Supplemental Information

Molecular Cell, *Volume 41*

Coupled 5′ Nucleotide Recognition and Processivity in Xrn1-Mediated mRNA Decay

Martin Jinek, Scott M. Coyle, and Jennifer A. Doudna

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Protein expression and purification

The design of constructs amenable for crystallization was guided by hydrophobic cluster analysis (Callebaut et al., 1997), which indicated that the C-terminal ~500 residues in Xrn1 proteins lack tertiary structure. The sequence corresponding to residues 1–1141 of *Drosophila melanogaster* Xrn1 (DmXrn1) was amplified from a full-length cDNA clone (LD22664, Drosophila Gene Collection Release 1, Berkeley Drosophila Genome Project) and subcloned in frame with a C-terminal hexahistidine tag into the pFastBac1 baculovirus shuttle vector. Point mutations were introduced using the QuikChange site-directed mutagenesis kit (Stratagene) and verified by DNA sequencing. Recombinant baculoviruses were generated in Sf9 cells using the Bac-to-Bac kit (Invitrogen) according to manufacturer's recommendations. For protein expression, Sf9 cells were grown in suspension culture in ESF 921 serum-free medium (Expression Systems) and infected at a density of 2.0 x 10^6 ml^{-1}. Cells were harvested 60 hr post infection and lysed by sonication in 20 mM Tris pH 8.5, 500 mM NaCl, 5 mM imidazole, 0.4% Triton X-100 (supplemented with protease inhibitors). The protein was purified by Ni-NTA affinity (Qiagen) and Q-sepharose ion exchange chromatographic steps, followed by size exclusion on a Superdex-200 column (GE Life Sciences) in 20 mM HEPES pH 7.5, 150 mM KCl, 4 mM $MgCl_2$, 1 mM TCEP. For protein samples used in enzymatic assays, the size exclusion was omitted and the proteins were instead dialyzed against 20 mM HEPES pH 7.5, 150 mM KCl, 1 mM TCEP after ion exchange chromatography. The wild type DmXrn1(1-1141) protein displays exonuclease activity *in vitro* that is indistinguishable from that of full-length DmXrn1 (data not shown). To produce selenomethionine-substituted Xrn1 protein, Sf9 cells were propagated for two passages in methionine-free ESF 921 medium (Expression systems) and infected at a density of 1x10^6 ml^{-1}. At 8 hr and 24 hr time points post infection, the medium was supplemented with 100 $mg{\cdot}l^{-1}$ L-selenomethionine. Cells were harvested 48 hr post infection and the protein was purified as for the native protein.

Complex crystallization

The Xrn1-substrate complex was reconstituted by incubating Xrn1ΔC-D207A mutant protein (in size exclusion chromatography buffer at a final concentration of 4 $mg{\cdot}ml^{-1}$) with 100 μM (final concentration) synthetic 5'-phosphorylated dT_{11} oligonucleotide (Integrated DNA Technologies) for 10 min on ice. The complex was crystallized using the hanging drop vapour diffusion method at 18°C. The reconstituted complex was combined in a 1:1 (v/v) ratio with 100 mM MES pH 6.2, 1% (v/v) PEG 300 on ice. The mixture was centrifuged to remove the resultant slight precipitate and 2 ml drops of the supernatant were equilibrated against a reservoir solution containing 50 mM MES pH 6.5. Plate-like crystals of the complex typically grew within 24 hr and were

propagated by microseeding for optimal quality. Crystals were briefly cryoprotected in a buffer containing 100 mM MES pH 6.5, 50 mM KCl, 4 mM $MgCl_2$ and 30% (v/v) ethylene glycol prior to flash cooling in liquid nitrogen. The crystals diffracted up to 3.2 Å using synchrotron X-ray radiation, belonged to space group $P4_22_12$, and contained one Xrn1 molecule in the asymmetric unit.

Structure determination

X-ray diffraction data were collected at beamline 8.2.2 of the Advanced Light Source (Lawrence Berkeley National Laboratory) and processed with XDS (Kabsch, 2010). Although the native data gave a robust molecular replacement solution in Phaser (McCoy et al., 2007) using a polyalanine search model derived from the structure of Rat1 in the Rat1-Rai1 complex (PDB 3FQD), the resulting difference electron density maps did not reveal any features attributable to the additional protein regions unique to Xrn1. The structure was subsequently solved using single-wavelength anomalous dispersion (SAD) data measured at the selenium K-edge (peak wavelength 0.979491 Å) from crystals containing selenomethionine-substituted Xrn1-substrate complex. Nineteen selenium sites were located by difference Fourier analysis using initial molecular replacement phases. The initial MR phases were then discarded and the selenium positions were used in phase calculations carried out with AutoSHARP (Vonrhein et al., 2007). The phases were further improved by solvent flipping and phase extension using native data extending to 3.2 Å resolution. The resulting electron density maps revealed clear regions of density corresponding to the Xrn1-specific regions as well as the nucleic acid in the active site of the protein (Fig. 2A). The molecular model was built using COOT (Emsley and Cowtan, 2004) and refined in Phenix.refine (Adams et al., 2010) against the 3.2 Å resolution native dataset. Simulated annealing was used in the initial round of refinement; subsequent rounds utilized individual coordinate refinement combined with TLS and grouped atomic displacement parameter (ADP) refinement (two ADP groups per residue). Secondary structure restraints were maintained throughout the refinement. The final model has a crystallographic R_{work} factor of 22.5% and an R_{free} factor of 27.1%, with good stereochemistry (Supplementary Table 1). The quality of the model was assessed using MolProbity (Chen et al., 2010); the final model has 94.7% of residues in the favoured region of the Ramachandran plot, 5.3% in the additionally allowed regions, and no outliers. The final model contains residues 1-1140 with the following exceptions. No electron density could be observed for regions 354-401, 897-909 and 1115-1126. In addition, a number of loops could not be built with confidence, although some residual electron density is present in F_o-F_c maps: residues 47-52, 119-125, 801-802, 836-841, 965-980 and 1027-1030. Molecular graphics illustrations were generated using PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC).

RNase protection assays

2.0 μg DmXrn1ΔC (740 nM final concentration) was incubated with 4 pmol [^{32}P] 5'-end-labeled RNA oligonucleotide (sequence $G(CU)_{10}C$) in 20 mM HEPES pH 7.5, 100 mM KCl, 4 mM $MgCl_2$, 5% (v/v) glycerol, 1 mM DTT in a total volume of 20 μl for 30 min at 0^{o}C. The mixtures were subsequently digested with 2.5 U RNase T1 and 1 μg RNase A (Fermentas) for 5 min at 20^{o}C. 2 μl aliquots were removed and quenched by the addition of 8 μl loading buffer (100 mM HEPES 7.5, 5 mM EDTA in formamide). The samples were resolved on a 16% denaturing polyacrylamide gel and visualized by phosphorimaging.

