DDX43 prefers single strand substrate and its full binding activity requires physical connection of all domains

Han Wu, Liu-Tao Zhai, Peng-Yang Chen, Xu-Guang Xi

HAL Id: hal-02389895

https://cnrs.hal.science/hal-02389895

Submitted on 21 Dec 2021

DDX43 prefers single strand substrate and its full binding activity requires physical connection of all domains

Han Wu[1], Liu-Tao Zhai[1], Peng-Yang Chen[1], Xu-Guang Xi[1,2*]

From the[1] State Key Laboratory of Crop Stress Biology in Arid Areas, College of Life Sciences, Northwest A&F University, Yangling, Shaanxi 712100, China; [2] the LBPA, Ecole Normale Supérieure Paris-Saclay, CNRS, Université Paris Saclay, 61 Avenue du Président Wilson, F-94235 Cachan, France.

*To whom correspondence should be addressed: Xu-Guang Xi: CNRS, 61 Ave. du Président Wilson, 94235 Cachan, France. E-mail: xxi01@ens-cachan.fr; Tel.: 33-47407754; Fax: 33-47407754.

Abstract

DDX43 is a cancer/testis antigen and is thought to stimulate oncogenic pathways in cell proliferation, while its specific function in cancer development is largely unexplored. DDX43 is the member of RNA helicase in DEAD-box family, consists of conserved helicase core and a single K-homology (KH) domain in its N-terminus. In this paper, we expressed and purified human DDX43 protein in *E. coli* and demonstrated that this protein is a homogeneous monomer. To understand the role and explore the substrates preference of DDX43 *in vitro*, we systematically studied its binding properties. We found that DDX43 prefers single-strand DNA or RNA with length longer than 12 nt and much prefers guanosine than the other three nucleotides. Achievement of the full binding affinity

1

of protein to substrate needs the existence of all domains, and they must be connected. The absence of either of them or the disjunction can result in a decreased binding affinity to substrates, approximately reduced 10-fold. We also found that the unwinding ability of DDX43 *in vitro* was neither efficient nor sustainable.

Keywords: DDX43, DEAD-box, KH domain, DNA binding, RNA helicase

1. Introduction

Helicase superfamily 2 (SF2) proteins are ubiquitous in RNA biology and have a very wide range of functions. All SF2 proteins share a conserved helicase core, which binds nucleoside triphosphate and nucleic acid and consists of two domains (D1 and D2 for short), each of which is similar to RecA [1]. The DEAD-box proteins are the largest family of RNA helicases in SF2, there are at least 26 different members of this family in yeast and more than 37 in human [1]. Members within families are highly conserved in nine motifs (Q, I, Ia, Ib, II, III, IV, V, VI) [2] but different at their N- or C-terminus, these additional distinct domains enriched their functional diversity. DEAD-box proteins are engaged in various metabolic processes, typically involve RNAs, such as translation (e.g. VASA [3]), mRNA splicing (e.g. Prp28 [4]), and RNA turnover (e.g. Dhh1 [5]), they also engaged in the biogenesis of pre-ribosomal/ribosomal subunits [6]. Only a limited numbers of full-length DEAD-box protein structures have been determined, mainly due to the long flexible linker between helicase core and the extra domain.

DDX43, also known as helicase antigen gene (HAGE), is a member of DEAD-box family and initially found as a cancer/testis antigen (CTA) [7]. The helicase core of DDX43 is flanked by

2

ancillary N-terminal KH domain which is the most prevalent RNA binding domain [8]. It consists of approximately 70 amino acids, with a conserved sequence VIGxxGxxI mapping to the middle. In all KH domain-nucleic acid complexes, the nucleic acid back bone interacts with the conserved GxxG loop, the hydrophobic groove of canonical KH domain can recognize up to four nucleotides [9], domains with a classical KH fold but lacking the conserved GxxG loop have shown no nucleic acid-binding activity [10].

DDX43 is overexpressed in many solid tumors but absent in normal tissues except testis, indicating its biomarker role in tumorigenesis. In uveal melanoma cells, the upregulation of *DDX43* leads to the constitutive hyper-activation of RAS [11] and the demethylation of *DDX43* may predict an unfavorable outcome in acute myeloid leukemia (AML) [12]. Despite the biological role of DDX43 in tumorigenesis and the successful structural characterization of both helicase core and KH domain, the structure of DDX43 remains unsolved and little is known about its activities *in vitro*.

In this experiment, we have purified human DDX43 protein and systematically studied the basic enzymatic properties of DDX43, hoping to further understand this new tumor marker. On the one hand, suitable nucleic acids can stabilize the helicase core and KH domain, reduce the entropy value of protein, increasing the probability of crystallization. On the other hand, it helps us speculate about the possible role of DDX43 in cancer by exploring the types of binding substrates.

2. Materials and Methods

2.1 Protein expression and purification

The amino acid sequence of human DDX43 was obtained from GenBank (Number:

NP_061135.2) and its gene was synthesized in General Biosystems, Inc. *DDX43* and a series of truncations were cloned into pET-15b-SUMO vector, and all recombinant plasmids were transformed into the *Escherichia coli* 2566(DE3). Cultures were cultivated at 37℃ until OD_{600} reached ~0.8 and then incubated for 16 h with 0.3 mM IPTG at 18℃. Cells were collected and pellets were re-suspended in lysis buffer (500 mM NaCl, 5 mM imidazole, 20 mM Tris-HCl pH 7.5, 5% glycerol, 2 mm EDTA) and lysed using an ultra-high pressure cell disrupter (JNBIO). EDTA is necessary in maintain protein stability during the purification process. After centrifugation, the supernatant were loaded onto Ni-NTA (GE Healthcare) and eluted using elution buffer (300 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl pH 7.5, 5% glycerol, 2mM EDTA). Eluted protein was digested with SUMO protease to remove the SUMO-tag. The SUMO-digested protein was loaded on Hitrap-Heparin (GE Healthcare) and eluted in a 150 to 1000 mM NaCl gradient in AKTA purifier in buffer Heparin (20 mM Tris-HCl pH 7.5, 5% glycerol, 2 mM EDTA). Target protein was collected, concentrated and further purified by gel filtration. Purified protein was concentrated to 20 mg/ml and stored at -80℃ until use.

