA-CGGGCGCAGAAATCTGACGGTACGTTT3' and 5'AAA-CGTACCGTCAGATTCTGCGCCCGTTCCCTATAGT-GAGTCGTATTA3'.

Analysis of the purity of the ADP

The purity of the ADP used in the crystallization experiments was determined by HPLC using a Supelcosil LC-18-T, 5 μm column (25 cm×4.6 mm) and isocratic elution with 0.1 M KH₂PO₄ (pH 6.0) and a flow-rate of 1 ml/min. ADP (100 μl of 100 μM solution) eluted at 5.7 min and AMP (100 μl of 100 μM solution) eluted at 9.6 min. Equal volumes of ADP solution and the reservoir solution used in crystallization (100 mM Tris-HCl (pH 8.0), 1.75 M sodium formate) were mixed and incubated for

24 h at 6 °C before the nucleotide content was determined as described above. This analysis showed that a small amount of ADP was hydrolyzed to AMP in 24 h (2% corresponding to 400 μM).

Crystallization and data collection

DDX3X(V168-G582) crystals were grown using vapor diffusion at 4 °C by mixing equal amounts of protein solution (21 mg/ml or 446 μM) containing 20 mM ATPγS or 20 mM ADP and 20 mM MgCl₂ and reservoir solution containing 100 mM Tris-HCl (pH 8.0), 1.75 M sodium formate). Crystals appeared after one day and had reached their maximal size in 12 days (approximately 300 μm×100 μm×50 μm). Unsuccessful attempts were made to crystallize DDX3X(V168-G582) with ANP, AMPPCP and without nucleotide. As cryoprotectant, motherliquor supplemented with of 20% PEG 400, 10% glycerol and 300 mM NaCl was added to the drop, then crystals were harvested and flash-cooled in liquid nitrogen. The diffraction data were collected at beamline ID23-1 at the ESRF synchrotron facility in Grenoble, France. All data were indexed and integrated in space group P3₁21 with the XDS package.⁴⁰ The cell parameters are $a = b = 67.288 \text{ \AA}$, $c = 253.470 \text{ \AA}$, $\alpha = \beta = 90^\circ$, $\chi = 120^\circ$.

Structure solution and refinement

The structure was solved by molecular replacement using the two domains separately. The N-terminal RecA-like domain of DjVLGB (PDB entry 1WRB) was used as the search model for domain 1, and the ATP-dependent RNA helicase VASA (PDB entry 2DB3) was used for

Table 1. Data collection statistics from XDS and refinement

<i>A. Data collection</i>	
Beamline	ESRF ID23-1
Wavelength (Å)	1.072
Space group	P3 ₁ 21
Cell parameters:	
<i>a</i> (Å)	67.3
<i>b</i> (Å)	67.3
<i>c</i> (Å)	253.5
α (deg.)	90
β (deg.)	90
χ (deg.)	120
Resolution (Å)	15.0–2.20 (2.30–2.20)
No. unique reflections	32,841
Redundancy	10.6
Completeness (%)	99.2 (98.8)
<i>R</i> _{sym} (%)	7.2 (47.6)
<i>I</i> / σ	19.3 (5.7)
<i>B. Refinement statistics</i>	
Resolution (Å)	15.0–2.20
No. unique reflections	32,841
No. reflections in test set	1751
<i>R</i> _{work} (%)	18.1
<i>R</i> _{free} (%)	21.6
RMSD from ideal ^a	
Bond lengths (Å)	0.020
Bond angles (deg.)	1.74

Values in parentheses refer to values in the highest resolution shell. There was no residue in the disallowed regions of the Ramachandran plot, as calculated using PROCHECK.⁴⁴

^a Ideal values from.⁴⁵

domain 2. The asymmetric unit contains one protein monomer. Refmac was used for refinement,⁴¹ and Coot was used for model building.⁴² TLS refinement using four TLS groups as suggested by *tlsmd*† was used with Refmac.⁴³ In the refinement, data in the interval 15–2.20 Å resolution were used and the progress of refinement was followed by decreasing *R* and *R*_{free} values. At the end of the refinement, *R*=0.1887 and *R*_{free}=0.2182. The final model of DDX3X(v168-G582) starts at valine 168 and ends at tyrosine 580. Residues 407–410, 535–536 and 581–582 are disordered and are not visible in the electron density. A Ramachandran plot shows 98% of all residues in favored regions and no outlier in the disallowed regions. The data collection and refinement statistics are given in Table 1.

