Supplemental Materials for

Structure and function of the 5’→3’ exoribonuclease Rat1

and its activating partner Rai1

Song Xiang,1 Amalene Cooper-Morgan,1 Xinfu Jiao,2
Megerditch Kiledjian,2 James L. Manley,1 Liang Tong1

1Department of Biological Sciences
Columbia University
New York, NY10027, USA

2Department of Cell Biology and Neuroscience
Rutgers University
Piscataway, NJ08854, USA

Correspondence information for Liang Tong

Phone: (212) 854-5203, FAX: (212) 865-8246
E-mail: ltong@columbia.edu
Supplemental Results and Discussion

Structure of Rat1

The first conserved region of Rat1 contains a seven-stranded ($\beta_1$ through $\beta_7$), mostly parallel $\beta$-sheet that is flanked by $\alpha$-helices on both sides (Supplemental Fig. 2). Strands $\beta_2$ through $\beta_6$ are arranged similar to the first five strands in a Rossmann fold. Strand $\beta_7$ occupies the position of the sixth strand in a Rossmann fold, but runs in the opposite direction. An extra helix is inserted after strands $\beta_2$ ($\alpha_B$) and $\beta_3$ ($\alpha_D$). Residues in the second conserved region form several helices and long loops, which are wrapped around the $\alpha_D$ helix.

The N-terminal residues of the $\alpha_D$ helix are strongly conserved in XRN2/Rat1 as well as XRN1 homologs (Supplemental Fig. 1), and they have important roles in the active site of these enzymes. The sequence conservation of the C-terminal residues of the $\alpha_D$ helix is weaker among the XRN2/Rat1 homologs, and the helix may be much shorter in the XRN1 homologs because of a deletion here (Supplemental Fig. 1). This suggests that the tower domain may be unique to the XRN2 enzymes, and may serve XRN2-specific functions, for example in the termination of Pol II transcription.

The functional importance of the $\alpha_D$ helix is supported by observations on temperature-sensitive mutations in S. cerevisiae Xrn1 and Rat1. A P90L mutation in Xrn1, equivalent to Pro108 in S. pombe Rat1 and located just prior to this helix, disrupts exonuclease activity at the non-permissive temperature $^1$. The rat1-1 temperature-sensitive mutation of Rat1 is Y657C $^2$, equivalent to Tyr643 in S. pombe Rat1, in the second conserved region. This residue is located in the $\alpha_M$-$\alpha_N$ loop (Supplemental Fig. 1), and its side-chain hydroxyl group is hydrogen-bonded to the side chain of Asp710, in a $\beta$-hairpin-like structure (between helices $\alpha_O$ and $\alpha_P$) that wraps around the base of the $\alpha_D$ helix (Supplemental Fig. 2). Therefore, both mutations appear to lead to destabilization of the $\alpha_D$ helix at the non-permissive temperature.

The segment immediately following the second conserved region adds four anti-parallel strands ($\beta_8$-$\beta_{11}$) to the central $\beta$-sheet in the first region, producing a highly-twisted eleven-stranded $\beta$-sheet (Supplemental Fig. 2). Residues at the extreme C-terminus of the current expression construct form a helix and interact with helices $\alpha_A$ (at the N-terminus) and $\alpha_H$ in the first conserved region (Supplemental Fig. 2).

Most of the segment linking the two conserved regions is not observed in our structure (Fig. 1b and Supplemental Fig. 1), probably due to disorder. The C-terminal end of the first region and the N-terminal end of the second region are located within 10 Å of each other, suggesting that this poorly conserved linker segment is likely an inserted cassette on the surface of the Rat1 structure, away from its active site. This linker contains one of the protease sensitive sites in Xrn1 $^1$. A second cluster of protease sensitive sites is equivalent to residue 810 in S. pombe Rat1, in an extended surface loop connecting strands $\beta_{10}$ and $\beta_{11}$ in the poorly conserved C-terminal segment (Supplemental Fig. 2).

Comparison to structural homologs of Rat1
Rat1 has several weak structural homologs in the Protein Data Bank, as identified by the SSM server. The structural homologs include the FEN-1 family of endonucleases (Supplemental Fig. 3), the 5' exonuclease from the phage T5, RNase H from the phage T4 (Supplemental Fig. 3), the 5' nuclease domain of Taq DNA polymerase, and other PIN domain-containing nucleases. The rms distance for equivalent Cα atoms between Rat1 and any of these other structures is about 3 Å, but the sequence identity is very low, between 8 and 15%. In addition, only about 200 Cα atoms in Rat1 can be aligned with residues in these other structures, as the structural similarity is observed only for residues in the first conserved region.

