Supplementary Information

The mechanism of RNA duplex recognition and unwinding by DEAD-box helicase DDX3X

He Song and Xinhua Ji

		→ NTE D1 –	-
DDX3X/131-609	131 SDEDDWSKPLPPSERLEQELFSG	GN TGINFEKYDD I PVEATGNNC PPHIESFSDVE	M 187
Ded1/92-563	92KHVPAPRNEKAEIAIFGVPED	PNFQSSGINFDNYDD I PVDASGKDVPEPITEFTSPP	L 149
DDX4/240-688	240 QGPKVTYIPPPPPED-EDSIFAH	YQTGINFDKYDT I LVEVSGHDAPPAILTFEEAN	L 295
Vasa/196-632	196 ERKREFYIPPEPSND-AIEIFSSG	I ASGIHFSKYNN I PVKVTGSDVPQPIQHFTSAD	L 252
	Q	l l	
DDX3X/131-609	188 GE I I MGN I EL TRYTR PTPVQKHA I PI	IKEKRDLMACAQTGSGKTAAFLLPILSQIYSDGPGE	A 250
Ded1/92-563	150 DGL L LEN IK LAR FTK PTPVQK Y SVPI	VANGRDLMACAQTGSGKTGGFLFPVLSESFKTGPSP	Q 212
DDX4/240-688	296 CQT LNNN I AKAGYTK L TPVQK Y SI PI	ILAGRDLMACAQTGSGKTAAFLLPILAHMMHDGITA	S 358
Vasa/196-632	253 RD I I I DNVNK SGYK I PTPIQKC SI PV	ISSGRDLMACAQTGSGKTAAFLLPILSKLLEDPHEL	E 315
	la	lb	
DDX3X/131-609	251 LRAMKENGRYGRRKQYPISLVLAPTR	ELA VQIYEEARKFSYRSRVRPCVVYGGAD I GQQIRD	L 313
Ded1/92-563	213 P ESQGSFYQRKAYPTAVIMAPTR	ELA TQIFDEAKKFTYRSWVKACVVYGGSPIGNQLRE	I 272
DDX4/240-688	359 R FKELQEPECIIVAPTR	ELVNQIYLEARKFSFGTCVRA VVIYGGTQLGHSIRQ	I 412
Vasa/196-632	316 L	ELA IQIFNEARKFAFESYLKIGIVYGGTSFRHQNEC	I 365
	lc	II	
DDX3X/131-609	314 ERGCHLLVATPGRLVDMMERGKIGLD	FCKYLVLDEADRMLDMGFEPQIRRIVEQDTMPPKGV	R 376
Ded1/92-563	273 ERGCDLLVATPGRLNDLLERGKISLA	NVKYLVLDEADRMLDMGFEPQIRHIVEDCDMTPVGE	R 335
DDX4/240-688	413 VQGCNILCATPGRLMDIIGKEKIGLK	QIKYLVLDEADRMLDMGFGPEMKKLISCPGMPSKEQ	R 475
Vasa/196-632	366 TRGCHVVIATPGRLLDFVDRTFITFE	DTRFV <mark>VLDEADRMLDMGF</mark> SEDMRRIMTHVTMRPE	H 426
		D1 D2→	
DDX3X/131-609	377 HTMMF SATF PKEIQMLARD FLD - EYI	FLAVGRVGSTSEN I TOKVVWVEESDKRSFLLDLLNA	T 438
Ded1/92-563	336 QTLMF SATF PADIQHLARD FLS - DYI	FLSVGRVGSTSEN I TOKVLYVENODKKSALLDLLSA	S 397
DDX4/240-688	476 QTLMF SATF PEEIQRLAAEFLK SNYL	FVAVGQVGGACRDVQQTVLQVGQFSKREKLVEILRN	I 538
Vasa/196-632	427 QTLMF SATF PEEIQRMAGEFLK - NYV	FVAIGIVGGACSDVKQTIYEVNKYAKRSKLIEILSE	Q 488
	IV	IVa V	
DDX3X/131-609	439 GKD SLTLVFVETKKGAD SLEDFLYHE	GYACT <mark>SIHGDR SQRDRE EALHQFR</mark> SGK SPIL VATAV	A 501
Ded1/92-563	398 T - DGLTL I FVETKRMADQLTDFL I MQ	NFRATAIHGDR TQ SERERALAAFR SGAATLL VATAV	A 459
DDX4/240-688	539 G - DER TMVFVETKKKADF I AT FLCQE	KISTTSIHGDR EQRER EQALGDFR FGKC PVL VATSV	A 600
Vasa/196-632	489 A - DG - TIVFVETKRGADFLASFLSEK	EFPTTSIHGDR LQ SQR EQALRDFK NG SMK VLIATSV	A 549
	Va		
DDX3X/131-609	502 ARGLD I SNVKHV INFDLPSD I EEYVH	R I GRTGR VGN LGLAT SF FN - ERN IN I TKDLLDLLVE	A 563
Ded1/92-563	460 ARGLD I PNVTHV INYDLPSDVDDYVH	R I GRTGRAGN TGLATAFFN - SEN SN I VKGLHE I LTE	A 521
DDX4/240-688	601 ARGLD I ENVQHV INFDLPST I DEYVH	R I GRTGRC GNTGRA I SF FDLE SDNH LAQPL VK VLTD	A 663
Vasa/196-632	550 SRGLD I KN I KHV I NYDMPSK I DDYVH	R I GRTGR VGNNGRAT SF FDPEKDRA I AADLVK I LEG	S 612
		-CTF	
DDX3X/131-609	564 KQEVPSWLENMAYEHHYKGSSRGRSK	SSRFSGGFGARDYRQSSGAS	609
Ded1/92-563	522 NQEVPSFLKDAMMSAPGSRS	NSR-RGGFGRNNNRDYRKAGGAS	563
DDX4/240-688	664 QQDVPAWLEEIAF	-STYIPGFSGSTR	688
Vasa/196-632	613 GQTVPDFLRTCGA	GGDGGYS	632