Nuclease activity assays

A ~550 nt long RNA harbouring five BoxB hairpins was prepared by in vitro transcription using T7 RNA polymerase and [α-^{32}P] ATP and purified using a RNA Clean & Concentrator-5 spin column (Zymo Research). The 5'-terminal triphosphate in the RNA was converted to monophosphate by incubation with tobacco acid pyrophosphatase (Epicentre Biotechnologies) and the RNA was repurified. 3'-end-labeled oligonucleotide substrates were prepared by labelling a 22-mer and 11-mer synthetic oligonucleotides, sequence $G(CU)_{10}C$ and $G(CU)_5C$, with [^{32}P]-pCp using T4 RNA ligase. The oligonucleotides were 5'-phosphorylated using T4 polynucleotide kinase. Unincorporated label was removed using a G25 spin column (GE Life Sciences). Exonuclease activity assays were carried out in a buffer containing 20 mM HEPES pH 7.5, 100 mM KCl, 4 mM $MgCl_2$, 5% (v/v) glycerol, 1 mM DTT. 1.0 μg of wild-type or mutant Xrn1 (370 mM final concentration) was incubated with 5 pmol RNA substrate in a volume of 20 μl at 20^{o}C. At indicated time points, 5 μl aliquots were removed and quenched by the addition of 15 μl loading buffer (100 mM HEPES 7.5, 5 mM EDTA in formamide). 4 μl samples were resolved on 8% (500-nt body labelled RNA) or 16% (22-mer and 11-mer oligonucleotides) denaturing polyacrylamide gels and visualized by phosphorimaging (Storm, GE Life Sciences).

For exonuclease assays using duplex substrates, the following oligonucleotides were used:

<u>RNA strands</u>

5'-AACCCCACCACCAUCACUU-3'

5'-AAAAACCCCACCACCAUCACUU-3'

5'-AAAAAAAACCCCACCACCAUCACUU-3'

<u>DNA strand</u>

5'-dAdAdGdTdGdAdTdGdGdTdGdGdTdGdGdGdG-3'

RNA oligonucleotides (50 pmol) were 3'-end labeled using [5'-^{32}P] pCp and T4 RNA ligase and subsequently 5'-phosphorylated using T4 polynucleotide kinase. The RNA oligonucleotides (adjusted to 2 μM with cold 5'-phosphorylated oligonucleotide) were annealed with 2-fold excess of the DNA oligonucleotide (4 μM) by heating to 95 ^{o}C for 3 min and cooling to room temperature. Assays were carried out in 20 mM HEPES pH 7.5, 100 mM KCl, 4 mM $MgCl_2$, 5% (v/v) glycerol, 1 mM DTT. 0.1 μg of wild-type or mutant Xrn1 (37 nM final concentration) was incubated with 200 nM duplex substrate in a volume of 20 μl at 20 ^{o}C for 2 minutes. 5 μl aliquots were removed and quenched by the addition of 15 μl loading buffer. 4 μl samples were resolved on a 16% denaturing polyacrylamide gels and visualized by phosphorimaging.

SUPPLEMENTAL REFERENCES

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res *38 Suppl*, W529-533.

Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B., and Mornon, J.P. (1997). Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cell Mol Life Sci *53*, 621-645.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr *66*, 12-21.

Devos, J.M., Tomanicek, S.J., Jones, C.E., Nossal, N.G., and Mueser, T.C. (2007). Crystal structure of bacteriophage T4 5' nuclease in complex with a branched DNA reveals how flap endonuclease-1 family nucleases bind their substrates. J Biol Chem *282*, 31713-31724.

Douangamath, A., Filipp, F.V., Klein, A.T., Barnett, P., Zou, P., Voorn-Brouwer, T., Vega, M.C., Mayans, O.M., Sattler, M., Distel, B., *et al.* (2002). Topography for independent binding of alpha-helical and PPII-helical ligands to a peroxisomal SH3 domain. Mol Cell *10*, 1007-1017.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J Mol Biol *233*, 123-138.

Hwang, K.Y., Baek, K., Kim, H.Y., and Cho, Y. (1998). The crystal structure of flap endonuclease-1 from Methanococcus jannaschii. Nat Struct Biol *5*, 707-713.

Kabsch, W. (2010). Xds. Acta Crystallogr D Biol Crystallogr *66*, 125-132.

Kaieda, S., Matsui, C., Mimori-Kiyosue, Y., and Ikegami, T. (2010). Structural basis of the recognition of the SAMP motif of adenomatous polyposis coli by the Src-homology 3 domain. Biochemistry *49*, 5143-5153.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr *60*, 2256-2268.

Ma, J.B., Ye, K., and Patel, D.J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature *429*, 318-322.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

Mueser, T.C., Nossal, N.G., and Hyde, C.C. (1996). Structure of bacteriophage T4 RNase H, a 5' to 3' RNA-DNA and DNA-DNA exonuclease with sequence similarity to the RAD2 family of eukaryotic proteins. Cell *85*, 1101-1112.

