doi:10.1016/j.jmb.2010.05.046 J. Mol. Biol. (2010) 400, 768–782

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Crystal Structure of Human RNA Helicase A (DHX9): Structural Basis for Unselective Nucleotide Base Binding in a DEAD-Box Variant Protein

Patrick Schütz, Elisabet Wahlberg, Tobias Karlberg, Martin Hammarström, Ruairi Collins, Alex Flores and Herwig Schüler*

Structural Genomics Consortium, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden

Received 4 February 2010; received in revised form 4 May 2010; accepted 19 May 2010
Available online 25 May 2010

RNA helicases of the DExD/H-box superfamily are critically involved in all RNA-related processes. No crystal structures of human DExH-box domains had been determined previously, and their structures were difficult to predict owing to the low level of homology among DExH-motif-containing proteins from diverse species. Here we present the crystal structures of the conserved domain 1 of the DEIH-motif-containing helicase DHX9 and of the DEAD-box helicase DDX20. Both contain a RecA-like core, but DHX9 differs from DEAD-box proteins in the arrangement of secondary structural elements and is more similar to viral helicases such as NS3. The N-terminus of the DHX9 core contains two long α-helices that reside on the surface of the core without contributing to nucleotide binding. The RNA-polymerase-II-interacting minimal transactivation domain sequence forms an extended loop structure that resides in a hydrophobic groove on the surface of the DEIH domain. DHX9 lacks base-selective contacts and forms an unspecific but important stacking interaction with the base of the bound nucleotide, and our biochemical analysis confirms that the protein can hydrolyze ATP, guanosine 5′-triphosphate, cytidine 5′-triphosphate, and uridine 5′-triphosphate. Together, these findings allow the localization of functional motifs within the three-dimensional structure of a human DEIH helicase and show how these enzymes can bind nucleotide with high affinity in the absence of a Q-motif.

Keywords: RNA helicase; RecA; DExD/H-box; nucleotide binding; X-ray crystallography

Edited by J. Doudna

Introduction

Ribonucleic acids, apart from storing genetic information, have multifaceted structural and catalytic functions in numerous vital processes. RNA folding and protein binding are essential aspects of RNA function.[1–3] RNA helicases remodel RNA and ribonucleoprotein (RNP) complexes, facilitate RNA translocation among cellular compartments, and mediate recognition of the RNAs of intruding viral pathogens. RNA helicases are divided into six superfamilies; those best described belong to helicase superfamily 2, which also includes DNA helicase enzymes.[4] Conserved sequence motifs are found in two RecA-like domains (conserved domains 1 and 2) that form the RNA and nucleotide binding core of all RNA helicases. Despite the conservation of motifs, RecA-like domains differ considerably in fold and function. Superfamily 2 helicases contain a characteristic DExD/H sequence motif. DEAD-box enzymes usually catalyze local unwinding of RNA structures, whereas DExH enzymes often translocate along RNA strands.[5] Members of both groups can displace and remodel paired RNA strands or RNA–protein interactions.[6] Most RNA helicases contain additional domains on either or both sides of the core domains, which mediate specific interactions or localization.

*Corresponding author. E-mail address: hersch@ki.se.

Abbreviations used: RNP, ribonucleoprotein; GTP, guanosine 5′-triphosphate; CTP, cytidine 5′-triphosphate; UTP, uridine 5′-triphosphate; PolII, RNA polymerase II; MTAD, minimal transactivation domain; PDB, Protein Data Bank; ITC, isothermal titration calorimetry; TEV, tobacco etch virus; TCEP, tris(2-carboxyethyl)phosphine; PEG, polyethylene glycol.

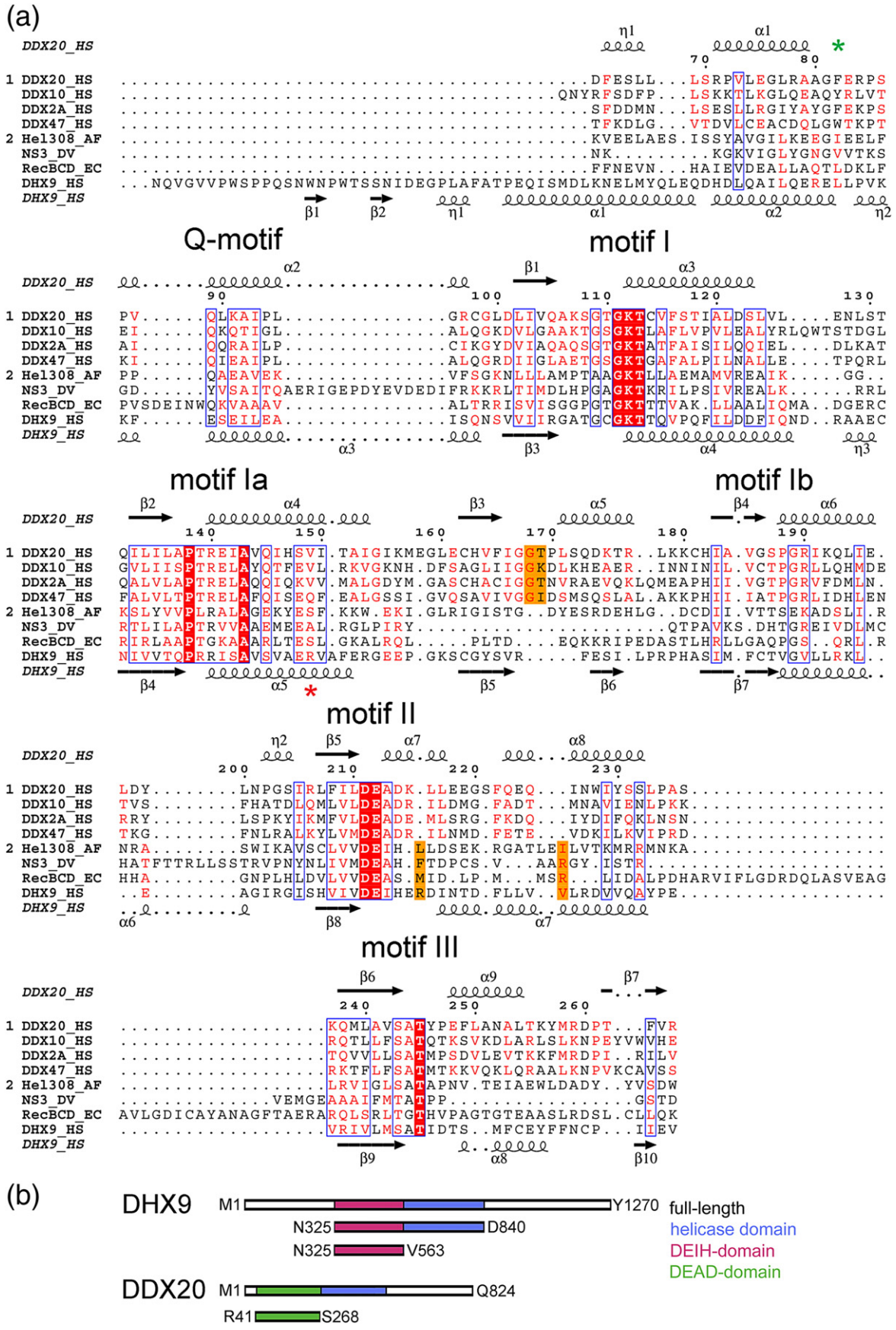

Fig. 1 (*legend on next page*)

Table 1. Crystal parameters, data collection, and refinement statistics

Structure	DHX9			DDX20
PDB entry		3LLM		2OXC
Ligand		MnADP		ADP
Data set	Peak	Inflection	Remote	Native
Beamline	DIAMOND I03	DIAMOND I03	DIAMOND I03	European Synchrotron Radiation Facility ID 14-2
Wavelength (Å)	0.98000	0.98020	0.96860	0.93300
Space group	$P3_221$	$P3_221$	$P3_221$	$P3_121$
Cell dimensions				
a, *b*, *c* (Å)	113.70, 113.70, 141.65	113.70, 113.70, 141.65	113.70, 113.70, 141.65	63.80, 63.80, 214.30
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120
Resolution (Å)	98.5–2.8 (2.95–2.8)	98.5–2.8 (2.95–2.8)	98.5–2.8 (2.95–2.8)	38.5–1.3 (1.5–1.3)
R_{sym}	0.099 (1.001)	0.105 (1.143)	0.094 (0.844)	0.057 (0.233)
$I/(\sigma,I)$	14.1 (2.4)	13.5 (2.2)	14.7 (2.8)	23.5 (5.2)
Completeness (%)	100 (100)	100 (100)	99.9 (100)	100 (99.9)
Redundancy	11 (11)	11 (11)	11 (11)	10 (11)
Refinement				
Resolution range (Å)		98.5–2.8		38.5–1.3
Number of reflections		25,255		119,028
R_{work}[a]/R_{free}[b]		21.0/23.8		15.4/17.5
Number of atoms				
Protein		3679		3227
Ligands		67		54
Water		27		668
B-factors (Å^2)				
Protein		65		12
Ligands		76		16
Water		52		29
RMSD				
Bond lengths (Å)		0.015		0.01
Bond angles (°)		1.457		1.443
Ramachandran plot				
Favored regions (%)		98.07		99.1
Allowed regions (%)		100		100

Numbers in parentheses refer to the highest-resolution shell.