Supplementary Figure 1 | Sequence alignment of founding members of the DDX3X/Ded1 and DDX4/Vasa subfamilies of DEAD-box RNA helicases. Identical residues and similar residues are shaded in blue and light blue, respectively. Boundaries of the NTE (N-terminal extension), D1 (RecA-like domain 1), D2 (RecA-like domain 2), and CTE (C-terminal extension) are indicated. The 12 highly conserved sequence motifs are boxed, indicated, and color coded: red, ATP binding and hydrolysis; blue, RNA binding; and orange, coordination between ATP and RNA binding.



Supplementary Figure 2 | **RNA duplex recognition by human DDX3X. a,** Asymmetric binding of dsRNA by the D1D2 core of human DDX3X as observed in the D1D2:dsRNA:D1D2 structure. The D1D2 core contains amino acid residues 132-607 of human DDX3X (Fig. 1a) and the RNA contains 26 base pairs with a 2-nucleotide 3' overhang at both terminal ends. The asymmetric binding results in different sequences for the two bound RNA strands. b, Symmetric binding of the dsRNA by the two D1D2 cores as imagined would results in the same sequence for the two bound RNA strands.



Supplementary Figure 3 | **Detailed interactions between the CTE of human DDX3X and dsRNA. a**, S520 and N551 of CTE recognize the 2'-hydroxyl groups of nucleotide residues 9 and 10 in the central region of RNA Strand 2 (backbone in red). **b**, H578 and H579 of CTE recognize the nucleotide residues 10 and 11 in the central region of RNA Strand 1 (backbone in blue).



Supplementary Figure 4 | The DDX3X D1D2:dsRNA:D1D2 assembly vs the asymmetric unit content of the Mss116p D2:dsRNA structure in stereo. The D1D2:dsRNA:D1D2 assembly of DDX3X (this work), which is revealed in crystal, also observed in solution, and validated by Hill kinetics data, is illustrated as a cartoon model in cyan for D1D2^{MolA}, orange for D1D2^{MolB}, blue for RNA Strand 1, and red for RNA Strand 2. The two D1D2 cores are also outlined with transparent molecular surfaces. The asymmetric unit of the Mss116p D2:dsRNA structure, containing four D2s (D2 α through δ) and a pseudocontinuous RNA duplex (the D2:D2:pseudo-dsRNA:D2:D2 ensemble, PDB: 4DB2) is shown as a cartoon model in palegreen. The two structures are superpositions on the bases of D2a in the D1D2:dsRNA:D1D2 assembly and D2 α in the D2:D2:dsRNA:D2:D2 ensemble.



Data-collection Parameters	
Beam line	12-ID-B (Advanced Photon Source)
Wavelength (Å)	0.8856
Detector	Pilatus 1M (SAXS), Pilatus 100K (WAXS)
$q \operatorname{range} (\operatorname{\AA}^{-1})$	0.006-2.8
Exposure time (s)	2
Concentration range (mg/ml)	1-4
Temperature	300
Structural Parameters	
<i>I</i> (0) from Guinier fitting	0.309 ± 0.003
$R_{\rm g}$ from Guinier fitting (Å)	33.8 ± 0.3
<i>I</i> (0) from GNOM	0.292 ± 0.004
<i>R</i> _g from GNOM (Å)	33.8 ± 0.3
D _{max} (Å)	105 ± 5
^a MW ^{pred} (kDa)	124
^b MW ^{saxs} (kDa)	120
Fit χ^2 value	1.72
Software Employed	
Primary Data Processing	Igor Pro
PDDF	GNOM
Ab initio Shape Analysis	DAMMIN
SAXS Profile Computation	CRYSOL
Superposition and averaging	SUPCOMB/DAMAVER
Molecular Visualization	PyMol

Supplementary Table 1. Data-collection and scattering-derived parameters

^a The MW^{pred} was calculated from the primary sequence.

^b The MW^{saxs} was calculated using the web portal "SAXS MoW".