2.2 Oligonucleotide substrates and fluorescence polarization binding assay

The substrates used in binding assays as well as in superdex200 were purchased from Sangon Biotech (Shanghai, China). The structure, sequences of unlabeled or fluorescently labeled DNA/RNA substrates were shown in Table S1. All synthetic structured oligonucleotides were purified using the same methods in [13].

The affinity between protein and substrates were analyzed by fluorescence polarization assay

using Infinite F200 PRO (TECAN) at a constant temperature of 20°C. Varying amounts of protein were added to 100 μl aliquot binding buffer (50 mM NaCl, 25 mM Tris–HCl, pH 7.5, 2 mM MgCl$_2$ and 2 mM DTT) containing 5 nM fluorescein-labeled substrate. Each sample was allowed to shake 20 s and equilibrate in solution for 2 min. After 2 min, the steady-state fluorescence anisotropy (*r*) was measured. A second reading was taken after 5 min, to ensure that the mixture was well equilibrated and stable. Less than 5% change was allowable between the 2- and 5- min measurements. If the binding curve is in sigmoidal shape, the curve can be fit by equation r = r$_{max}$ * P/ (*K*d + P), where r$_{max}$ is the maximal amplitude of the anisotropy (=r$_{max,complex}$ −r$_{free DNA}$), P is the protein concentration, *K*d is the mid-point of the curve corresponding to the apparent dissociation constant which could reflects the binding affinity of the protein with the fluorescence-labelled substrates, the lower the *K*d value is, the higher affinity is between the protein and the substrate [13].

2.3 Dynamic light scattering (DLS) and gel filtration analysis

DLS measurements were performed at 20°C using the DynaPro NanoStar instrument (Wyatt Technology Corporation) with a thermostat cell holder. The protein concentration was 5 to 50 μM in DLS buffer (500 mM NaCl, 20 mM Tris-HCl pH 7.5, 5% glycerol, 2 mM EDTA). A measurement time period of 50 s—10 acquisitions, 5 s each, were applied. Data were analyzed using Dynamics7.0 software with regularization methods as described in [13].

The apparent molecular weight of DDX43 was measured using superdex200 10/300 column (GE Healthcare) as described in [14]. Equation log (*Mw*) =-2.9716x+7.5407 (R^2=0.9936) was used to calculate the corresponding molecular weight of DDX43, where x represents V$_x$/V$_0$, V$_x$ represent

protein elution volume and V_0 equals 8.0 ml for bed volume. For binding analysis, the elution volume of single protein and protein-DNA mixture in superdex200 were compared, if the elution volume of protein-DNA mixture was shifted forward than single protein, then protein-DNA complex was formed *in vitro*. All loading samples were in the same volume (80 μl) and eluted in same buffer (50 mM NaCl, 20 mM Tris-Cl pH 7.5, 2 mM EDTA, 5% glycerol). DDX43 and its substrates were mixed in a ratio of 1:1.1 to avoid free protein.

2.4 Stopped-flow FRET measurement

A stopped-flow FRET assay was used [13] for measuring unwinding kinetic rate constants of DDX43 using doubly labeled forked RNA substrates (FA-FD). Unwinding kinetics were measured in a two-syringe mode, 500 nM DDX43 and 5 nM RNA substrates were pre-incubated in the unwinding reaction buffer (25 mM Tris–HCl pH 7.5, 15 mM NaCl, 2 mM $MgCl_2$ and 2 mM DTT) at 37°C for 5 min (in syringe 3) and the unwinding reaction was initiated by rapid mixing with 2 mM ATP (in syringe 4). All reported concentrations correspond to final concentrations, i.e. after mixing.

3. Results

3.1 DDX43 is monomeric in solution with high homogeneity

KH domain in N-terminal of DDX43 is a small αβ nucleic acid recognition domain. According to the prediction results of IUpred2A and ANCHOR2 (https://iupred2a.elte.hu/), we found that DDX43 was overall ordered except the initial ~60 aa (Figure 1A). The recombinant DDX43 protein was purified to near homogeneity as judged by their appearance on SDS-PAGE gel (Figure 1B) and

matched its theoretical molecular weight of 73 kDa. In order to explore the oligomeric state of

DDX43 in solution, DLS and analytical size exclusion chromatography were performed. In the 5-50

μM concentration range, the hydrodynamic radii of DDX43 was monomodal monodispersed in 4.632

nm with 4.6% polydispersity (Figure 1C). In gel filtration analysis, DDX43 was eluted as a single

peak in 15 ml (Figure 1D), with its apparent molecular weight of 93 kDa. These data suggest that

DDX43 is monomeric in solution with high homogeneity but might not present as a regular sphere

shape because we got a slightly larger value than the assumed spherical protein.