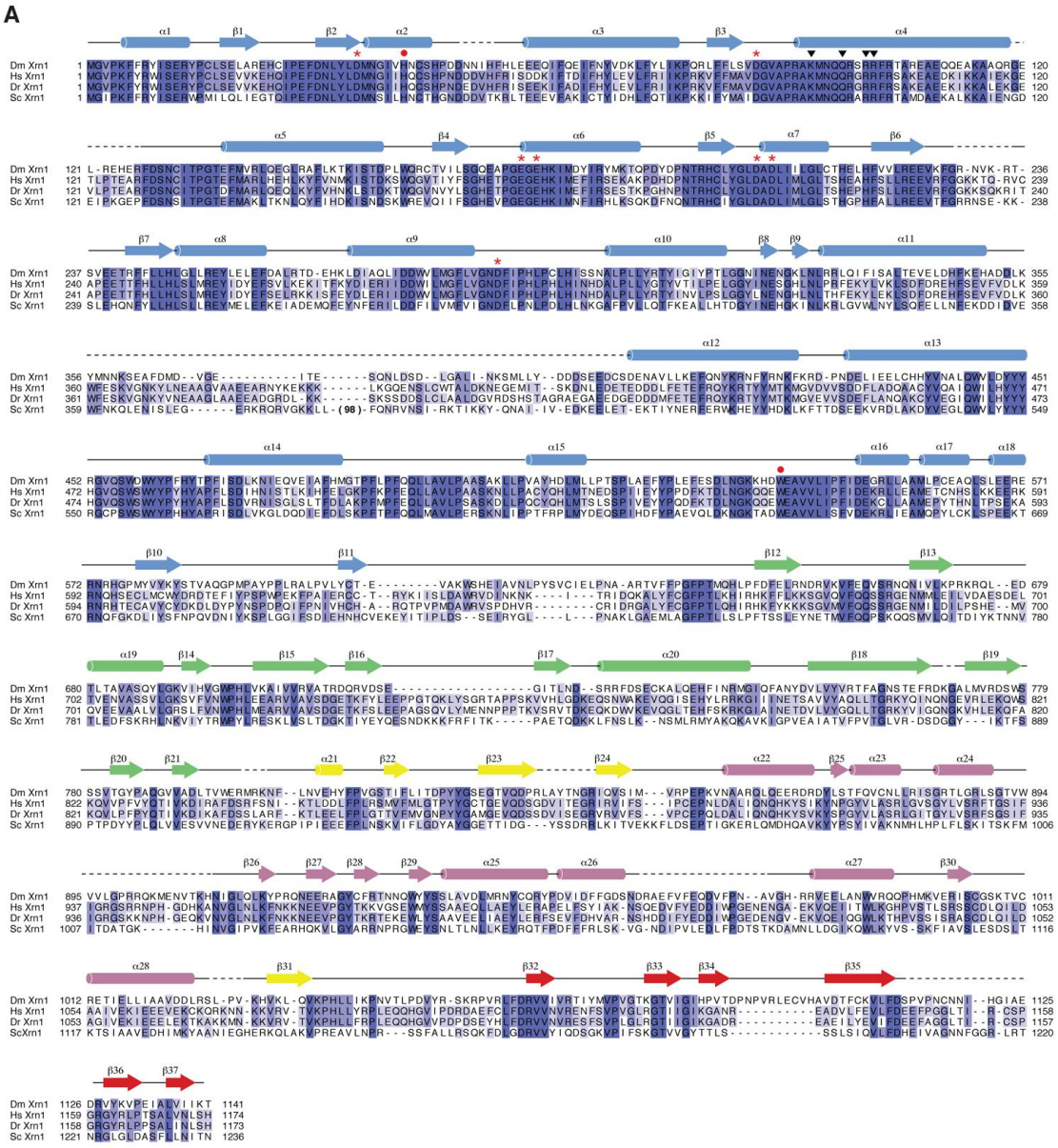
Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol *302*, 205-217.

Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007). Automated structure solution with autoSHARP. Methods Mol Biol *364*, 215-230.

Xiang, S., Cooper-Morgan, A., Jiao, X., Kiledjian, M., Manley, J.L., and Tong, L. (2009). Structure and function of the 5'-->3' exoribonuclease Rat1 and its activating partner Rai1. Nature *458*, 784-788.

Figure S1.