[a] R_{work} is defined as $\sum||F_{obs}|-|F_{calc}||/\sum|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

[b] R_{free} is the *R*-factor for the test set (5–10% of the data).

DHX9 (also known as RNA helicase A or *maleless*) binds single-stranded RNA and single-stranded DNA, and unwinds double-stranded nucleic acids driven by hydrolysis of ATP, guanosine 5′-triphosphate (GTP), cytidine 5′-triphosphate (CTP), or uridine 5′-triphosphate (UTP).[7,8] It links the transcription and translation of both cellular and retroviral mRNAs.[9,10] DHX9 can interact with several transcription regulators.[11] It mediates the association of CREB (*c*AMP-*r*esponsive *e*lement *b*inding)-binding protein with RNA polymerase II (PolII)[12] and acts as an RNA loading factor in the RNA-induced silencing complex.[13] Many of these interactions are mediated by different sections on the N-terminal side of the RecA-like domains.[11,14,15] Interaction with PolII is mediated by the minimal transactivation domain (MTAD) immediately before the DEIH domain.[16] Understanding DHX9 function is of explicit medical importance owing to its role in the replication of various viruses such as human immunodeficiency virus and hepatitis C virus.[17,18]

DDX20 (also known as Gemin3 or DP103)[11] is a component of the SMN (*s*urvival of *m*otor *n*eurons) complex that is involved in the assembly and reconstruction of RNP complexes.[19,20] Interaction with SMN resides in the C-terminus of DDX20. DDX20, Gemin4, and eIF2C2 form a separate complex that contains numerous micro RNAs.[21] DDX20 also binds to the Epstein–Barr virus nuclear proteins EBNA2 and EBNA3C.[22] DDX20 cleavage by the poliovirus-encoded proteinase 2Apro results in DDX20 inactivation and reduced small nuclear RNP assembly.[23]

Here we present the first crystal structure of a human DEIH-motif-containing helicase, the structure of conserved domain 1 of DHX9. Our results illustrate the nature of the PolII-interacting MTAD sequence, allow the localization of functional motifs, and provide the structural basis for the lack of nucleotide selectivity within this subclass of helicases. The

Fig. 1. (a) Sequence alignment of conserved domains 1 of selected DEAD-box and DExH-box enzymes. Secondary structural information is given for DHX9 above the alignment and for DDX20 below the alignment. The positions of conserved sequence motifs are indicated. Green asterisk, base-stacking aromatic side chain in DEAD-box proteins; red asterisk, base-stacking Arg456 in DHX9. Sequence database entry codes are given in Materials and Methods. (b) Schematic drawing of the DHX9 and DDX20 constructs used in this study.

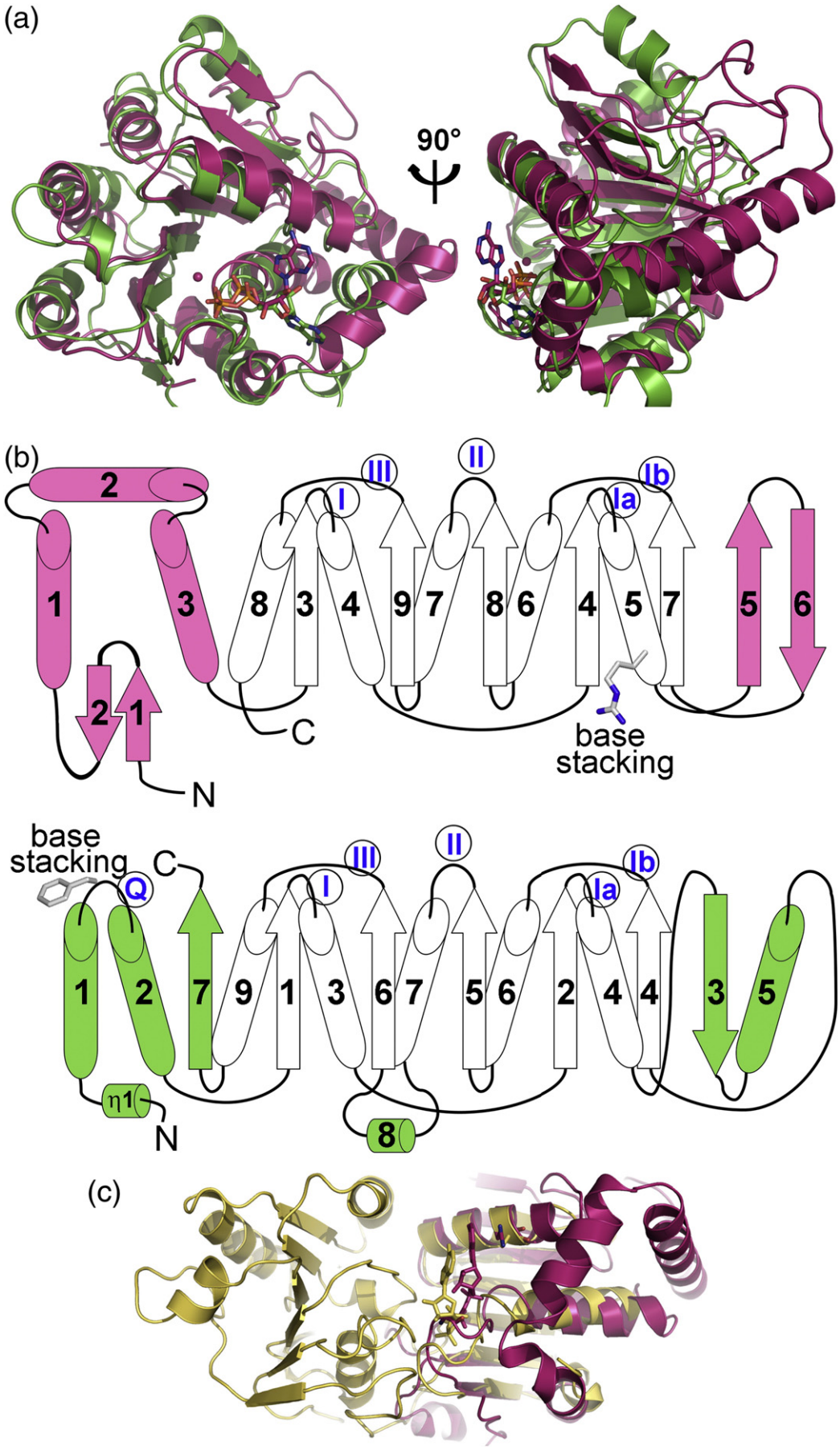

Fig. 2 (*legend on next page*)

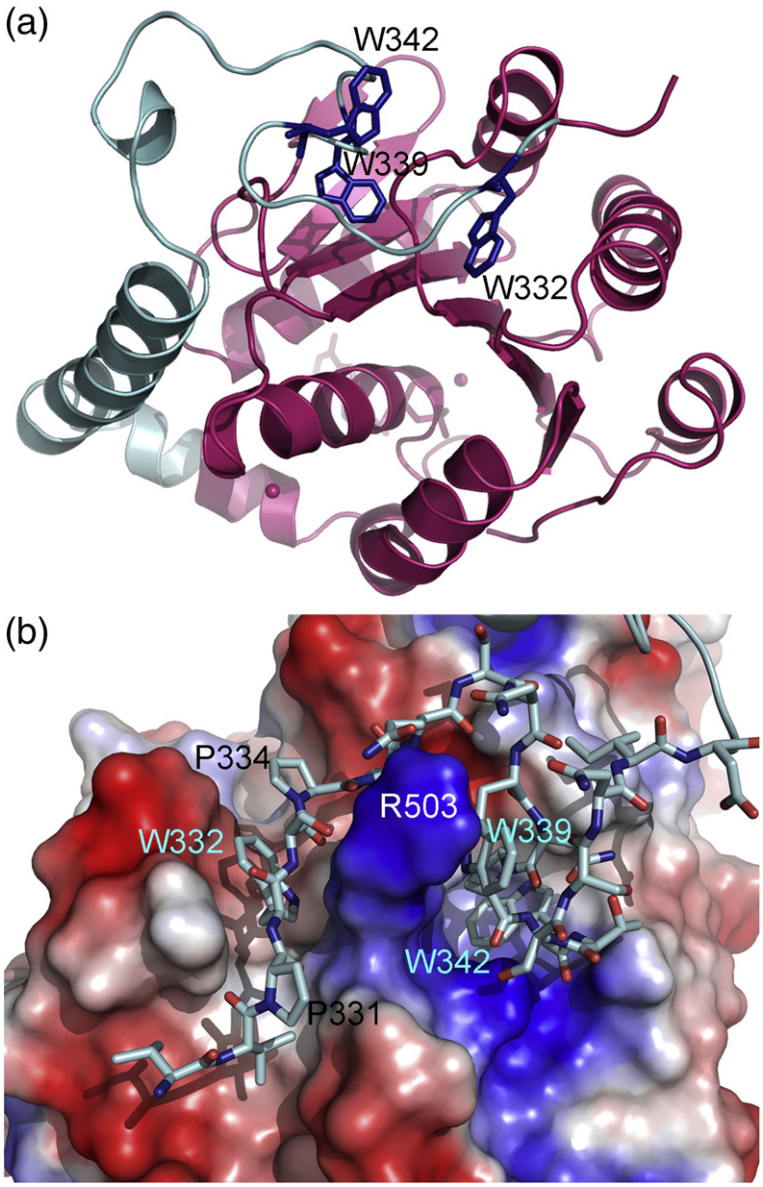

Fig. 3. Structure of MTAD. (a) Cartoon representation of the DHX9 structure, with the conserved domain shown in pink and with the N-terminal extension shown in blue. The three tryptophans that are essential for DHX9 interaction with PolII are shown as sticks. (b) Electrostatic surface representation of the core of the protein illustrating the hydrophobic groove that accommodates the MTAD.

structure suggests that the human DEIH motif helicases are most similar to viral RNA helicase NS3. We also present a structure of the corresponding domain of the human DEAD-box helicase DDX20. Structural and functional comparisons of the two structures contribute to our understanding of these medically important proteins.