3.2 DDX43 prefers to bind single strand substrate

In order to explore the possible physiological function of DDX43, we first studied its binding

property using thirteen types of substrates (Figure 2A, Figure S1A). The binding dissociation constant

(Kd) for each substrate was summarized in Table S2. The binding affinity of DDX43 to different

substrates can be classified into three levels according to the determined Kd value. (a) The highest

binding affinity of DDX43 was single strand substrate with Kd value less than 10 nM and exhibited

no differences between ssDNA and ssRNA. This characteristic was consistent with the properties of

KH domain. (b) DDX43 also preferred forked duplex DNA (Kd= 40.3 ± 3.0 nM) and 12 nt bubble

DNA (Kd= 41.6 ± 5.1 nM) with a slight increase Kd value. The differences between 12nt bubble DNA

and 4 nt bubble DNA (Kd= 95.9 ± 18.4 nM) reminded that DDX43 preferred substrate with longer

single-strand region. It also favored 3G3 DNA than 3G4 DNA, mostly because 3G3 is unstable in

solution and tend to destructed to single strand DNA. DDX43 shown polarity preferences to 3'-tailed

DNA substrates than 5'-tailed, with Kd value increased about twice. (c) DDX43 exhibited the least

sensitive for blunt-ended dsDNA and stabled 4 layer G4 DNA, with Kd value larger than 200 nM. In summary, DDX43 formed protein–substrate complexes with all of the substrates tested and with the overwhelming affinity for the single strand DNA/RNA and the lowest affinity for the blunt-end dsDNA.

To elucidate potential sequence specificity binding for DDX43, four substrates in same length but with different nucleotides, named poly12T, poly12C, poly12A and $(GT)_6$ were used. The absence of poly12G was due to the limitation of synthesis. We found DDX43 exhibited binding affinity with nucleotides in the order of G> T> C> A, shown the preferences for guanine much than the other three nucleotides (Figure 2B, Figure S1B).

The affinity difference between 12nt bubble DNA and 4 nt bubble DNA reminded us the necessity of the determination for substrate length. Single strand DNA in random sequences, ranging from 10 nt to 42 nt, were used to test the minimum binding length of DDX43 (Table S1). No binding affinity was observed when ssDNA was 10 nt and DDX43 didn't exhibited high affinity with ssDNA until the length reached 16 nt. DDX43 showed similar affinity to ssDNA ranging from 16 nt to 42 nt with Kd value less than 15 nM (Figure 2C, Figure S1C). The minimum ssDNA substrate length was further confirmed through the elution volume difference of S200. When DDX43 and 11nt ssDNA were mixed and incubated at 20℃ for 30 min, the elution volume of DDX43 among the mixture remained the same to DDX43 alone, both were eluted at 15 ml. When DDX43 mixed with 12 nt ssDNA, the elution volume of DDX43 was shifted from 15 ml to around 14.4 ml together with the majority of 12 nt ssDNA, suggesting DDX43 exhibited least affinity with 11nt ssDNA (Figure 2D). Therefore, we proved that the least length of ssDNA required for DDX43 binding was 12nt.

3.3 DDX43 has two equally important substrate binding regions

DDX43 has two theoretically function domains, KH domain and helicase core. To find out which domain plays a greater role in substrate binding, we have constructed and purified eight truncated proteins (Figure 3A). All of them were in the same homogeneity as wild DDX43 as shown on SDS-PAGE (Figure 3B).

The Kd value of eight truncated proteins binding to 12 nt $(GT)_6$, 24 nt ssDNA and 22 nt ssRNA were measured (Figure 3C). Above three substrates were all of high affinity to wild DDX43 and representing ssDNA in specific sequence, ssDNA in random sequence and ssRNA, respectively. Based on the Kd values from truncated proteins, we can draw three conclusions. First, the integrity of N-terminus was essential to maintain high affinity to substrates. For example, DDX43^{30-R} and DDX43^{65-R}, two proteins that both retained complete function domains yet their Kd values of all three substrates were increased evidently, suggesting that the unstructured N-terminal region beyond KH domain was also engaged in substrates binding. Second, the loop area between KH domain and helicase core was not fully involved in substrate binding. Two mutated proteins, DDX43 Δ140-166 and DDX43 Δ140-172, were constructed with retention of complete function domains but deletion parts of linker, still shared the same Kd values with wild DDX43. Third, KH domain was stronger than helicase core in terms of substrate binding ability, which was proved by DDX43^{F-144} and DDX43^{171-R}. Although both of their affinity with substrates was declined sharply, the Kd value of DDX43^{F-144} was about half of the DDX43^{171-R} when binding to the same substrates. Furthermore, the two RecA-like domains of DDX43 were synergistic in binding substrates, DDX43$^{170-456}$ possessed

RecA-like domain 1, was barely bound to substrates and the same to DDX43[457-R], which occupies another RecA-like domain. Besides the extreme low substrates affinity of DDX43[457-R], protein self-degradation and instability (Figure 3B) also lead to the missing Kd value of DDX43[457-R] in Figure 3C.

3.4 The realization of DDX43 high affinity with substrates requires linked domains

We already known that the deletion of the flexible area between KH domain and helicase core would not affect the binding affinity of DDX43 to substrates. Full length DDX43 tagged with 3C protease fusion domain in flexible area were designed to explore the necessity of two linked function domains for retention of protein' high affinity with substrates. Proteins with 3C protease insertion sites in 140 aa and 157 aa were designed and purified, named 140-3C and 157-3C for short, respectively.

We have validated that two proteins were successfully constructed and can be cleaved by 3C protease (Figure 4A). Series molar ratio of protein to 3C protease were set to explore the optimal ratio of protein to be digested with the least amount of 3C protease at 20°C in 10 min, we found that when the molar ratio was 100 to 1, more than 90% of the fusion protein was cleaved (Figure 4A). Binding experiment clearly showed the differences between the fusion protein and the digested protein in binding to three representing substrates. Fusion proteins were first fully digested with 3C proteases in a ratio of 100:1, following measured the Kd values of binding to different substrates, and compared it with the untreated proteins (Figure 4B, Figure S2), we found that the Kd values of 3C digested proteins increased significantly, suggesting they have lost the high affinity to substrates.