A


α1 β1 β2 α2 α3 β3 α4
Dm Xrn1 1 MGVPKFFRYISERYPCLSELAREHCIPEFDNLYLDMNGIVRNCSHPDDNNIHFHLEEEQITQEITFNYVDKLFYLIKPQRLFFLSVDGVAPRAKMNQQRSRRFRTAREAEQQEAKAAQRGE 120
Hs Xrn1 1 MGVPKFFRYISERYPCLSEVVKEHQIPEFDNLYLDMNGIIHQCSHPNDDDVHFRISDDKITFTDIFHYLEVLFRIIKPRKVFFMAVDGVAPRAKMNQQRGRRFRSAKEAEDKIKKAIEKGE 120
Dr Xrn1 1 MGVPKFFRYISERYPCLSEVVKEHQIPEFDNLYLDMNGIIHQCSHPNDEDVHFRISEKIFADIFHYLEVLFRIIKPRKVFFMAVDGVAPRAKMNQQRGRRFRSAKEAEEKIKKALEKGE 120
Sc Xrn1 1 MGIPKFFRYISERWEMILQLIEGTQIPEFDNLYLDMNSILHNCTHGNDDDVTKRLTEEEVEAKLCTYIDHLFQTIKPKKIFYMAIDGVAPRAKMNQQRARRFRTAMDAEKALKKAIENGD 120
α5 β4 α6 β5 α7 β6
Dm Xrn1 121 L-REHERFDSNCITPGTEFMVRLQEGIRAFLKTKISTDPLWQRCTVILSGQEAPGEGEHKIMDYIRYMKTQPDYDPNTRHCLYGLDADLIILGLCTHELHFVVLREEVKFGR-NVK-RT- 236
Hs Xrn1 121 TLPTEARFDSNCITPGTEFMARLHEHLKYFVNMKLSTDKSWQGVTIYFSGHETPGEGEHKIMEFIRSEKAKPDHDPNTRHCLYGLDADLIMLGLTSHEAHFSLLREEVRFGGKKTQ-RVC 239
Dr Xrn1 121 VLPTEARFDSNCITPGTDFMARLQEQLKYFVHNKLSTDKTWQGVNVYLSGHETPGEGEHKIMEFIRSESTKFGHNPNTRHCLYGLDADLIMLGLTSHEPHFSLLREEVRFGGKKSQKRIT 240
Sc Xrn1 121 EIPKGEPFDSNSITPGTEFMAKLTKNLQYFIHDKISNDSKMREVQIIFSGHEVPGEGEHKIMNFIRHLRSQKDFNQNTRHCIYGLDADLIMLGLSTLGPHFALLREEVTFGRRNSE-KK- 238
β7 α8 α9 α10 β8 β9 α11
Dm Xrn1 237 SVEETRFFLHLGILREYLELEFDALRTD-EHKLDIAQLIDDWVIMGFLVGNDFIPHLPCLHISSNALPLLYRTYIGIYPTLGGNINENGKLNLRRLQIFISALTEVELDHFKEHADDLK 355
Hs Xrn1 240 APEETTFHLLHLSLMREYIDYERSVLKEKITEKYDIERIIDDWVIMGFLVGNDFIPHLPHLHINHDALPLLYGTYVTILPELGGYINESGHLNLPRFEKYLEKLSDFDREHFSEVFVDLK 359
Dr Xrn1 241 APEETTFHLLHLSLMREYIDYERSELRKKISEEYDLERIIDDWVIMGFLVGNDFIPHLPHLHINHDALPLLYRTYINVLPSLGGYLNENGHLNLTHFEKYLEKLSDFDREHFSEVFVDLK 360
Sc Xrn1 239 SLEHQNFYLLHLSLLREYMELEFKEIADEMQFEYNFERILDDFILVMEVIGNDELPNLRDLHLNKGAFPVLLQIFKEALLHTDGYINEHGKINLKRLGVVWLNYLSQFELLNFEKDDIDVE 358
α12 α13
Dm Xrn1 356 YMNNKSEAFDMD--VGE----------ITE------SQNLDSD--LGALI-NKSMLLY--DDDSEEDCSDENAVLLKEFFQNYKRNFYRNKFKRD-PNDELIEELCHHYVNALQWVLDYYY 451
Hs Xrn1 360 WFESKVGNKYLNEAAGVAAEEARNYKEKKK------LKGQENSLCWTALDKNEGEMIT--SKDNLEDETEDDDLFETEFRQYKRTYYMTKMGVDVVSDDFLADQAACYVQAIQWILHYYY 471
Dr Xrn1 361 WFESKVGNKYVNEAAGLAAEEADGRDL-KK------SKSSDDSLCLAALDGVRDSHSTAGRAEGAEEDGEDDDMFETEFRQYKRTYYMTKMGVEVVSDEFLANQAKCYVEGIQWILHYYY 473
Sc Xrn1 359 WFNKQLENISLEG------ERKRQRVGKKLL-(98)-FQNRVNSI-RKTIKKY-QNAI-IV-EDKEELET-EKTIYNERFEERWKHEYYHDKLKFTTDSEEKVRDLAKDVVEGLQWVLYYYY 549
α14 α15 α16 α17 α18
Dm Xrn1 452 RGVQSWDWYYPFHYTPFISDLKNIEQVEIAFHMGTPFLPFQDLLAVLPAASAKLLPVAYHDTMLLPTSFLAEFYPLEFESDLNGKKHDWEAVVLIPFIDEGRILLAAMLPCEAQLSLEERE 571
Hs Xrn1 472 HGVQSWSWYYPYHYAPFLSDIHNISTLKIHFELGKPFKPFEQLLAVLPAASKNLLPACYQHLMTNEDSPTIEYYPPDFKTDLNGKQQEWEAVVLIPFIDEKRLLEAMETCNHSLKKEERK 591
Dr Xrn1 474 HGVQSWSWYYPYHYAPFLSDVRNISGLSLTFDLAKPFMPFEQLLAVLPSASKDLLPQCYQHLMTSLSSPIVEYYPQDFKTDLNGKQQEWEAVVLIPFIDEKCLLAAMEPYTHNLTPSEKA 593
Sc Xrn1 550 RGCPSWSWYYPHHYAPRISDLVKGLDQDIEFDLSKPETPFQQLMAVLPERSKNLIPPTFRPLMYDEQSPIHDFYPAEVQLIKNGKTADWEAVVLISFVDEKRLIEAMQPYLCKLSPEEKT 669
β10 β11 β12 β13
Dm Xrn1 572 RNRHGPMYVYKYSTVAQGPMPAYPPLRALPVLYCT-E--------VAKWSHEIAVNLPYSVCIELPNA-ARTVFPGFPTMQHLPFDFELRNDRVKVFEQVSRNQNIVLKPRKRQL--ED 679