Results

We used a structural genomics platform to study human RNA helicases by X-ray crystallography. We set out to determine the structural features that divide the human DEAD-box helicases from the closely related proteins containing a DExH motif. DHX9/RNA helicase A was targeted because it is a DEIH-box protein of considerable interest, with well-characterized nonhuman orthologs. As an example for a DEAD-box domain, we chose to include DDX20 in this study. We had already solved crystal structures of DDX20, and it was available for biochemical analysis. Both proteins are transcription regulators of medical importance.[11]

The borders of the conserved core domains contained in DDX20 and many of the well-

Fig. 2. Overview of the crystal structures of human DHX9 and DDX20. (a) Structural alignment of DHX9 (pink) and DDX20 (green). Note the differences in secondary structure arrangement outside the core domains. (b) Topology cartoons of DHX9 (upper diagram) and DDX20 (lower diagram). The conserved RecA-like domain is shown in white, and the unique structural elements are shown in color. Secondary structural elements are indicated in Arabic numerals, and sequence motifs are indicated in Roman numerals as in Fig. 1. The positions of the base-stacking residues are indicated. (c) Superposition of the structures of DHX9 (pink) and dengue virus helicase NS3 (yellow). Note the similar secondary structure arrangement in the DEIH domains.

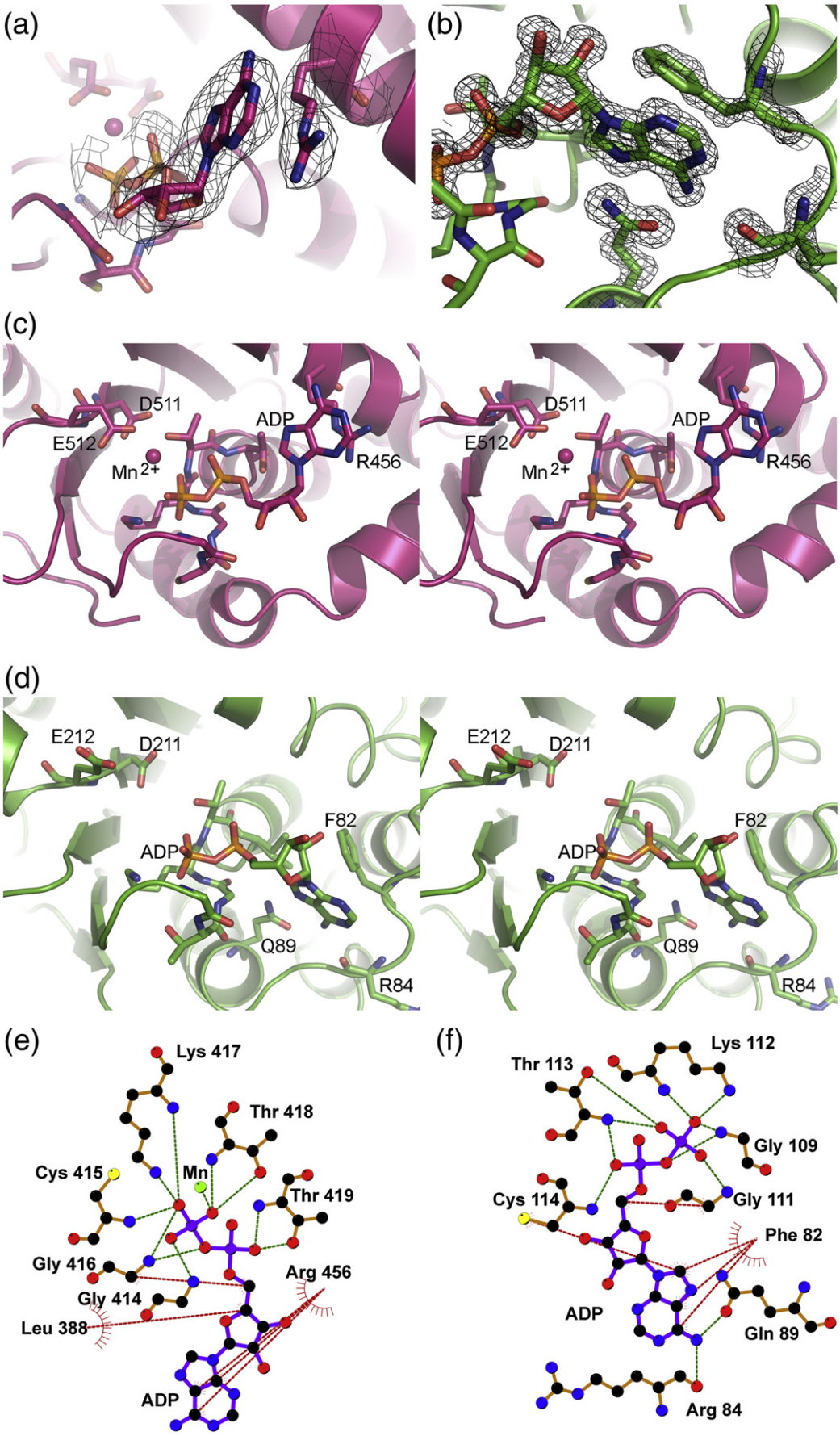

Fig. 4 (*legend on next page*)

characterized RecA-like proteins, as well as some of the conserved sequence motifs, are not readily recognized in DHX9 and other DEAD-box variant proteins (Fig. 1a). We used a multiconstruct approach[24] to test for the soluble expression of protein constructs that were designed based on sequence alignments, secondary structure predictions, and structural homology searches. For both DHX9 and DDX20, we identified protein constructs, spanning either one or both conserved core domains, which were straightforward to produce in a standard *Escherichia coli* expression system (Fig. 1b). To select protein constructs and buffer conditions for crystallography, we employed thermal stability assays.[25]

DHX9 overall structure

We obtained crystals of DHX9$^{325–563}$ in the presence of ADP and Mn^{2+}, and collected native diffraction data. The structure could not be solved by molecular replacement owing to lack of a suitable model. The DHX9 crystals were then reproduced using selenomethionine-labeled protein, and the structure was solved by selenium-based multiwavelength anomalous dispersion phasing and refined to 2.8 Å (Table 1). The final model comprised residues Gly328-Val563 and could be subdivided into three sections (Figs. 2 and 3). (i) The first ~30 residues make up the MTAD of human DHX9. They form an extended loop that is discussed in Structure of MTAD. (ii) The following ~45-residue segment preceding the N-terminus of the DEIH domain constitutes α-helices 1–3 that arrange perpendicular to each other on the surface of the α/β core (Fig. 2a). (iii) The remainder of the protein has a RecA-like α/β topology core that is typical of DExD/H helicases. However, the conserved central β-sheet is extended by two strands (β-strands 5 and 6). All β-strands run parallel, with the exception of β-strand 6 (Fig. 2b).

This arrangement of secondary structural elements in DHX9 is clearly different from that of known DEAD-box domains, explaining why we were unable to find a suitable model structure for molecular replacement. A 3D-Blast search carried out with BioXGEM[26] identified three structures with similar secondary structural elements, but none with equal arrangement of secondary structural elements: The recently determined DEAH domain of yeast Prp43p[27] [Protein Data Bank (PDB) entry 3KX2] has the greatest similarity (RMSD of 2.0 Å, 37% sequence identity), followed by the DEIH domain of the *Archaeoglobus fulgidus* DNA helicase Hel308[28] (PDB entry 2P6U; RMSD of 2.47 Å, 15% sequence identity) and the DECH domain of the flavivirus NS3[29] (PDB entry 2JLR; RMSD of 2.46 Å, 17% sequence identity) and of the *E. coli* SF1 helicase RecBCD[30] (PDB entry 1W36; RMSD of 3.04 Å, 15% sequence identity). None of these structures contains an N-terminus that folds as in our DHX9 structure, which is the first crystal structure of a human DEIH motif RNA helicase (Fig. 2c).