By measuring the anisotropy value using Infinite F200 PRO also exhibited the process about fusion protein gradually losing their high affinity with substrates when increasing amount of protease. The Kd values of fusion proteins binding to the representing substrate were pre-measured (Figure 4B, Figure S2) and we know when the concentration of 140-3C and 157-3C reached about 500 nM, the anisotropy value no longer increase substantially. Thus, 500 nM protein and 5 nM fluorescent substrates were used in experiment. Mixture of the protein and fluorescent substrate were divided to 16 wells into 96-well microplates, following added 3C protease in series of proportion. After harvesting for 10 minutes at 20°C, the anisotropy value of each well were measured. Taking the total amount of 3C protease as X axis and the corresponding decreased anisotropy value as Y axis, here we obtained figure 4C. Each measurement was repeated at least three times. In figure 4C, we found that with the increasing amount of 3C proteases, the anisotropy value gradually decreased and when the amount of 3C protease was 10 nM, that is, the ratio of protein to 3C reached 50 to 1, the anisotropy eventually stabilizing at the same value.

Previously, we have demonstrated that deleting a portion of the flexible region did not affect the substrates affinity of the protein, which suggest that this region might not engaged in substrates binding. This experiment proves that although the intermediate flexible region does not play roles in substrate binding, it helps to achieve the ability of DDX43 in substrate binding through connecting the two function domains.

4. Discussion

In this study, we found that DDX43 was a weak DEAD-box helicase that could not efficiently

unwind dsRNA duplex, as measured by stopped-flow FRET assay (Figure S3). It is reported that DDX43 has the ability to rescue the expression of NRAS by unwinding NRAS siRNA in a DDX43 knockdown melanoma cell line [15], but cannot unwind 16 bp dsRNA *in vitro* [16]. It is common that DEAD-box RNA helicases do not display an unwinding activity *in vitro* [17]. Even if the unwinding activity was observed, the RNA duplex region is generally no longer than 10 bp, those short paired RNAs are the common feature of structured RNAs in cells [18], suggesting that DEAD-box helicases might serve more of a remodeling function [17,19] than a strand-displacement function. Which might also explain the reason DDX43 preferred ssDNA/RNA and required two linked function domain to bind to the substrates, probably because ssDNA/RNA was much more flexible and manageable. The linked domains could contribute to the stability of the proper conformation of DDX43 after binding to single-strand DNA/RNA.

Generally, the function of the protein depends on its structure. We have screened full-length DDX43 and all its truncations mentioned in the text for crystallization, several DNA-DDX43 and RNA-DDX43 complex were also sifted, yet with no harvest. Structure of known DEAD-box proteins revealed that conformational changes of D1 and D2 were occurred upon substrate binding [3,20], which suggest that an appropriate substrate would largely contribute to the solution of protein structure. It is reported that DDX43 was interacted with DDX53 and DHX15 [21,22]. Thus co-crystallization of proteins-DDX43 is an alternative choose. Present data suggest that the precise function of cancer-related DEAD-box proteins, such as DDX3X, may be influenced by their interacting partners and were highly tumor or background-dependent [23,24], which could be the same to DDX43.

Acknowledgement

References

[1] M.E. Fairman-Williams, U.P. Guenther, E. Jankowsky, SF1 and SF2 helicases: family matters, Curr Opin Struct Biol 20 (2010) 313-324.

[2] O. Cordin, J. Banroques, N.K. Tanner, P. Linder, The DEAD-box protein family of RNA helicases, Gene 367 (2006) 17-37.

[3] T. Sengoku, O. Nureki, A. Nakamura, S. Kobayashi, S. Yokoyama, Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa, Cell 125 (2006) 287-300.

[4] M.J. Tauchert, R. Ficner, Structural analysis of the spliceosomal RNA helicase Prp28 from the thermophilic eukaryote Chaetomium thermophilum, Acta Crystallogr F Struct Biol Commun 72 (2016) 409-416.

[5] J.M. Coller, M. Tucker, U. Sheth, M.A. Valencia-Sanchez, R. Parker, The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes, Rna 7 (2001) 1717-1727.

[6] I. Jarmoskaite, R. Russell, RNA helicase proteins as chaperones and remodelers, Annu Rev Biochem 83 (2014) 697-725.

[7] V. Martelange, C. De Smet, E. De Plaen, C. Lurquin, T. Boon, Identification on a human sarcoma of two new genes with tumor-specific expression, Cancer Res 60 (2000) 3848-3855.

[8] S. Gerstberger, M. Hafner, T. Tuschl, A census of human RNA-binding proteins, Nat Rev Genet 15 (2014) 829-845.

[9] G. Nicastro, I.A. Taylor, A. Ramos, KH-RNA interactions: back in the groove, Curr Opin Struct Biol 30 (2015) 63-70.

[10] A. Oddone, E. Lorentzen, J. Basquin, A. Gasch, V. Rybin, E. Conti, M. Sattler, Structural and biochemical characterization of the yeast exosome component Rrp40, EMBO Rep 8 (2007) 63-69.

[11] G. Ambrosini, R. Khanin, R.D. Carvajal, G.K. Schwartz, Overexpression of DDX43 mediates MEK inhibitor resistance through RAS Upregulation in uveal melanoma cells, Mol Cancer Ther 13 (2014) 2073-2080.

[12] J. Lin, Q. Chen, J. Yang, J. Qian, Z.Q. Deng, W. Qian, X.X. Chen, J.C. Ma, D.S. Xiong, Y.J. Ma, C. An, C.Y. Tang, DDX43 promoter is frequently hypomethylated and may predict a favorable outcome in acute myeloid leukemia, Leuk Res 38 (2014) 601-607.