Hs Xrn1 592 RNQHSECLMCWYDRDTEFIYPSPWPEKFPAIERCC-T--RYKIISLDAWRVDINKNK----I---TRIDQKALYFCGFPTLKHIRHKFFLKKSGVQVFQQSSRGENMMLEILVDAESDEL 701
Dr Xrn1 594 RNRHTECAVYCYDKDLDYPYNSPDPQIFPNIVHCH-A--RQTPVPMDAWRVSPDHVR--------CRIDRGALYFCGFPTLQHIRHKEYKKKSGVMVFQQSSRGENMILDILPSHE--MV 700
Sc Xrn1 670 RNQFGKDLIYSFNPQVDNIYKSPLGGIFSDIEHNHCVEKEYITIPLDS--SETRYGL----L---PNAKLGAEMLAGFPTLLSLPFTSSLEYNETMVFQQPSKQQSMVLQITDIYKTNNV 780
α19 β14 β15 β16 β17 α20 β18 β19
Dm Xrn1 680 TLTAVASQYLGKVIHVGNPHLVKAIVVRVATRDQRVDSE----------GITLND--SRRFDSECKALQEHFINRMGIQFANYDVLVYVRTFAGNSTEFRDKGALMVRDSWS 779
Hs Xrn1 702 IVENVASSVLGKSVFVNMPHLEEARVVAVSDGETKFYLEEPPGTQKLYSGRTAPPSKVVHLGDKEQSNWAKEVQGISEHYLRRKGIITNETSAVVYAQLLTGRKYQINQNGEVRLEKQVS 821
Dr Xrn1 701 QVEEVAALVLQRSLFVNMPHLEMARVVAVSDGETKFSLEEPAGSQVLYMENNPPPTKVSRVTDKEQKDWKEVQGLTEHFSKRKGIAINETDVLLVGQLLTGRKYVIGQNGKVHLEKQFA 820
Sc Xrn1 781 TLEDFSKRHLNKVIYTRLMYLRESKLLSLTDGKTIYEYQESNDKKKFRFITK-----PAETQDKKLFNSLK--NSMLRMYAKCKAVKIGPVEAIATVFPVTGLVR-DSDGIY---IKTFS 889
β20 β21 α21 β22 β23 β24 α22 β25 α23 α24
Dm Xrn1 780 SSVTGYPADGVVADLTVWERMRKNF--LNVEHYFRVGSTIFLITDPYYGSEGTVQDPRLAYTNGRIQVSIM---VRPEPKVNAARDLQEERDRDRVLSTFQVCNLLRISGRTLGRLSGTVW 894
Hs Xrn1 822 KQVVPFVYQTIVKDIRAFDSRFSNI--KTLDDLFFLRSMVFMLGTPYYGCTGEVQDSGDVITEGRIRVIFS---IPCEPNLDALIQNQHKYSIKVNPGYVLASRLGVSGYLVSRFTGSIF 936
Dr Xrn1 821 KQVLPFPYQTIVKDIKAFDSSLARF--KTLEELFPLGTTVFMVGNPYYGAMGEVQDSSDVISEGRVRVVFS---VPCEPQLDALIQNQHKYSVKVSPGYVLASRLGITGYLVSRFSGSIF 935
Sc Xrn1 890 PTPDYYPLQLVVESVVNEDERYKERGPIPIEEEFFLNSKVIFLGDYAYGGETTIDG---YSSDRRLKITVEKKFLDSEPTIGKERLQMDHQAVKMYPSYIVAKNMHLHPLFLSSKITSKFM 1006
β26 β27 β28 β29 α25 α26 α27 β30
Dm Xrn1 895 VVLGPRRQKMENVTKHNIGLQLKYPRQNEERAGYCFRTNNQMYSSLAVDLMRNMCQRYPDVIDFFGDSNDRAEFVFEQDVFFN--AVGH-RRVEELANWVRQQPHMKVERISGGSKTVC 1011
Hs Xrn1 937 TGRGSRRNPH-GDHKANVGLNLKFNKKNEEVPGYTKKVGSEVMYSSAAEQLLAEYLERAPELFSYIAK-NSQEDVFYEDDIWPGENENGA-EKVQETITWLKGHPVSTLSRSSCDLQILD 1053
Dr Xrn1 936 IGRGSKKNPH-GEQKVNVGLNLKFNKKNEEVPGYTKRTEKEWLYSAAVEELIAEYLERFSEVFDHVAR-NSHDDIFYEDDIWPGEDENGV-EKVQEIQGWLKTHPVSSISRASCDLQILD 1052
Sc Xrn1 1007 ITDATGK-------HINVGIPVKFEARHQKVLGYARRNPRGWEYSNLTLNLLKEYRQTFPDFFRLSK-VG-NDIPVLEDLFFDTSTKDAMNLLDGIKQWLKYVS-SKFIAVSLESDSLT 1116
α28 β31 β32 β33 β34 β35
Dm Xrn1 1012 RETIELLIAAVDDLRSL-PV-KHVKL-QVKPHLLIKPNVTLPDVYR-SKRPVRLFDRVVIVRTIYMVPVGTKGTVIGIHPVTDPNPVRLECVHAVDTFCKVLFDSPVPNCNNI--HGIAE 1125
Hs Xrn1 1054 AAIVEKIEEEVEKCKQRKNN-KKVRV-TVKPHLLYRPLEQQHGVIPDRDAEFCLFDRVVNVRENFSVPVGLRGTIIGIKGANR----------EADVLFEVLFDEEFPGGLTI--R-CSP 1158
Dr Xrn1 1053 AGIVEKIEEELEKTKAKKMN-KKVRV-TVKPHLLFRPLEQQHGVVPDPDSEYHLFDRVVNVRESFSVPLGLRGTIIGIKGADR----------EAEILYEVIFDEEFAGGLTI--R-CSP 1157
ScXrn1 1117 KTSIAAVEDHIMKYAANIEGHERKQLAKVPREAVLNER---SSFALLRSQKFDLGDRVVYIQDSGKVPIFSKGTVVGYTTLS----------SSLSIQVLFDHEIVAGNNFGGR-LRT 1220
β36 β37
Dm Xrn1 1126 DRVYKVPEIALVIIKT 1141
Hs Xrn1 1159 GRGYRLPTSALVNLSH 1174
Dr Xrn1 1158 GRGYRLPPSALINLSH 1173
Sc Xrn1 1221 NRGLGLDASFLLNITN 1236