Protein Data Bank, accession code

Coordinates for the crystal structure have been deposited in the Protein Data Bank with accession code 2I4I.

Acknowledgements

We thank Vildan Dincbas Renqvist for help with HPLC analysis, Linda Svensson for *in vitro* transcription of RNA, Helena Berglund and Johan Weigelt for critical reading of the manuscript, and the European Synchrotron Radiation Facility and its staff for providing us with radiation time, help and guidance. R.-M.J. was supported by money from the Carl Tryggers Foundation awarded to L.H.S. The Structural Genomics Consortium is a registered charity (number 1097737) funded by the Karolinska Institutet, the Swedish Governmental Agency for Innovation Systems (VINNOVA), the Knut and Alice Wallenberg Foundation, the Swedish Foundation for Strategic Research, the Wellcome Trust, GlaxoSmithKline, Genome Canada, the Canadian Institutes of Health Research, the Ontario Innovation Trust, and the Ontario Research and Development Challenge Foundation for Innovation.

References

- Cordin, O., Banroques, J., Tanner, N. K. & Linder, P. (2006). The DEAD-box protein family of RNA helicases. *Gene*, **367**, 17–37.
- Abdelhaleem, M. (2005). RNA helicases: regulators of differentiation. *Clin. Biochem.* **38**, 499–503.
- Foresta, C., Ferlin, A. & Moro, E. (2000). Deletion and expression analysis of AZFa genes on the human Y chromosome revealed a major role for DBY in male infertility. *Hum. Mol. Genet.* **9**, 1161–1169.
- Ditton, H. J., Zimmer, J., Kamp, C., Rajpert-De Meyts, E. & Vogt, P. H. (2004). The AZFa gene DBY (DDX3Y) is widely transcribed but the protein is limited to the male germ cells by translation control. *Hum. Mol. Genet.* **13**, 2333–2341.
- Chang, P. C., Chi, C. W., Chau, G. Y., Li, F. Y., Tsai, Y. H., Wu, J. C. & Wu Lee, Y. H. (2006). DDX3, a DEAD box RNA helicase, is deregulated in hepatitis virus-associated hepatocellular carcinoma and is involved in cell growth control. *Oncogene*, **25**, 1991–2003.
- Kanai, Y., Dohmae, N. & Hirokawa, N. (2004). Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. *Neuron*, **43**, 513–525.
- de la Cruz, J., Iost, I., Kressler, D. & Linder, P. (1997). The p20 and Ded1 proteins have antagonistic roles in eIF4E-dependent translation in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **94**, 5201–5206.
- Chuang, R. Y., Weaver, P. L., Liu, Z. & Chang, T. H. (1997). Requirement of the DEAD-Box protein ded1p for messenger RNA translation. *Science*, **275**, 1468–1471.
- Yedavalli, V. S., Neuveut, C., Chi, Y. H., Kleiman, L. & Jeang, K. T. (2004). Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. *Cell*, **119**, 381–392.
- Kwong, A. D., Rao, B. G. & Jeang, K. T. (2005). Viral and cellular RNA helicases as antiviral targets. *Nature Rev. Drug Discov.* **4**, 845–853.
- Shi, H., Cordin, O., Minder, C. M., Linder, P. & Xu, R. M. (2004). Crystal structure of the human ATP-dependent splicing and export factor UAP56. *Proc. Natl Acad. Sci. USA*, **101**, 17628–17633.
- Andersen, C. B., Ballut, L., Johansen, J. S., Chamieh, H., Nielsen, K. H., Oliveira, C. L. *et al.* (2006). Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science*, **313**, 1968–1972.
- Bono, F., Ebert, J., Lorentzen, E. & Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell*, **126**, 713–725.
- Cheng, Z., Coller, J., Parker, R. & Song, H. (2005). Crystal structure and functional analysis of DEAD-box protein Dhh1p. *RNA*, **11**, 1258–1270.
- Story, R. M., Li, H. & Abelson, J. N. (2001). Crystal structure of a DEAD box protein from the hyperthermophile *Methanococcus jannaschii*. *Proc. Natl Acad. Sci. USA*, **98**, 1465–1470.
- Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. (2006). Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* VASA. *Cell*, **125**, 287–300.
- Caruthers, J. M., Johnson, E. R. & McKay, D. B. (2000). Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc. Natl Acad. Sci. USA*, **97**, 13080–13085.
- Johnson, E. R. & McKay, D. B. (1999). Crystallographic structure of the amino terminal domain of yeast initiation factor 4A, a representative DEAD-box RNA helicase. *RNA*, **5**, 1526–1534.
- 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.
- Zhao, R., Shen, J., Green, M. R., MacMorris, M. & Blumenthal, T. (2004). Crystal structure of UAP56, a DEXD/H-box protein involved in pre-mRNA splicing and mRNA export. *Structure*, **12**, 1373–1381.
- Kurimoto, K., Muto, Y., Obayashi, N., Terada, T., Shirouzu, M., Yabuki, T. *et al.* (2005). Crystal structure of the N-terminal RecA-like domain of a DEAD-box RNA helicase, the *Dugesia japonica* vasa-like gene B protein. *J. Struct. Biol.* **150**, 58–68.