[13] N.N. Liu, X.L. Duan, X. Ai, Y.T. Yang, M. Li, S.X. Dou, S. Rety, E. Deprez, X.G. Xi, The Bacteroides sp. 3_1_23 Pif1 protein is a multifunctional helicase, Nucleic Acids Res 43 (2015) 8942-8954.

[14] J. Shi, W.F. Chen, B. Zhang, S.H. Fan, X. Ai, N.N. Liu, S. Rety, X.G. Xi, A helical bundle in the N-terminal domain of the BLM helicase mediates dimer and potentially hexamer formation, J Biol Chem 292 (2017) 5909-5920.

[15] A.J. Linley, M.G. Mathieu, A.K. Miles, R.C. Rees, S.E. McArdle, T. Regad, The helicase HAGE expressed by malignant melanoma-initiating cells is required for tumor cell proliferation in vivo, J Biol Chem 287 (2012) 13633-13643.

[16] T. Talwar, V. Vidhyasagar, J. Qing, M. Guo, A. Kariem, Y. Lu, R.S. Singh, K.E. Lukong, Y. Wu, The DEAD-box protein DDX43 (HAGE) is a dual RNA-DNA helicase and has a K-homology domain required for full nucleic acid unwinding activity, J Biol Chem 292 (2017) 10429-10443.

[17] J. Banroques, O. Cordin, M. Doere, P. Linder, N.K. Tanner, A conserved phenylalanine of motif IV in superfamily 2 helicases is required for cooperative, ATP-dependent binding of RNA substrates in DEAD-box proteins, Mol Cell Biol 28 (2008) 3359-3371.

[18] N.K. Tanner, P. Linder, DExD/H box RNA helicases: From generic motors to specific dissociation functions, Molecular Cell 8 (2001) 251-262.

[19] E. Jankowsky, H. Bowers, Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases, Nucleic Acids Res 34 (2006) 4181-4188.

[20] J. Xiol, P. Spinelli, M.A. Laussmann, D. Homolka, Z. Yang, E. Cora, Y. Coute, S. Conn, J. Kadlec, R. Sachidanandam, M. Kaksonen, S. Cusack, A. Ephrussi, R.S. Pillai, RNA clamping by Vasa assembles a piRNA amplifier complex on transposon transcripts, Cell 157 (2014) 1698-1711.

[21] E.L. Huttlin, R.J. Bruckner, J.A. Paulo, J.R. Cannon, L. Ting, K. Baltier, G. Colby, F.

Gebreab, M.P. Gygi, H. Parzen, J. Szpyt, S. Tam, G. Zarraga, L. Pontano-Vaites, S. Swarup, A.E. White, D.K. Schweppe, R. Rad, B.K. Erickson, R.A. Obar, K.G. Guruharsha, K. Li, S. Artavanis-Tsakonas, S.P. Gygi, J.W. Harper, Architecture of the human interactome defines protein communities and disease networks, Nature 545 (2017) 505-509.

[22] C. Wan, B. Borgeson, S. Phanse, F. Tu, K. Drew, G. Clark, X. Xiong, O. Kagan, J. Kwan, A. Bezginov, K. Chessman, S. Pal, G. Cromar, O. Papoulas, Z. Ni, D.R. Boutz, S. Stoilova, P.C. Havugimana, X. Guo, R.H. Malty, M. Sarov, J. Greenblatt, M. Babu, W.B. Derry, E.R. Tillier, J.B. Wallingford, J. Parkinson, E.M. Marcotte, A. Emili, Panorama of ancient metazoan macromolecular complexes, Nature 525 (2015) 339-344.

[23] F.V. Fuller-Pace, DEAD box RNA helicase functions in cancer, RNA Biol 10 (2013) 121-132.

[24] L.B. Epling, C.R. Grace, B.R. Lowe, J.F. Partridge, E.J. Enemark, Cancer-associated mutants of RNA helicase DDX3X are defective in RNA-stimulated ATP hydrolysis, J Mol Biol 427 (2015) 1779-1796.

Figure 1. Basic information of purified human DDX43. (A) Sequence scanning and structure conservation prediction of human DDX43. Disorder tendency of each position was the average value of all available sites in the same column provided by Anchor2 (blue) and IUPred2 (red). (B) 12% SDS-PAGE gel of purified DDX43. (C) Size distribution histogram of DDX43 measured by DLS. (D) Size exclusion chromatography analysis of DDX43 by superdex200 10/300 GL. Molecular weight used for the calibration are indicated and marked on the top axis. The insert graph shows the plot of standard log (molecular weight) versus V_x/V_0 value.

Figure 2. Binding characteristics of human DDX43 helicase. (A) DDX43 binds to thirtheen types of substrates. The dissociation constants (Kd) were derived from the fits of experimental curves as described in 'Materials and Methods' section. (B) DDX43 prefers guanine rich sequence $(GT)_6$ than the other three polynucleotides. (C) DDX43 binds to different length ssDNA. The Kd values were summarized in Table S1. (D) Gel filtration profiles of DDX43 (red), DNA (black) and DDX43&DNA mixture (blue). The elution volume of DDX43 and the mixture were labeled in corresponding color.

Figure 3. Helicase core and KH domain are both essential for the full binding affinity of DDX43. (A) Schematic representation of DDX43 (1–648 aa) and eight truncated DDX43 proteins. Numbers represent the sites of amino acid, D1 and D2 are short for RecA-like domain 1 and 2, respectively. (B) 12% SDS-PAGE gel of DDX43 and eight truncated proteins. DDX43[457-R] (black arrow) was degraded spontaneously (red arrow). (C) Histogram comparative analysis of Kd values of eight DDX43 proteins with 12 nt $(GT)_6$, 24 nt ssDNA and 22 nt ssRNA, respectively. No binding affinity was observed of

DDX43[457-R]. Kd values were summarized in Table S1.