Structure of MTAD

At the N-terminus of the DEIH domain, the first 30 residues that could be resolved in the DHX9 structure constitute the MTAD that facilitates interaction with PolII. The MTAD forms a coil that contains two short β-strands and meanders across the surface of the α/β core (Fig. 3a). The section is highly hydrophobic and contains no charged side chains. Eighteen residues of the N-terminal coil interact with 29 residues of the DHX9 core mainly through hydrophobic interactions; 10 hydrogen bounds and no salt bridge are formed between them. The residues that bury the largest surface are Val330 (103 Å^2), Trp332 (181 Å^2), Pro335 (105 Å^2), Trp339 (65 Å^2), Pro341 (79 Å^2), and Trp342 (212 Å^2). Together, they make up 745 Å^2 of the total 973 Å^2 of the interface between the core and the N-terminus. This surface binding mode of the MTAD can be illustrated by electrostatic surface rendering of the core surface, which shows how the proline and tryptophan side chains reach into hydrophobic pockets (Fig. 3b). No crystal contacts are formed in this region.

Overall structure of DDX20

We crystallized human DDX20$^{41–268}$ and solved structures of the DDX20 DEAD domain in complex with bound ADP (Table 1). The final model comprised residues Ala62-Leu266, and the structure was refined to 1.3 Å. It shows the DEAD domain α/β topology, and the seven β-strands all run parallel in the order 7-1-6-5-2-4-3 (Fig. 2a and b). The N-terminus folds into a bundle of three nearly parallel α-helices. The loop between α-helix 1 and α-helix 2 contains the Q-motif and the phenylalanine that stacks with the nucleotide. The overall fold of this domain is similar to those of other DEAD domains, and the closest structural homolog is yeast eIF4A (PDB entry 1FUU; *Z*-score of 7.2, RMSD of 1.31 Å, 38.4% sequence identity).

Comparison of the nucleotide binding sites of DHX9 and DDX20

Both proteins bind nucleotide at a corresponding surface location, and nucleotide-buried solvent-accessible surface areas are very similar in both

Fig. 4. Details of the nucleotide binding sites and nucleotide-interacting motifs in DHX9 and DDX20. (a) DHX9: Example of the electron density ($2F_{obs}-F_{calc}$) for ADP and the Arg456 side chain, rendered at 1σ. (b) DDX20: Example of the electron density ($2F_{obs}-F_{calc}$) for ADP and selected side chains, rendered at 1.5σ. (c) Stereo view of the nucleotide binding site of DHX9. (d) Stereo view of the nucleotide binding site of DDX20. (e and f) Schematic of the interactions of DHX9 (e) and DDX20 (f) with bound ADP.

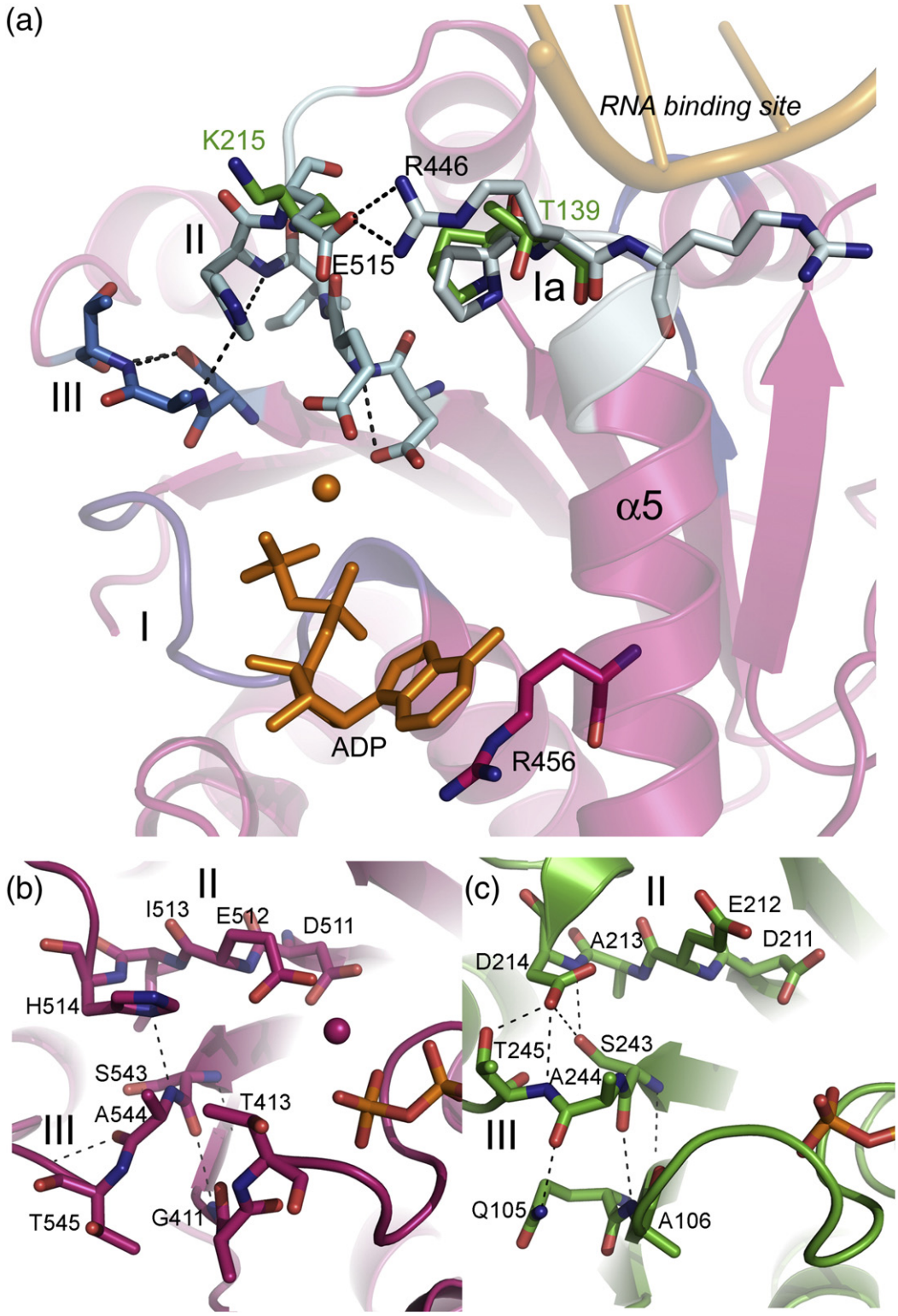

Fig. 5. Structure of the conserved DExD/H family sequence motifs in human DHX9. (a) Surroundings of the conserved motifs of DHX9. Selected DHX9 residues are shown as blue sticks. The DDX20 residues that align with the salt bridge forming DHX9–Lys215 and DHX9–Arg446 are shown as blue sticks. Base-stacking DHX9–Arg456 on α-helix 5 is shown as pink sticks. Motifs are numbered, and the position of the RNA binding site is indicated by a single-stranded RNA of the superposed DDX19–RNA complex (PDB entry 3G0H). (b) Details of the interactions of motifs I, II, and III of DHX9. (c) Details of the interactions of motifs I, II, and III of DDX20.

structures (214.3 Å^2 for DHX9 and 212.2 Å^2 for DDX20). However, the ribose and nucleotide base moieties are in entirely different orientations (Fig. 4). In DHX9, the orientation of the nucleotide ribose and base is determined by a stacking interaction of the purine rings with Arg456, located on α-helix 5

(Fig. 4a, c, and e). In other known DExH protein structures, stacking interactions with arginine, stacking interactions with tyrosine, or no side chains at all are observed, and the orientations of the nucleotide base vary (Supplementary Data, Fig. S1).[27,29,31] No other interactions between the nucleoside and the protein are formed. By contrast, in DDX20, the nucleotide base stacks with Phe82, located in the loop between α-helix 2 and α-helix 3, and this position is occupied by an aromatic side chain also in most other DEAD-box domains,[32] but may be aliphatic.[33] Further DDX20 side chains of the Q-motif contribute to the orientation of the nucleotide base: Gln89 hydrogen bonds with the ring nitrogens N6 and N7, and the backbone oxygen of Arg84 also bonds with N6 (Fig. 4b, d, and f). The Q-motif is not present in DHX9 and homologous proteins.

DHX9 coordinates nucleotide phosphates by the P-loop, as in DDX20 and the known DEAD-box domains (Fig. 4). α-Phosphate is coordinated by DHX9–Lys417, Thr418, and Thr419, and β-phosphate is coordinated by Gly414, Lys417, Gly416, and Thr418. Both phosphates are further coordinated by the manganese ion, which is tethered by the Glu512 side chain of the DEIH-box. The *B*-factors in the nucleotide-coordinating N-terminal part of DHX9 (α-helix 1, α-helix 2, and the loop between α-helix 2 and α-helix 3) are higher than the average over the entire structure, showing that this region has greater flexibility. In the DDX20–ADP complex, α-phosphate is coordinated by Lys112, Thr113, and Cys114, and β-phosphate is coordinated by Gly109, Lys112, Gly111, and Thr113. No divalent cation was found in the nucleotide binding site.

Conserved motifs

The conserved sequence motifs shared by all DExD/H helicases locate to five central α/β repeats, where they occupy the loops between the C-terminal end of each β-sheet and the beginning of each α-helix (Fig. 1a). These are also regions that superimpose best when the DHX9 and DDX20 structures are aligned. Motif I (the P-loop) coordinates nucleotide phosphates; motifs Ia and Ib contribute to the positively charged RNA binding site; and motif II coordinates the divalent cation. All five motifs are in close proximity to each other or linked to each other: motif III (the SAT motif) connects motif I to motif II, and motifs Ia and Ib are connected. Individual family members have different secondary elements inserted into the α–β core (β-sheets 5 and 6 in DHX9; α-helix 5 and β-sheet 3 in DDX20), but these contain no conserved sequence motifs. The loop regions between these secondary structure elements are occupied by positively charged side chains and border to the RNA binding motifs Ia and Ib.