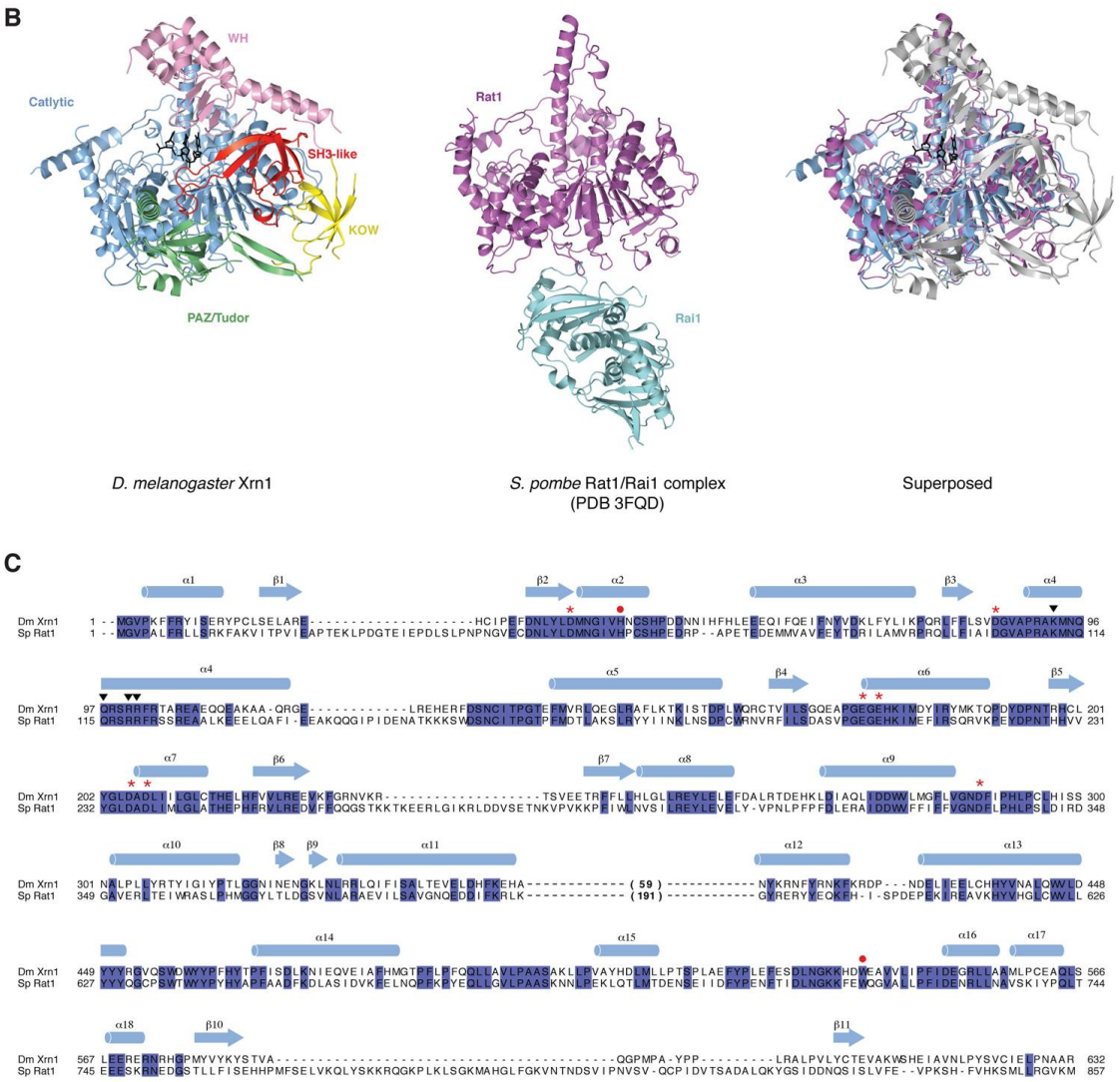


Figure S1, related to Figure 1. Sequence and Structural Homology of DmXrn1

Amino acid sequences of *Drosophila melanogaster* (Dm), human (Hs), *Danio rerio* (Dr) and *Saccharomyces cerevisiae* (Sc) Xrn1 proteins were aligned using T-Coffee (Notredame et al., 2000). Invariant and significantly conserved residues are highlighted in dark and light blue, respectively. Red asterisks indicate acidic residues involved in Mg^{2+} binding; red circles indicate aromatic residues stacking with the bases of the nucleic acid substrate; black inverted triangles indicate residues involved in 5' phosphate recognition. Secondary structure of DmXrn1 is shown above the sequence. Coloring of the secondary structure elements follows the color scheme from Figure 1A. (B) Ribbon diagram representations of DmXrn1 (left) and *S. pombe* Rat1-Rai1 complex (Xiang et al., 2009) (center). Structures were superposed using DALI (Holm and Sander, 1993) and are shown in the same orientation. Right, structural superposition of the catalytic domains of DmXrn1 (catalytic domain in blue, remainder in grey) and *S. pombe* Rat1 (magenta). (C) Structure-based sequence alignment of the catalytic domains of DmXrn1 and *S. pombe* Rat1. Alignment was generated using the SSM server within PDBeFold (Krissinel and Henrick, 2004). Red asterisks indicate acidic residues involved in Mg^{2+} binding; red circles indicate aromatic residues stacking with the bases of the nucleic acid substrate; black inverted triangles indicate residues involved in 5' phosphate recognition. Secondary structure of DmXrn1 is shown above the sequence. Conserved residues are highlighted in blue.

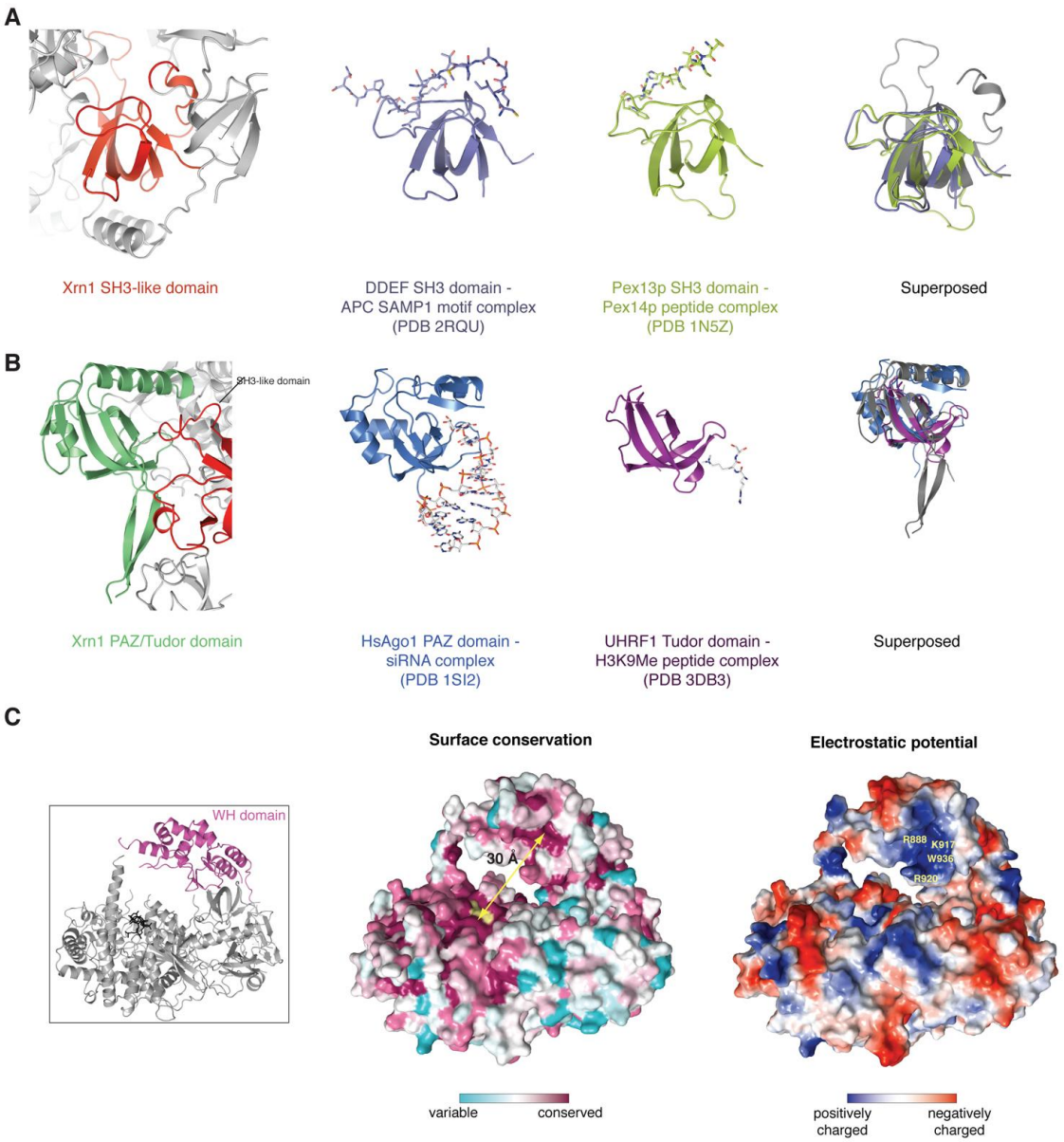


Figure S2, related to Figure 1. Structural Features of the Noncatalytic Domains of DmXrn1
(A) Ribbon diagram representations of the DmXrn1 SH3-like domain (far left, colored red), the human DDEF1 SH3 domain in complex with the SAMP1 motif APC (Kaieda et al., 2010) (center left, colored blue), and the SH3 domain of the peroxisome import factor Pex13p in complex with a Pex14p peptide (Douangamath et al., 2002) (center right, colored green). The SH3-like domain superposes with the two structures with RMSD of 1.7 Å (over 51 Cα atoms) and 2.2 Å (over 53 Cα atoms), respectively. In the far right panel, the three protein domains are shown superposed, with the DmXrn1 SH3-like domain coloured in grey. (B) Ribbon diagram representations of the DmXrn1 PAZ/Tudor domain (far left, PAZ/Tudor coloured green, SH3-like domain loops coloured red), the human Argonaute 1 (HsAgo1) PAZ domain in complex with an siRNA oligonucleotide (Ma et al., 2004) (center left, colored blue), and the Tudor domain of UHRF1 in complex with a trimethylated lysine peptide (center right, colored purple). Far right; the three protein domains superposed (DmXrn1 PAZ/Tudor domain colored grey). The structures were superposed using DALI (Holm and