Figure 4. Physical connection between helicase core and KH domain is necessary for DDX43 bind to substrates with high affinity. (A) 12% SDS-PAGE gel of 140-3C [upper] and 157-3C [below] digestion with 3C protease in indicated proportion. Schematic representation of each protein was lie down the gel, the orange line represent the insertion site of 3C. (B) Histogram comparative of Kd values between wild type 140-3C and digested 140-3C terms on binding substrates [upper]. Same to 157-3C [below]. (C) After binding to 12 nt $(GT)_6$ (red), 24 nt ssDNA (blue) and 22 nt ssRNA FD (black), with the increasing amount of 3C protease, the anisotropy value of 140-3C&substrate [upper] and 157-3C&substrate [below] were decreased.
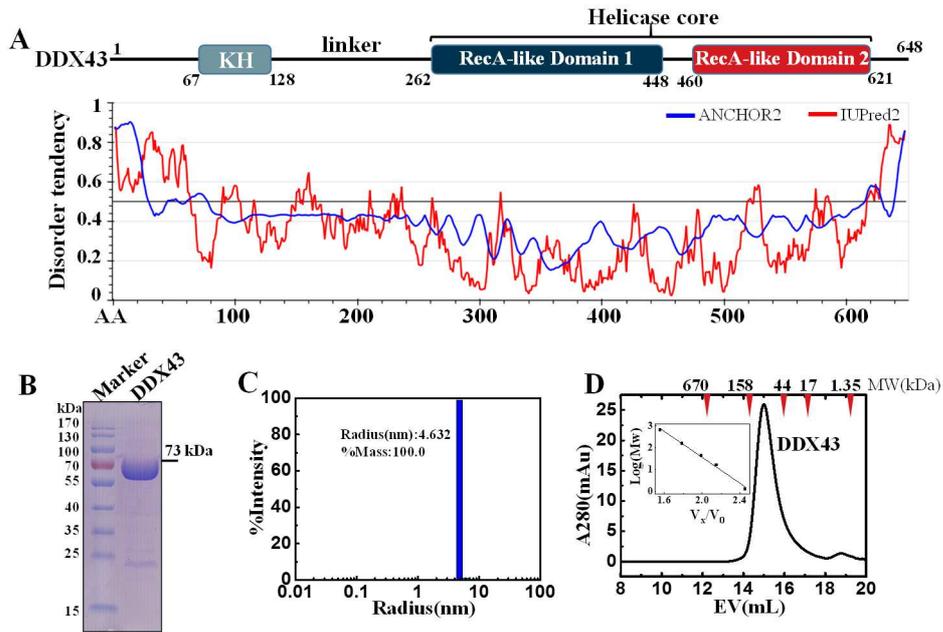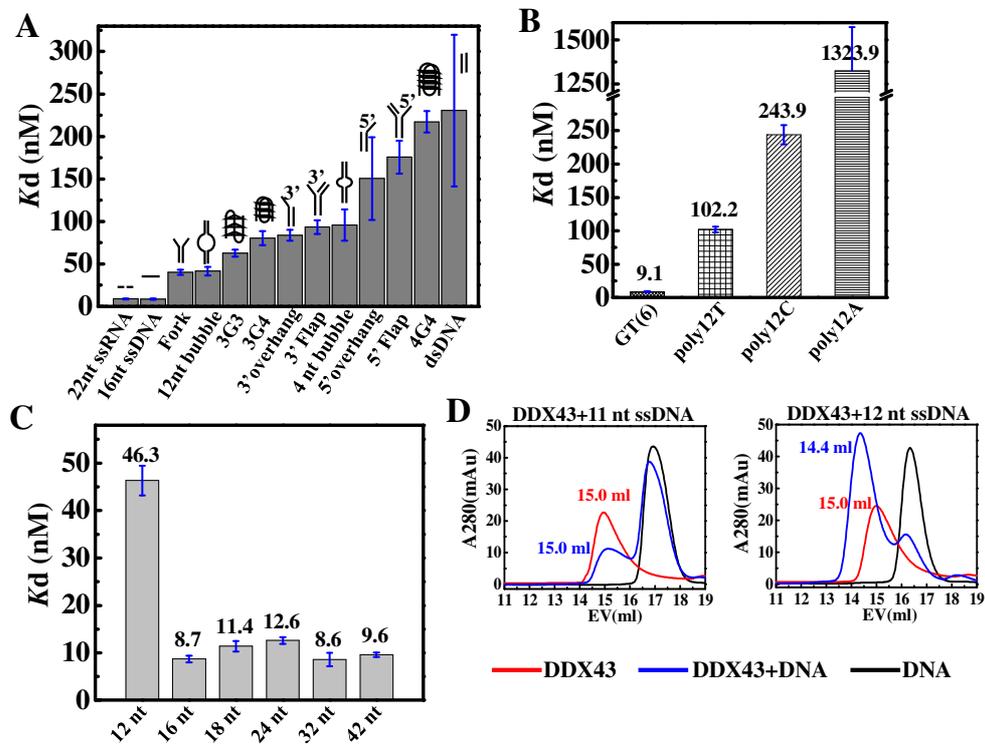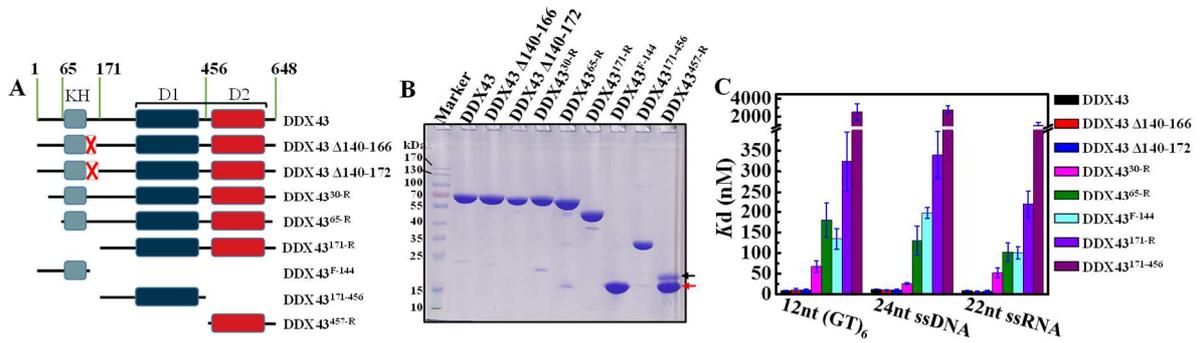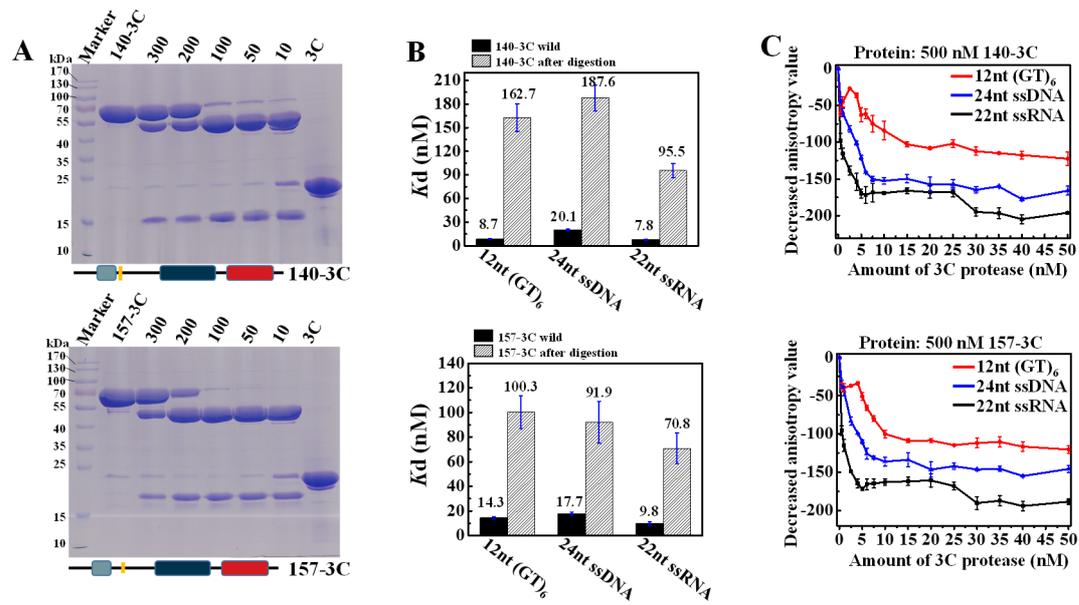
Figure 1