Conserved motif II contains the sequences that give the RecA-like helicase subfamilies their designations. Our structural analysis shows that the DEIH motif of DHX9 and the DEAD motif of DDX20 both function to connect motif Ia to the SAT motif, but this is achieved by fundamentally different interactions in the two proteins (Fig. 5). In DDX20 and other DEAD-box proteins, the C-terminal aspartic acid of the DEAD motif interacts with the serine and threonine of the SAT motif. A hydrogen-bonding network is formed through the aspartic acid side chain and backbone nitrogen to the serine and threonine alcohol groups and the threonine backbone nitrogen. By contrast, in DHX9, the histidine of the DEIH motif interacts with the alanine of the SAT motif by a hydrogen bond between the histidine N^{δ} and the alanine backbone amide, and by van der Waals contact between the imidazole C2 carbon and the alanine side chain. Both side chains are strictly conserved in all human DExH proteins (Fig. 6).

Nucleotide base specificity of DHX9

Drosophila DHX9 can use several nucleotides for enzymatic activity.[7,8] We conducted NTPase measurements with a human DHX9 construct that spanned the DEIH and helicase domains ($DHX9^{325–840}$). ATP, GTP, CTP, and UTP were all suitable substrates for this protein construct, and we determined an ATPase rate of 1.5 s^{-1} (Fig. 7a).

Our adenine nucleotide crystal complexes of the conserved domains 1 of DXH9 and DDX20 show that these protein fragments can bind nucleotide in the absence of the helicase domains. Furthermore, com-

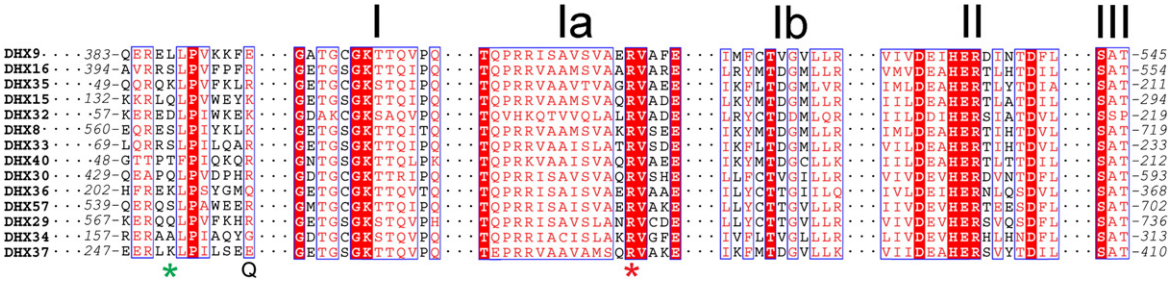

Fig. 6. Alignment of the generally conserved sequence motifs of human DExH proteins. Green asterisks, positions of base-stacking aromatic side chain in DEAD-box proteins; red asterisk, position of the base-stacking Arg456 in DHX9. "Q" denotes the position of the conserved glutamate in DEAD-box proteins. Sequence database entry codes are given in Materials and Methods.

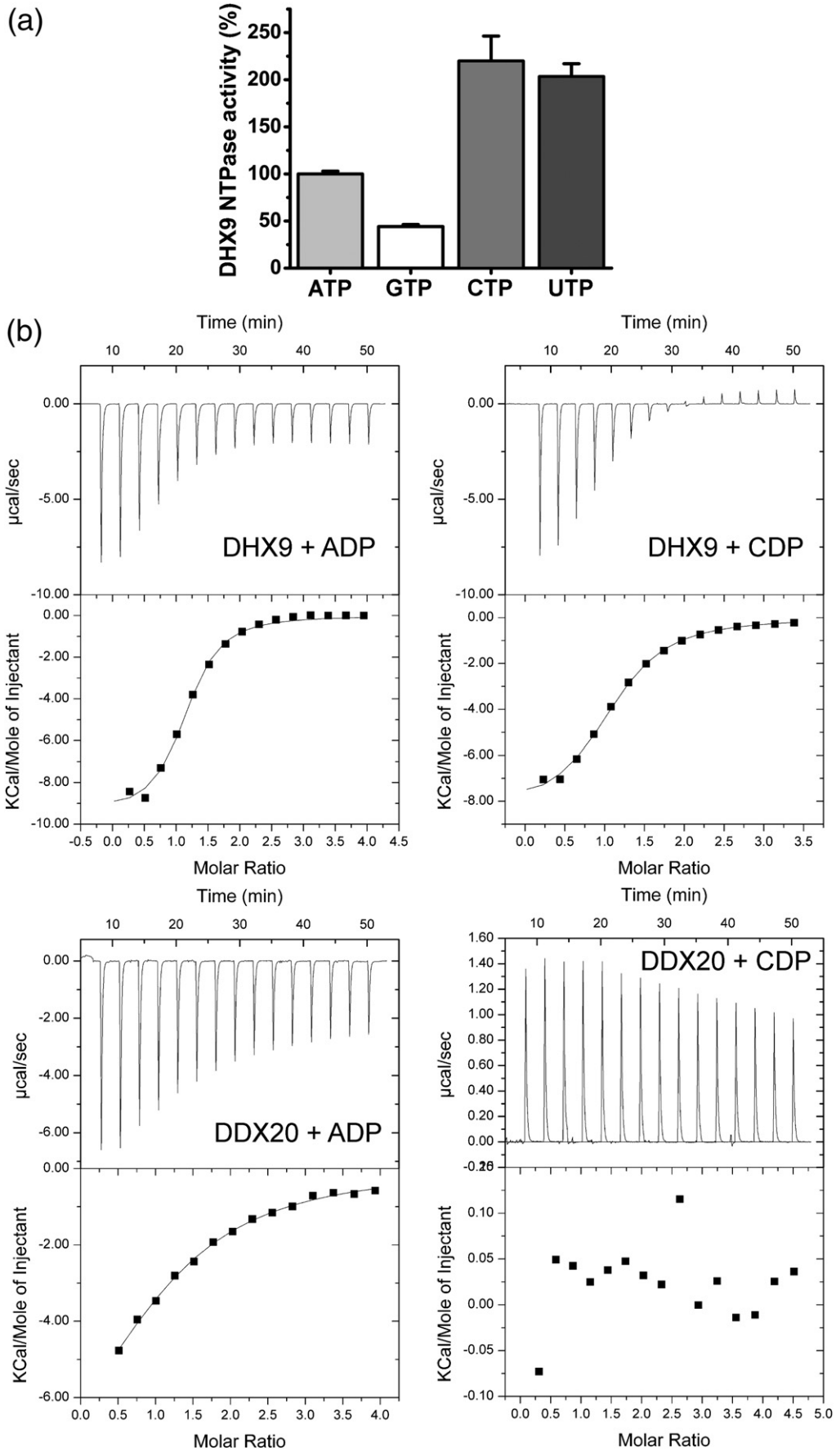

Fig. 7 (*legend on next page*)

Table 2. Nucleotide diphosphate binding of DHX9 and DDX20, as assessed by ITC

	Protein							
	$DHX9^{325–563}$				$DDX20^{41–268}$			
	ADP	GDP	CDP	UDP	ADP	GDP	CDP	UDP
K_d^{app} (μM)	18 [16, 21]	26 [22, 30]	58 [54, 63]	58 [53, 64]	230 [210, 260]	NB	NB	NB
ΔH_{app} (kcal/mol)	−9.5±0.3	−10.0±0.2	−8.4±0.2	−7.7±0.2	−9.8±0.9			
N	1.11±0.02	1.09±0.02	1.08±0.02	1.07±0.02	1.11±0.08			

Values in brackets represent the lower and upper bounds of the K_d value. Data were deduced from one experiment under each condition. NB, no detectable binding.

parison of side -chain interactions with the nucleotide base in the two crystal structures suggested that DHX9 may be capable of binding different nucleotides. To evaluate nucleotide binding selectivity by the DEIH domain, we conducted isothermal titration calorimetry (ITC) experiments, where corresponding constructs of DHX9 and DDX20 were titrated with pyrimidine and purine nucleotide diphosphates. We found that $DHX9^{325–563}$ bound ADP, GDP, CDP, and UDP with similar affinities in the micromolar range, while $DDX20^{41–268}$ showed no detectable binding to any of these nucleotides, except to ADP (Fig. 7b and Table 2). In thermal stabilization assays by differential scanning fluorimetry,[25] different DHX9 constructs were stabilized by ADP, GDP, CDP, and UDP, and DDX20 constructs were stabilized by ADP (results not shown). Taken together, these results confirmed that the DEAD domain itself confers nucleotide selectivity, which can be readily explained by a number of interactions observed in the crystal structures of the DDX20 and by previously published DEAD domains. These results also suggest that DHX9 lacks the selectivity filter for the nucleotide base without having a lower affinity for nucleotides lacking an adenine base.