Sander, 1993) and are shown in identical orientations. The structures superpose with root-mean-square deviations (RMSD) of 3.2 Å (over 85 Cα atoms) and 3.4 Å (over 62 Cα atoms) for 1SI2 and 3DB3, respectively. (C) Surface features of the winged-helix domain of DmXrn1. Left; Ribbon diagram representation of DmXrn1 with the winged-helix (WH) domain highlighted in pink. Center; Surface representation of DmXrn1ΔC (same orientation as in inset) coloured according to phylogenetic conservation. Surface conservation was calculated with CONSURF (Ashkenazy et al., 2010) using a T-Coffee sequence alignment comprising Xrn1 sequences from *Drosophila melanogaster,* human, *Danio rerio*, *Gallus gallus*, *Apis mellifera*, *Schizosaccharomyces pombe and Saccharomyces cerevisiae*. Increasing conservation is indicated with a gradient from turquoise through white to magenta. The substrate is colored yellow. The distance between the 3' end of the substrate in the active site and the conserved patch (K917) in the WH domain is indicated. Right; Surface representation of DmXrn1ΔC colored according to electrostatic potential. Electrostatic potential was calculated using PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC). Strictly conserved residues are labelled.

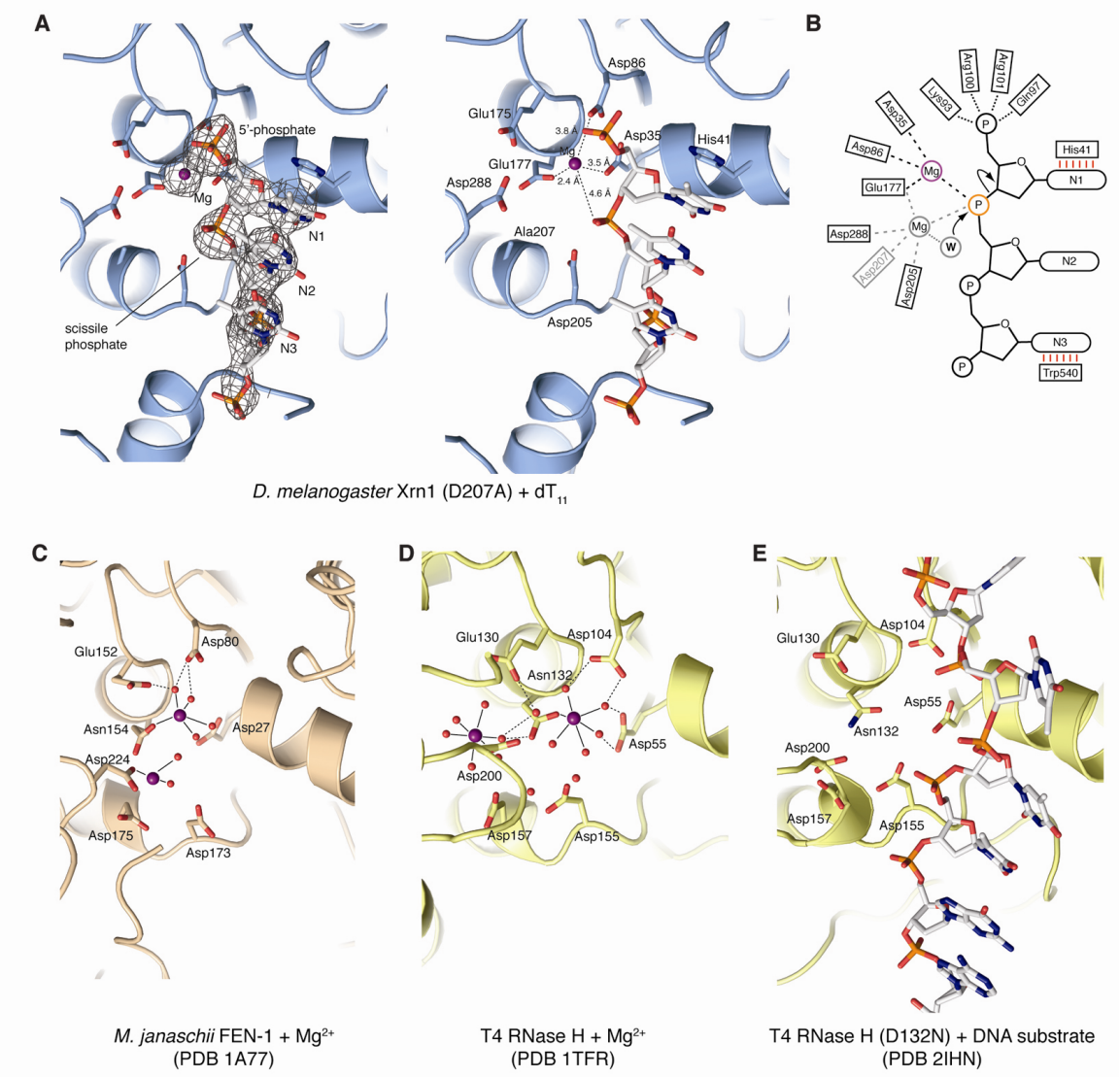


Figure S3, related to Figure 2. Comparisons of the Active Sites of DmXrn1 and Its Structural Homologs

(A) The active site of DmXrn1. Left; the final $2F_o$-F_c map, contoured at 1.0 σ, is shown superposed over the final model. The Mg^{2+} ion is shown as a purple sphere. The tower domain helix (residues 90-119) is omitted for clarity. Right; interatomic distances between active site residues, the substrate and the Mg^{2+} ion are indicated with dashed lines. (B) Schematic representation of DmXrn1 substrate binding and catalysis. Dotted lines indicate electrostatic contacts, red bars indicate stacking interactions. Contacts that are too distant to be mediated by direct inner-sphere coordination are indicated with dashed lines. E177 is coordinated directly. The second Mg^{2+} ion, absent from the structure due to the inactivating D207A mutation, is depicted in grey. This ion is thought to activate a water molecule (W) for an in-line attack at the scissile phosphate (orange circle). (C) Mg^{2+} ion binding in the 2.0 Å structure of *Methanococcus janaschii* FEN-1 (Hwang et al., 1998). Water molecules are shown as red spheres. Inner-sphere coordination interactions are indicated with solid lines. Hydrogen bonding is depicted with dashed lines. (D) Mg^{2+} ion binding in the 2.06 Å structure of T4 RNase H (Mueser et al., 1996). Hydrogen bonding and coordination interactions are depicted

as in B. (E) DNA binding in T4 RNase H substrate complex (Devos et al., 2007). Structures were superposed using DALI (Holm and Sander, 1993) and are shown in identical orientations.