Figure 2

Figure 3

Figure 4

Supplementary data

Table S1. Substrates used in the binding and unwinding experiment

Name	Structure	Sequence	Name	Sequence
D20S12 Fork	20bp / 12nt	CTCTGGCCGTCTTACGGTCGCTCTGCTCGACG-F * CGTCGAGCAGAGCGACCGTATTATTTTTTTTT	22 nt ssRNA (FD)	UUUUUUUUUUCUCUGCUCGACG-F
12nt bubble	14bp / 14bp / 12nt	CCATGCAGCTGTCAGTCCATTGTCATGCTAGGCCTACTGC-F GCAGTAGGCCTAGCTACTGTTACCTGTGACAGCTGCATGG	18bp dsDNA	TGGCGACGGCAGCGAGGC-F GCCTCGCTGCCGTCGCCA
3G3		gggTTAgggTTAggg-F	(GT)6	GTGTGTGTGTGT-F
3G4		TgggTTAgggTTAggggTTAggg-F	10 nt ssDNA	CTCGTGCTCG-F
3' overhang	20bp / 3' / 12nt	CACTGGCCGTCTTACGGTCGCTCTGCTCGACG-F CGACCGTAAGACGGCCAGTG	12 nt ssDNA	ACGGATGTCTAA-F
3' Flap	20bp / 3' / 12bp	CACTGGCCGTCTTACGGTCGCTCTGCTCGACG-F CGTCGAGCAGAGCGACCGTATTATTTTTTTTT AGACGGCCAGTG	16 nt ssDNA	CTCTGCTCGACGGATT-F
4nt bubble	18bp / 18bp / 4nt	CCATGCAGCTGTCAGTCCATTGTCATGCTAGGCCTACTGC-F GCAGTAGGCCTAGCATGAGTTAGGACTGACAGCTGCATGG	18 nt ssDNA	GCCTCGCTGCCGTCGCCA-F
5' overhang	20bp / 20nt 5'	CACTGGCCGTCTTACGGTCGAAAAAAAAAAACTCTGCTCGACG-F CGTCGAGCAGAGTTTTTTTT	24 nt ssDNA	GCCCTGGTGCCGACCAACGAAGGT-F
5' Flap	20bp / 12bp 5'	CACTGGCCGTCTTACGGTCGCTCTGCTCGACG-F CGTCGAGCAGAGCGACCGTATTATTTTTTTTT AAAAAAAAATAA	32 nt ssDNA	CACTGGCCGTCTTACGGTCGCTCTGCTCGACG-F
4G4		ggggTTAggggTTAggggTTAgggg-F	42 nt ssDNA	CACTGGCCGTCTTACGGTCGAAAAAAAAAAACTCTGCTCGACG-F
FA-FD	12bp / 10nt	HF*-CGUCGAGCAGAGUUUUUUUUUU UUUUUUUUUUCUCUGCUCGACG-F	11 nt ssDNA	CGAGCACTGCG