Discussion

We presented the crystal structures of the human DEIH helicase DHX9 and the DEAD helicase DDX20. Comparison of these structures identified three important features for DHX9: (i) Its overall topology differs from those of the DEAD-box proteins, and some of the motifs that have a conserved position differ in their functions. (ii) DHX9 binds the nucleotide base by stacking with an arginine side chain and does not select for adenine nucleotides. (iii) The N-terminal section of the DEIH domain consists of three α-helices that reside on the surface of the protein, enclosing the core like a claw.

Structural basis for missing base selectivity

It was known previously that *Drosophila* DHX9 can unwind double-stranded RNA using ATP, GTP, CTP, and UTP (as well as their 2′-deoxy forms) as substrates.[7,8] Here we show that also human DHX9 can hydrolyze these nucleotides, and that the DEIH domain can bind the respective nucleoside diphosphates. By contrast, DDX20 and all DEAD helicases that have been studied selectively bind ATP. Thus, both purine and pyrimidine nucleotides can drive DHX9 activity. The structural basis for this promiscuity is the missing selectivity filter for the adenine base: While DDX20 and other DEAD enzymes make extensive specific main-chain and side-chain contacts with positions in the adenine rings, these contacts are replaced in DHX9 by a single π-stacking interaction between an arginine side chain and the aromatic ring electrons of the base. We speculate that the inherent flexibility of the long arginine side chain may contribute to its ability to stack with both purine and pyrimidine bases.

Nucleotide selectivity in DExD/H proteins has been studied extensively and is known to be mediated by the Q-motif in DEAD proteins[32,34] Selectivity for ATP and dATP has also been shown for the prototypical DEAD-box RNA helicase eIF4A[35] and, conversely, mutagenesis of the pyrimidine-interacting glutamate abolishes base selectivity.[36] By contrast, several DExH helicases, including the flavivirus NS3 helicase, have been demonstrated to have a general NTPase activity.[35,37–40] The NS3 protein consists only of the α/β core with no N-terminal extension, and the Q-motif is absent. On the other hand, the arginine side chain that stacks with the base is strictly conserved in all human DExH helicases (Fig. 6). Thus, it seems likely that DExH-box helicases are general NTPases. Under normal conditions in the cell, nucleotide levels are such that DExH proteins probably use ATP. Base promiscuity is possibly of advantage under situations of ATP depletion, when it is a means to ensure continued

Fig. 7. Nucleoside triphosphatase activity and nucleoside diphosphate binding. (a) $DHX9^{325–840}$ hydrolyzed ATP at a rate of 1. 5 s^{-1}. DHX9 also hydrolyzed GTP, CTP, and UTP. Error bars represent standard errors of five determinations. (b) Nucleoside diphosphate binding to the DExD/H domain of DHX9 and DDX20. ITC was used to study the interactions between $DHX9^{325–563}$ and ADP, between $DDX20^{41–268}$ and ADP, between $DHX9^{325–563}$ and CDP, and between $DDX20^{41–268}$ and CDP. Upper panels show the baseline-corrected calorimetric signals, and lower panels show the integrated normalized heat effects (squares) overlaid with the best fit to a one-site binding model (lines). Further data from this experiment are presented in Table 2.

functioning of DExH helicases. It has been observed that GTP depletion leads to translocation of DHX9 from the nucleolus to the nucleoplasm,[8] and the authors speculated that binding of GTP may regulate the localization of DHX9.

Our ITC study shows that for both DHX9 and DDX20, the conserved domain 1 in isolation is able to bind nucleotide with affinity in the physiologically relevant range. The same is also known for other RNA helicases, although some need both conserved domains for efficient nucleotide binding (reviewed by Hilbert *et al.*[41]). Crystal structures of DEAD-box tandem domains[42–44] suggest that upon RNA binding and cleft closure, conserved domain 2 affects nucleotide binding by bringing in the catalytic residues and side chains that interact with the ribose hydroxyl(s).

Notably, we determined that DHX9 binds ADP (and other nucleoside diphosphates) an order of magnitude tighter than DDX20 binds ADP. The crystal structures suggested that the major differences in nucleotide binding between the two enzymes are the orientations of the base and its interactions with protein side chains DHX9–Arg456 and DDX20–Phe82 (Fig. 4). A ligand interface analysis using the PISA server[45] showed that the interface between protein and ADP was somewhat larger in DDX20. However, the predicted solvation energy effect was larger for DHX9 than for DDX20 ($\Delta G = -6.4$ kcal/mol *versus* -5.7 kcal/mol), and the greatest contribution to this difference was made by the base-stacking side chains. We conclude that the missing Q-motif in DExH proteins does not imply reduced affinity for nucleotides, and that the arginine side chain of DHX9 makes a significant contribution to nucleotide affinity.

Link between nucleotide and RNA binding sites

The DHX9 structure contains two linkages between the nucleotide binding site and the RNA binding site, which are both missing in DEAD-box proteins: (i) The DEIH motif is linked to the RNA binding motif Ia through a salt bridge between the Glu515 carboxyl (the residue following the DEIH sequence) and the Arg446 side chain. In DDX20, there is no ionic interaction between the DEAD motif and motif Ia. (ii) DHX9 α-helix 5 harbors not only motif Ia but also the base-stacking Arg456 that is conserved in all human DExH proteins. A base-stacking side chain in the corresponding position has been shown for at least one other DExH protein, NS3, from the hepatitis C virus.[31] This link is missing in DEAD-box proteins, since the base-stacking side chain originates from the Q-motif and not from the α-helix corresponding to DHX9–α5.

The SAT motif has been shown by mutagenesis to uncouple ATPase activity from RNA helicase activity in both DEAD-box[46–48] and DExH proteins.[49,50] Our crystal structures show that the SAT motifs of DHX9 and DDX20 mediate a similar structural connectivity, but the SAT side chains fulfill different functions (Fig. 5b and c). It seems that in all DExD/H proteins, the first position (usually a serine; rarely a threonine) forms a contact with motif VI in conserved domain 2 across the nucleotide binding cleft.[29,44] In DDX20 and other DEAD-box proteins, the serine and threonine side chains link to the DEAD motif, while the alanine side chain links to the P-loop. The AAA mutation thus breaks the linkage between the SAT motif and the DEAD motif. In DHX9 (and other DExH proteins[29]), the DExH and SAT motifs are connected through the alanine side chain and main-chain interactions in the SAT motif. The serine side chain maintains SAT-loop geometry, and the threonine side chain links to the P-loop. Here, the AAA mutation thus breaks the P-loop connection and destabilizes the entire loop. Thus, SAT-loop mutations in DEAD and DExH proteins disturb the coupling between ATPase and RNA unwinding by slightly different mechanisms.

Localization of the PolII binding site

DHX9 binds to PolII and activates CREB-dependent transcription.[12] The MTAD that interacts with PolII has been identified between residues 331 and 380. Three tryptophan mutants (W332A, W339A, and W342A) each confer a reduction in transcriptional activity to 20% of wild-type levels.[16] We localized the MTAD in the crystal structure (Fig. 3). It consists of an extended loop that contains a small β–turn–β element on the N-terminal side of the core domain. All three tryptophans and a number of proline side chains anchor this loop to the surface of the core domain. As there are no crystal contacts formed in this region, the observed placement of the MTAD is likely not a crystallographic artifact. It was shown that the MTAD domain alone is sufficient for transactivation.[16] We speculate that anchoring of the MTAD to the core domain might be an aspect of PolII regulation, possibly involving MTAD release from its surface binding groove.

Materials and Methods

Protein expression and purification

The cDNA encoding full-length human DHX9 was PCR amplified from pooled human brain, liver, placenta, and thymus cDNA libraries (Ambion) using oligonucleotide primers 5′-ATGGGTGACGTTAAAAATTTTC-3′ and 5′-TTACTGTCTACACACAGAAC-3′. The cDNA encoding full-length human DDX20 was obtained from the Mammalian Gene Collection (BC034953). Various expression constructs were obtained by subcloning the coding regions into vector pNIC-Bsa4 by ligation-independent cloning. The resulting expression constructs contained a hexahistidine tag and a tobacco etch virus (TEV) protease cleavage site (MHHHHHHSSGVDLGTENLYFQS) at the N-termini.