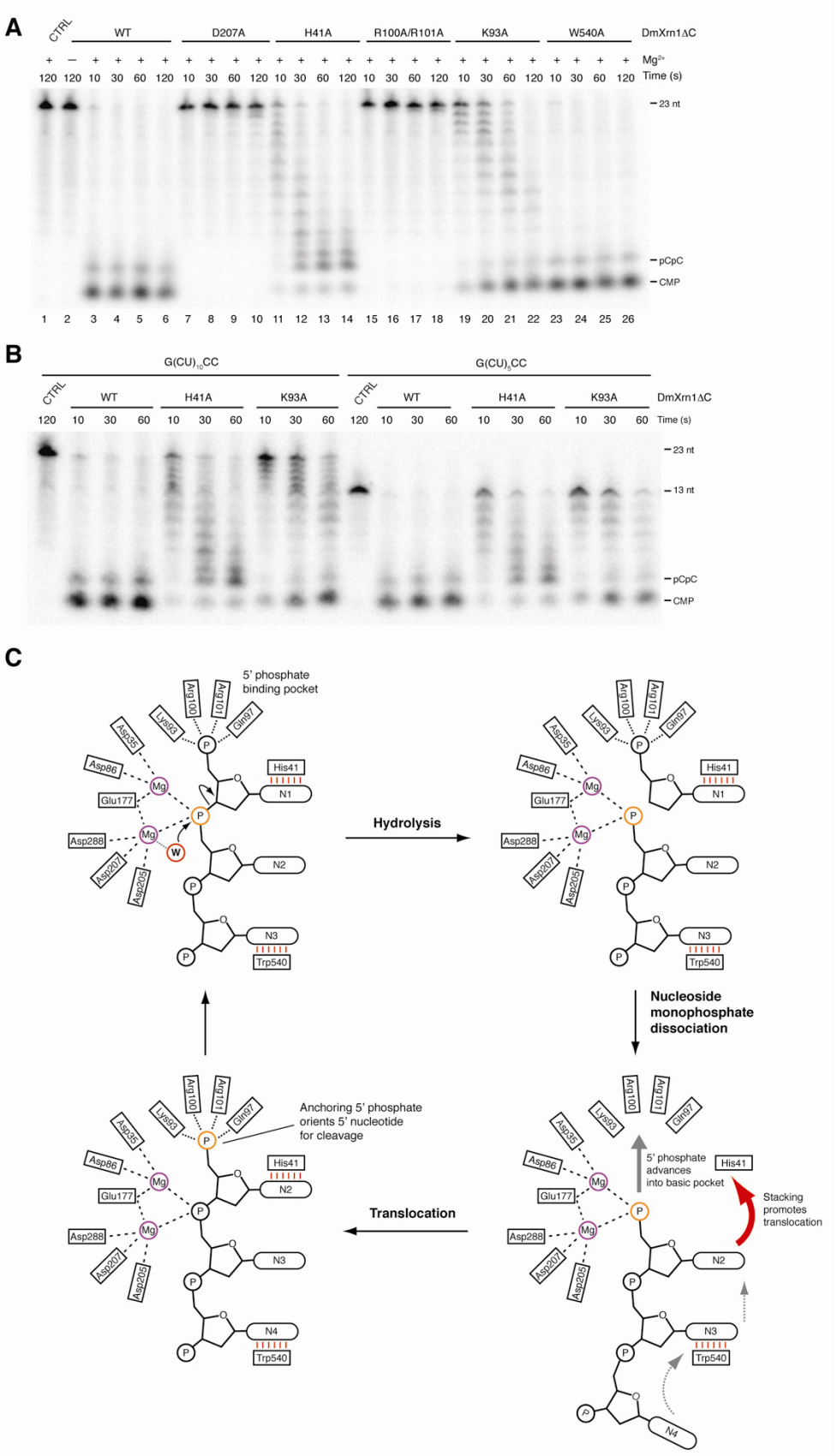

A
CTRL
WT
D207A
H41A
R100A/R101A
K93A
W540A
DmXrn1ΔC
Mg2+
Time (s)
23 nt
pCpC
CMP
B
G(CU)10CC
G(CU)5CC
13 nt
C
5' phosphate binding pocket
Hydrolysis
Nucleoside monophosphate dissociation
Translocation
Anchoring 5' phosphate orients 5' nucleotide for cleavage
5' phosphate advances into basic pocket
Stacking promotes translocation
Arg100
Arg101
Gln97
Lys93
Asp35
Asp86
Glu177
Asp288
Asp207
Asp205
His41
Trp540
N1
N2
N3
N4

Figure S4.

Figure S4, related to Figure 3. Residues Involved in Recognizing the 5'-terminal Nucleotide Contribute to Xrn1 Processivity

(A) Exonuclease activity assay of wild-type (WT) and mutant DmXrn1ΔC proteins. DmXrn1ΔC constructs were incubated with a 3'-[^{32}P]-pCp labeled, 5'-phosphorylated G(CU)$_{10}$C oligonucleotide. The enzyme and substrate concentrations were 370 nM and 200 nM, respectively. Aliquots of reactions were removed at 10, 30, 60 and 120 s time points, as indicated. Products were analyzed by electrophoresis on a 16% polyacrylamide and 8M urea gel and visualized by phosphorimaging. (B) The processivity defects in H41A and K93A mutants are independent of substrate length. Assays were performed in parallel with with 3' end-labeled, 5'-phosphorylated RNA oligonucleotides G(CU)$_{10}$C and G(CU)$_5$C. The assays were analyzed as in A. (C) A putative model for processive RNA degradation by Xrn1. Xrn1 achieves processive degradation by a Brownian ratchet mechanism in which His41, in concert with the 5' phosphate binding pocket, impart directionality to the process. The RNA substrate is bound in the active site with the scissile phosphate (marked by an orange circle) coordinated to the two magnesium ions, one of which activates a water molecule (red circle) for an in-line attack at the scissile phopshate. Following hydrolysis and nucleoside monophosphate dissociation, π-π stacking between the base of the 5'-terminal nucleotide of the product (N2) and His41 drives substrate translocation (indicated with a red arrow). As the substrate translocates, the base stacking between the base of nucleotide N3 and Trp540 is disrupted and replaced by a stacking interaction with the base of the downstream nucleotide (N4). Substrate translocation is reinforced by electrostatic contacts between the 5' phosphate of the substrate and the basic phosphate binding pocket. This ensures correct orientation of the 5'-terminal nucleotide for the next round of cleavage.