*F, Fluorescein; *HF, Hexachlorofluorescein;

Table S2. Binding parameters for different substrates*

P \ S	22nt ssRNA	16nt ssDNA	Fork	12nt bubble	3G3	3G4	3' overhang	3' Flap	4 nt bubble
DDX43	8.9±0.6	8.7±0.7	40.3±3.0	41.6±5.1	62.6±4.1	80.4±8.3	83.8±6.5	93.5±8.1	95.9±18.4

P \ S	5'overhang	5' Flap	4G4	dsDNA	(GT)6	poly12T	poly12C	poly12A
DDX43	150.6±48.6	175.7±19.5	217.2±12.5	230.6±89.2	9.1±0.6	102.2±4.4	243.9±14.5	1323.9±250.8

P \ S	10 nt ssDNA	12ntssDNA	16 nt ssDNA	18 nt ssDNA	24 nt ssDNA	32 nt ssDNA	42 nt ssDNA
DDX43	—	46.3±3.1	8.7±0.7	11.4±1.1	12.6±0.7	8.6±1.4	9.6±0.5

S \ P	Δ140-166	Δ140-172	30-R	65-R	F-144	171-R	170-456	457-R
12nt (GT)6	10.2±4.0	11.3±1.7	67.3±13.5	180.1±40.5	134.7±25.1	324.0±71.1	2494.2±969.7	—
24nt ssDNA	10.0±1.4	12.3±0.7	26.6±2.3	129.9±35.4	197.6±13.2	339.6±56.6	2762.0±452.1	—
22nt ssRNA	6.4±1.6	6.9±2.6	52.2±11.1	102.4±22.5	100.5±14.5	219.7±33.1	982.0±343.7	—

The mean value and standard deviations were determined from 3-5 independent measurements.
Unit, Kd (nM); —, no binding affinity was observed

Figure S1 The corresponding binding graphs of DDX43 to various substrates measured by steady-state fluorescence anisotropy. (A) DDX43 binds to thirtheen substrates. (B) DDX43 prefers guanine rich sequence-(GT)6 than the other three polynucleotides. (C) DDX43 binds to different

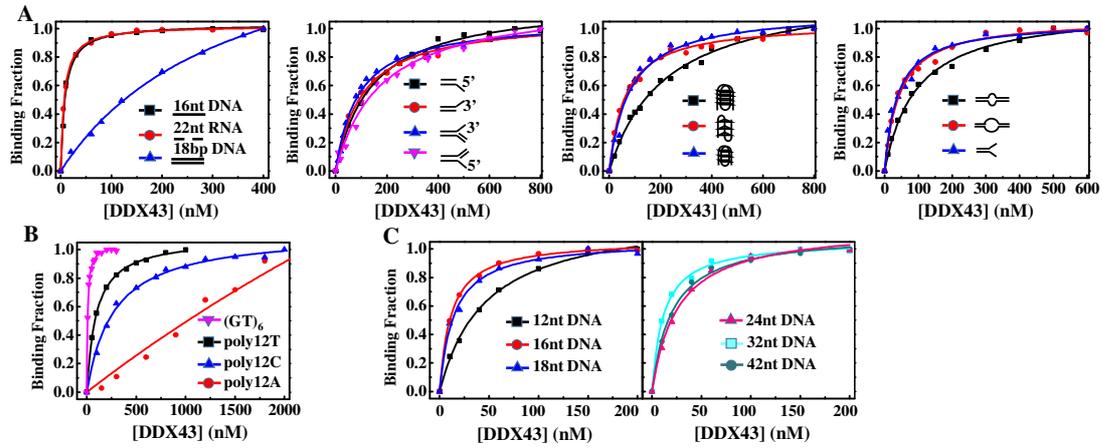length ssDNA. 5 nM of fluorescein-labeled substrate was titrated by DDX43 at 20°C.

Figure S2 The binding graphs of wild 3C-insertion DDX43 (black line) and digested 3C-insertion

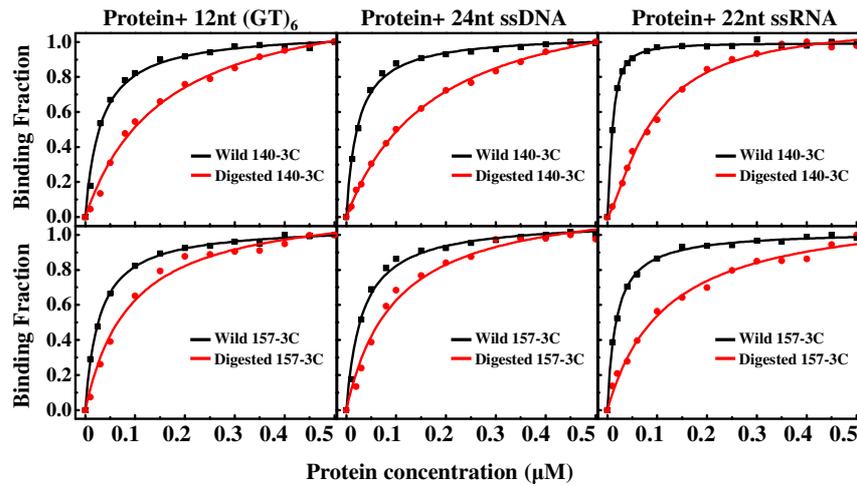DDX43 (red line) to three different substrates.

Figure S3 DDX43-catalyzed unwinding kinetics with forked FA-FD RNA substrates measured with

stopped-flow.