Expression constructs were transformed into *E. coli* strain BL21(DE3)R3 pRARE (Novagen), and cultures were grown in Terrific Broth supplemented with 8 g/l glycerol and Antifoam-204 at 37 °C in a LEX bioreactor (Harbinger

Biotechnology and Engineering). At an absorbance at 600 nm (A_{600}) of between 1 and 2, the temperature was lowered to 18 °C, recombinant protein production was induced by addition of 0.5 mM IPTG, and cell growth was continued for 18 h at 18 °C. Cells were harvested by centrifugation, and cell pellets were resuspended in 1.5 vol/wet cell weight of lysis buffer [100 mM Hepes, 500 mM NaCl, 10% glycerol, 10 mM imidazole, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP) (pH 8.0), and one tablet of Complete EDTA-Free protease inhibitor (Roche Biosciences) per 50 ml of cell suspension]. Before lysis, 4 μ l (1000 U) of Benzonase (Novagen) was added per 50 ml of cell suspension, and lysis was achieved by freeze–thaw cycles and sonication. Cell debris was removed by centrifugation, and soluble fractions were filtered through a syringe filter (pore size, 0.45 μm). Selenomethionine-labeled DHX9$^{325–563}$ protein was produced by the pathway inhibition method,[51] and cells were lysed by osmotic shock.[52]

Cleared lysates were passed over 1-ml HiTrap Chelating columns (GE Healthcare) preequilibrated with buffer 1 [30 mM Hepes, 500 mM NaCl, 10% glycerol, 10 mM imidazole (pH 7.5), and 0.5 mM TCEP]. The columns were washed sequentially with buffer 1 and with buffer 1 supplemented with 25 mM imidazole. Bound protein was eluted with buffer 1 containing 500 mM imidazole, the concentration of TCEP was adjusted to 2 mM, the protein was loaded onto 16/60 Superdex-200 HiLoad columns (GE Healthcare), and gel filtration was performed in buffer 2 [20 mM Hepes, 300 mM NaCl, 10% glycerol (pH 7.5), and 0.5 mM TCEP]. Fractions were pooled based on gel-filtration profiles and purity, TCEP was added to 2 mM, and hexahistidine tag was removed by overnight incubation with His_6-tagged TEV protease at a molar ratio of 50:1 at 4 °C or room temperature. After the TEV cleavage reactions had passed over 1-ml HisTrap Chelating columns in buffer 2, the proteins were concentrated using spin concentrators. Proteins were typically more than 90% pure, as judged by SDS-PAGE analysis. Protein construct masses were verified by liquid chromatography electrospray ionization mass spectrometry analysis (results not shown). Aliquots were flash frozen and stored at −80 °C.

Crystallization and structure determination

Crystals of selenomethionine-labeled DHX9$^{325–563}$ were obtained by sitting-drop vapor diffusion at 4 °C after mixing 0.3 μ l of protein solution [20 mg/ml in 20 mM Hepes, 50 mM NaCl, 10% glycerol (pH 7.5), 2 mM TCEP, 1 mM ADP, and 1 mM $MnCl_2$] with 0.2 μ l of well solution [80 mM sodium cacodylate (pH 6.5), 160 mM calcium acetate hydrate, 14.4% polyethylene glycol (PEG) 8000, and 20% glycerol]. Crystals appeared after 3–5 days and continued to grow for 1–2 weeks. Crystals were transferred to a cryosolution [112 mM calcium acetate hydrate, 56 mM sodium cacodylate (pH 6.5), 12% wt/vol PEG 8000, and 30% glycerol] and frozen in liquid nitrogen. Diffraction data were collected at the DIAMOND synchrotron radiation light source. The crystal diffracted to 2.8 Å, and data were integrated into the space group *P*3221. A three-wavelength anomalous dispersion (multi-wavelength anomalous dispersion) experiment was performed. Selenide positions were determined using SHELXD.[53] Phases were computed and improved with autoSHARP.[54] Model building was performed with Buccaneer[55] and Coot,[56] and refinement was performed with Refmac5.[57] Data collection and refinement statistics are summarized in Table 1. Two proteins were found in the asymmetric unit. For both monomers, one ADP molecule and one manganese ion were visible in the electron density.

Crystals of DDX20$^{41–268}$ were obtained by sitting-drop vapor diffusion at 4 °C after mixing 0.1 μ l of protein solution (20 mg/ml), including 20 mM ADP and 10 mM $MgCl_2$, with 0.1 μ l of well solution [0.1 M bis-Tris (pH 5.5), 0.2 M NaCl, and 12% PEG 3350]. Crystals appeared after 1 day and continued to grow for 1 week. Crystals were briefly transferred to a cryosolution (well solution complemented with 28% glycerol) before flash freezing in liquid nitrogen. Diffraction data were collected at the European Synchrotron Radiation Facility. The crystal diffracted to 1.3 Å, and data were integrated into space group $P3_121$. Phasing was performed, and the structure was solved by molecular replacement with MOLREP using PDB entry 1QVA (yeast elF4A-NTD) as search model. Data were integrated into space group $P3_121$, and the asymmetric unit contained two protein monomers. Refmac5 was used for refinement, and Coot was used for model building. TLS-restrained refinement using four TLS groups per monomer was used in the refinement process. The TLS groups were selected using the tlsmd server†. Refinement statistics are given in Table 1. A few residues in the N-termini and C-termini were not visible in the electron density map.

NTPase assay

DHX9$^{325–840}$ was diluted into an assay buffer [20 mM Hepes (pH 7.5), 3 mM $MgCl_2$, 300 mM NaCl, 10% glycerol, and 2 mM TCEP] up to final concentrations of 7–80 nM. The assay was performed in the presence of 0. 1 mg/ml poly(U)-RNA (Sigma-Aldrich). The reaction was started by addition of nucleotides (Sigma) up to a final concentration of 1 mM. NTPase activity was assessed by continuous measurement of the concentration of released inorganic phosphate using the EnzChek phosphate assay kit (Invitrogen) at 22 °C.

Isothermal titration calorimetry

DHX9$^{325–563}$ and DDX20$^{41–268}$ proteins were dialyzed overnight at 4 °C in 20 mM Hepes (pH 7.5), 3 mM $MgCl_2$, 300 mM NaCl, 10% glycerol, and 2 mM TCEP. Protein concentration after dialysis was determined using extension coefficients ε_{280}(DHX9) = 23, 950 M^{-1} cm^{-1} and ε_{280} (DDX20) = 12, 950 M^{-1} cm^{-1}. The nucleotides (Sigma) were dissolved in dialysis buffer, and their concentrations were determined using extinction coefficients ε_{259} = 15, 400 M^{-1} cm^{-1} for ADP, ε_{252} = 13, 700 M^{-1} cm^{-1} for GDP, ε_{271} = 9000 M^{-1} cm^{-1} for CDP, and ε_{262} = 10, 000 M^{-1} cm^{-1} for UDP. ITC was performed at 25 °C using an ITC200 microcalorimeter (MicroCal™ GE Healthcare Life Sciences) with a cell volume of 200 μ l. A protein solution (300–433 μM) was loaded into the sample cell and titrated with nucleotide solution (4.5–7.7 mM; initial injection of 0.3 μ l and 15 subsequent 2.47-μ l injections) at a stirring rate of 1000 rpm. Results were analyzed with Origin™ software (MicroCal™ GE Healthcare Life Sciences). The binding enthalpy values obtained were corrected for heat of dilution of the ligand into the buffer and fitted to a one-site binding model.

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

Sequence alignments

Sequences were aligned using T-Coffee[58] and manually adjusted. The following sequences were included: Fig. 1: DDX20_HS (human DDX20; accession code Q9UHI6), DDX10_HS (human DDX10; Q13206), DDX2A_HS (human eukaryotic initiation factor 4A-1; P60842), DDX47_HS (human DDX47; Q9H0S4), HEL308_AF (*A. fulgidus* Hel308; NP_071282), NS3_DV (dengue virus 4 nonstructural protein 3; AAW30973), RecBCD_EC (*E. coli* RecBCD/exonuclease V β-subunit; NP_417297), DHX9_HS (human DHX9; Q08211); Fig. 6: human DHX9 (Q08211), DHX16 (O60231), DHX35 (Q9H5Z1), DHX15 (O43143), DHX32 (Q7L7V1), DHX8 (Q14562), DHX33 (Q9R6H0), DHX40 (Q8IX18), DHX30 (Q7L2E3), DHX36 (Q9H2U1), DHX57 (Q6P158), DHX29 (Q7Z478), DHX34 (Q14147), DHX37 (Q8IY37).

Accession numbers

Coordinates and structure factors have been deposited in the PDB with accession numbers 2OXC and 3LLM.

Acknowledgements

We thank Susanne van den Berg and Mailén Andersson for technical assistance. The Structural Genomics Consortium is a registered charity (no. 1097737) that receives funds from the Canadian Institutes for Health Research, the Canada Foundation for Innovation, Genome Canada, through the Ontario Genomics Institute, GlaxoSmithKline, Karolinska Institutet, the Knut and Alice Wallenberg Foundation, the Ontario Innovation Trust, the Ontario Ministry for Research and Innovation, Merck and Co., Inc., the Novartis Research Foundation, the Swedish Agency for Innovation Systems, the Swedish Foundation for Strategic Research, and the Wellcome Trust. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2010.05.046