† <http://skuld.bmsc.washington.edu/~tlsmd/>

22. Carmel, A. B. & Matthews, B. W. (2004). Crystal structure of the BstDEAD N-terminal domain: a novel DEAD protein from *Bacillus stearothermophilus*. *RNA*, **10**, 66–74.
23. Rudolph, M. G., Heissmann, R., Wittmann, J. G. & Klostermeier, D. (2006). Crystal structure and nucleotide binding of the *Thermus thermophilus* RNA helicase Hera N-terminal domain. *J. Mol. Biol.* **361**, 731–743.
24. Linder, P. & Lasko, P. (2006). Bent out of shape: RNA unwinding by the DEAD-box helicase Vasa. *Cell*, **125**, 219–221.
25. Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E. *et al.* (2001). High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J. Biomol. Screen*, **6**, 429–440.
26. Franca, R., Belfiore, A., Spadari, S. & Maga, G. (2007). Human DEAD-box ATPase DDX3 shows a relaxed nucleoside substrate specificity. *Proteins: Struct. Funct. Genet.* **67**, 1128–1137.
27. Sekiguchi, T., Kurihara, Y. & Fukumura, J. (2007). Phosphorylation of threonine 204 of DEAD-box RNA helicase DDX3 by cyclin B/cdc2 in vitro. *Biochem. Biophys. Res. Commun.* **356**, 668–673.
28. Yan, X., Mouillet, J. F., Ou, Q. & Sadovsky, Y. (2003). A novel domain within the DEAD-box protein DP103 is essential for transcriptional repression and helicase activity. *Mol. Cell. Biol.* **23**, 414–423.
29. Schneider, S. & Schwer, B. (2001). Functional domains of the yeast splicing factor Prp22p. *J. Biol. Chem.* **276**, 21184–21191.
30. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S. *et al.* (2004). The Pfam protein families database. *Nucl. Acids Res.* **32**, D138–D141.
31. Bryson, K., McGuffin, L. J., Marsden, R. L., Ward, J. J., Sodhi, J. S. & Jones, D. T. (2005). Protein structure prediction servers at University College London. *Nucl. Acids Res.* **33**, W36–W38.
32. Wright, P. E. & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331.
33. Dyson, H. J. & Wright, P. E. (2002). Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12**, 54–60.
34. Chene, P. (2003). The ATPases: a new family for a family-based drug design approach. *Expert Opin. Ther. Targets*, **7**, 453–461.
35. Jeang, K. T. & Yedavalli, V. (2006). Role of RNA helicases in HIV-1 replication. *Nucl. Acids Res.* **34**, 4198–4205.
36. Kursula, P., Flodin, S., Ehn, M., Hammarstrom, M., Schuler, H., Nordlund, P. & Stenmark, P. (2006). Structure of the synthetase domain of human CTP synthetase, a target for anticancer therapy. *Acta Crystallog. sect. F*, **62**, 613–617.
37. Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. & Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357**, 289–298.
38. Fuller-Pace, F. V., Nicol, S. M., Reid, A. D. & Lane, D. P. (1993). DbpA: a DEAD box protein specifically activated by 23s rRNA. *EMBO J.* **12**, 3619–3626.
39. Ippolito, J. A. & Steitz, T. A. (2000). The structure of the HIV-1 RRE high affinity rev binding site at 1.6 Å resolution. *J. Mol. Biol.* **295**, 711–717.
40. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constant. *J. Appl. Crystallog.* **26**, 795–800.
41. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
42. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallog. sect. D*, **60**, 2126–2132.
43. Painter, J. & Merritt, E. A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallog. sect. D*, **62**, 439–450.
44. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291.
45. Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst. sect. A*, **47**, 392–400.

Edited by J. Doudna

(Received 12 April 2007; received in revised form 12 June 2007; accepted 14 June 2007)
Available online 26 June 2007