References

1. Cordin, O., Banroques, J., Tanner, N. K. & Linder, P. (2006). The DEAD-box protein family of RNA helicases. *Gene*, **367**, 17–37.
2. Chu, V. B. & Herschlag, D. (2008). Unwinding RNA's secrets: advances in the biology, physics, and modeling of complex RNAs. *Curr. Opin. Struct. Biol.* **18**, 305–314.
3. Gubaev, A., Hilbert, M. & Klostermeier, D. (2009). The DNA-gate of *Bacillus subtilis* gyrase is predominantly in the closed conformation during the DNA supercoiling reaction. *Proc. Natl Acad. Sci. USA*, **106**, 13278–13283.
4. Singleton, M. R., Dillingham, M. S. & Wigley, D. B. (2007). Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76**, 23–50.
5. Pyle, A. M. (2008). Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu. Rev. Biophys.* **37**, 317–336.
6. Beran, R. K. F., Bruno, M. M., Bowers, H. A., Jankowsky, E. & Pyle, A. M. (2006). Robust translocation along a molecular monorail: the NS3 helicase from hepatitis C virus traverses unusually large disruptions in its track. *J. Mol. Biol.* **358**, 974–982.
7. Lee, C. G. & Hurwitz, J. (1992). A new RNA helicase isolated from HeLa cells that catalytically translocates in the 3′ to 5′ direction. *J. Biol. Chem.* **267**, 4398–4407.
8. Huang, M. & Mitchell, B. S. (2008). Guanine nucleotide depletion mediates translocation of nucleolar proteins, including RNA helicase A (DHX-9). *Nucleosides Nucleotides Nucleic Acids*, **27**, 704–711.
9. Tettweiler, G. & Lasko, P. (2006). A new model for translational regulation of specific mRNAs. *Trends Biochem. Sci.* **31**, 607–610.
10. Hartman, T. R., Qian, S., Bolinger, C., Fernandez, S., Schoenberg, D. R. & Boris-Lawrie, K. (2006). RNA helicase A is necessary for translation of selected messenger RNAs. *Nat. Struct. Mol. Biol.* **13**, 509–516.
11. Fuller-Pace, F. V. (2006). DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. *Nucleic Acids Res.* **34**, 4206–4215.
12. Nakajima, T., Uchida, C., Anderson, S. F., Lee, C. G., Hurwitz, J., Parvin, J. D. & Montminy, M. (1997). RNA helicase A mediates association of CBP with RNA polymerase II. *Cell*, **90**, 1107–1112.
13. Robb, G. B. & Rana, T. M. (2007). RNA helicase A interacts with RISC in human cells and functions in RISC loading. *Mol. Cell*, **26**, 523–537.
14. Schlegel, B. P., Starita, L. M. & Parvin, J. D. (2003). Overexpression of a protein fragment of RNA helicase A causes inhibition of endogenous BRCA1 function and defects in ploidy and cytokinesis in mammary epithelial cells. *Oncogene*, **22**, 983–991.
15. Mischo, H. E., Hemmerich, P., Grosse, F. & Zhang, S. (2005). Actinomycin D induces histone gamma-H2AX foci and complex formation of gamma-H2AX with Ku70 and nuclear DNA helicase II. *J. Biol. Chem.* **280**, 9586–9594.
16. Aratani, S., Fujii, R., Oishi, T., Fujita, H., Amano, T., Ohshima, T. *et al.* (2001). Dual roles of RNA helicase A in CREB-dependent transcription. *Mol. Cell Biol.* **21**, 4460–4469.
17. Tang, H., Gaietta, G. M., Fischer, W. H., Ellisman, M. H. & Wong-Staal, F. (1997). A cellular cofactor for the constitutive transport element of type D retrovirus. *Science*, **276**, 1412–1415.
18. Isken, O., Grassmann, C. W., Sarisky, R. T., Kann, M., Zhang, S., Grosse, F. *et al.* (2003). Members of the NF90/NFAR protein group are involved in the life cycle of a positive-strand RNA virus. *EMBO J.* **22**, 5655–5665.
19. Shpargel, K. B. & Matera, A. G. (2005). Gemin proteins are required for efficient assembly of Sm-class ribonucleoproteins. *Proc. Natl Acad. Sci. USA*, **102**, 17372–17377.
20. Cauchi, R. J., Davies, K. E. & Liu, J. L. (2008). A motor function for the DEAD-box RNA helicase, Gemin3, in *Drosophila*. *PLoS Genet.* **4**, e1000265.
21. Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L. *et al.* (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* **16**, 720–728.

22. Grundhoff, A. T., Kremmer, E., Tureci, O., Glieden, A., Gindorf, C., Atz, J. *et al.* (1999). Characterization of DP103, a novel DEAD box protein that binds to the Epstein–Barr virus nuclear proteins EBNA2 and EBNA3C. *J. Biol. Chem.* **274**, 19136–19144.
23. Almstead, L. L. & Sarnow, P. (2007). Inhibition of U snRNP assembly by a virus-encoded proteinase. *Genes Dev.* **21**, 1086–1097.
24. Graslund, S., Sagemark, J., Berglund, H., Dahlgren, L. G., Flores, A., Hammarstrom, M. *et al.* (2008). The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. *Protein Expression Purif.* **58**, 210–221.
25. Niesen, F. H., Berglund, H. & Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212–2221.
26. Yang, J. M. & Tung, C. H. (2006). Protein structure database search and evolutionary classification. *Nucleic Acids Res.* **34**, 3646–3659.
27. He, Y., Andersen, G. R. & Nielsen, K. H. (2010). Structural basis for the function of DEAH helicases. *EMBO Rep.* **11**, 180–186.
28. Büttner, K., Nehring, S. & Hopfner, K. P. (2007). Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat. Struct. Mol. Biol.* **14**, 647–652.
29. Luo, D. H., Xu, T., Watson, R. P., Scherer-Becker, D., Sampath, A., Jahnke, W. *et al.* (2008). Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein. *EMBO J.* **27**, 3209–3219.
30. Singleton, M. R., Dillingham, M. S., Gaudier, M., Kowalczykowski, S. C. & Wigley, D. B. (2004). Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature*, **432**, 187–193.
31. Gu, B., Gates, A. T., Isken, O., Behrens, S. E. & Sarisky, R. T. (2003). Replication studies using genotype 1a subgenomic hepatitis C virus replicons. *J. Virol.* **77**, 5352–5359.
32. Tanner, N. K. (2003). The newly identified Q motif of DEAD box helicases is involved in adenine recognition. *Cell Cycle*, **2**, 18–91.
33. 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.
34. 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.
35. Du, M. X., Johnson, R. B., Sun, X. L., Staschke, K. A., Colacino, J. & Wang, Q. M. (2002). Comparative characterization of two DEAD-box RNA helicases in superfamily II: human translation-initiation factor 4A and hepatitis C virus non-structural protein 3 (NS3) helicase. *Biochem. J.* **363**, 147–155.
36. Sinha, K. M., Glickman, M. S. & Shuman, S. (2009). Mutational analysis of *Mycobacterium* UvrD1 identifies functional groups required for ATP hydrolysis, DNA unwinding, and chemomechanical coupling. *Biochemistry*, **48**, 4019–4030.
37. Claude, A., Arenas, J. & Hurwitz, J. (1991). The isolation and characterization of an RNA helicase from nuclear extracts of HeLa cells. *J. Biol. Chem.* **266**, 10358–10367.
38. Shuman, S. (1993). Vaccinia virus RNA helicase. Directionality and substrate specificity. *J. Biol. Chem.* **268**, 11798–11802.
39. Zhang, S. & Grosse, F. (1994). Nuclear DNA helicase II unwinds both DNA and RNA. *Biochemistry*, **33**, 3906–3912.
40. Lin, C. & Kim, J. L. (1999). Structure-based mutagenesis study of hepatitis C virus NS3 helicase. *J. Virol.* **73**, 8798–8807.
41. Hilbert, M., Karow, A. R. & Klostermeier, D. (2009). The mechanism of ATP-dependent RNA unwinding by DEAD box proteins. *Biol. Chem.* **390**, 1237–1250.
42. 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.
43. 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.
44. Collins, R., Karlberg, T., Lehtio, L., Schutz, P., van den Berg, S., Dahlgren, L. G. *et al.* (2009). The DEXD/H-box RNA helicase DDX19 is regulated by an {alpha}-helical switch. *J. Biol. Chem.* **284**, 10296–10300.
45. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
46. Scheffner, M., Knippers, R. & Stahl, H. (1989). RNA unwinding activity of SV40 large T antigen. *Cell*, **57**, 955–963.
47. Flores-Rozas, H. & Hurwitz, J. (1993). Characterization of a new RNA helicase from nuclear extracts of HeLa cells which translocates in the 5′ to 3′ direction. *J. Biol. Chem.* **268**, 21372–21383.
48. 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.
49. Gross, C. H. & Shuman, S. (1998). The nucleoside triphosphatase and helicase activities of vaccinia virus NPH-II are essential for virus replication. *J. Virol.* **72**, 4729–4736.
50. Kim, D. W., Gwack, Y., Han, J. H. & Choe, J. (1997). Towards defining a minimal functional domain for NTPase and RNA helicase activities of the hepatitis C virus NS3 protein. *Virus Res.* **49**, 17–25.
51. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124.
52. Magnusdottir, A., Johansson, I., Dahlgren, L. G., Nordlund, P. & Berglund, H. (2009). Enabling IMAC purification of low abundance recombinant proteins from *E. coli* lysates. *Nat. Methods*, **6**, 477–478.
53. Sheldrick, G. M. (2008). A short history of SHELX. *Acta Crystallogr. Sect. A*, **64**, 112–122.
54. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230.
55. Cowtan, K. (2006). The Buccaneer software for automated model building: 1. Tracing protein chains. *Acta Crystallogr. Sect. D*, **62**, 1002–1011.
56. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D*, **60**, 2126–2132.
57. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D*, **53**, 240–255.
58. Notredame, C., Higgins, D. G. & Heringa, J. (2000). T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217.