Strands β2-β7 in the central β-sheet and a few of the flanking helices have equivalents in most of these other structures (Supplemental Fig. 3). The tower domain in Rat1 is equivalent to the helical clamp in A. fulgidus FEN-1 (although helix αD in Rat1 is much longer) (Supplemental Fig. 3) and the helical arch in T5 exonuclease, but the equivalent region is a long loop in M. jannaschii FEN-1 and is disordered in T4 RNase H, even when in complex with a fork DNA substrate (Supplemental Fig. 3).

The active site of Rat1

The base of the αD helix in the tower domain contributes three positively-charged (Lys111, Arg118, Arg119) and two polar (Gln114, Gln115) side chains to the active site, and a helix (αB) after strand β2 contributes two His residues (His61 and His65) and Asn57. These residues are clustered together and form a steep wall on one side of the active site pocket (Supplemental Fig. 4), and might be important for interacting with the phosphate backbone of the substrate. Mutation of these residues, as well as of those in the acidic cluster, can abolish exonuclease activity or produce defect in Pol II termination, confirming their functional importance. In addition, structural comparisons with the RNA complexes of FEN-1 and RNase H suggest that other residues in both conserved regions are likely involved in binding the RNA substrate of XRN1 and XRN2.

There are few direct contributions from amino acid side chains in the second conserved region of Rat1 to the active site. The side-chain hydroxyl groups of Tyr627 and Tyr628 (in helix αM) have hydrogen-bonding interactions with the acidic residues Glu205 and Asp237 in the active site, respectively, and the side chain of Gln671 (helix αN) is located in the cluster of side chains from the αB and αD helices (Supplemental Fig. 4). These three residues are strictly conserved among the XRN1 and XRN2 homologs (Supplemental Fig. 1). The poorly conserved C-terminal segment of Rat1 is located on the opposite face of Rat1 from the active site and does not make any direct contribution to it (Supplemental Fig. 4).

In related nucleases, two divalent cations are associated with the cluster of acidic residues in the active site and the metal ions are important for catalysis. We soaked the Rat1-Rai1 crystals in 50 mM MnCl₂ but did not observe any binding in this region based on the crystallographic analysis. A more detailed structural comparison with RNase H, in complex with two metal ions, showed that there are large differences in the positions of some of the equivalent acidic residues in the cluster, especially for residues Asp235 and Asp336 in Rat1 (Supplemental Fig. 5). It is possible that a conformational change in the active site region of Rat1 is needed to bring the acidic residues into better alignment for the binding of divalent cations. In addition, this binding may also be
dependent on the presence of the RNA substrate, as has been shown for related enzymes. We soaked the Rat1-Rai1 crystals with various mononucleotides and oligonucleotides but did not observe any binding, which may be consistent with the need for a conformational change.

The structure suggests two possible reasons why RNAs with a 5’ cap are poor substrates of XRNs — (1) The cap structure is too bulky to be accommodated in the active site. (2) The 5’ phosphate group of the RNA substrate is recognized specifically by the enzyme. While the latter possibility is more likely, further studies, especially the structure of an RNA complex, are needed to definitively answer this question.

**Structure of the Rat1-Rai1 complex**

There are ion-pair, hydrogen-bonding as well as van der Waals interactions in the interface between Rat1 and Rai1 (Supplemental Fig. 6). The side chain ammonium ion of Lys848 (in the C-terminal loop) in Rat1 is hydrogen-bonded to the main-chain carbonyl oxygen atoms of residues 159, 160 and 162 (in the β8-αE loop) in Rai1, and the main-chain carbonyl atom of Leu851 (C-terminal loop) in Rat1 is hydrogen-bonded to the side-chain hydroxyl of Tyr46 (β4) in Rai1 (Supplemental Fig. 6). An ion pair between Asp372 (αJ-αK loop) of Rat1 and Arg164 (αE) of Rai1 should provide further stabilization of this complex. The side chains of Val219, Pro221 (αF-β5 loop), Phe845, Met850, Leu851, and Leu852 (C-terminal loop) in Rat1, and the side chains of Tyr45, Tyr46 (β4), Ile158, Trp159 (β8-αE loop) and Trp204 (β11-β12 loop) in Rai1 contribute to the van der Waals interactions across the interface.

It was reported earlier that deletion of the C-terminal 115 residues of Rai1 abolishes the interaction with Rat1. The structure shows that none of these C-terminal residues are directly in the interaction interface with Rat1. Instead, it is clear from the structure that removal of these residues will greatly destabilize the Rai1 protein, and the loss of interaction with Rat1 is most likely due to the loss of the native structure of the deletion mutant.

To ascertain whether Rai1 undergoes a conformational change upon binding to Rat1, we determined the structures at 2.0 Å resolution of *S. pombe* Rai1 and mouse DOM3Z alone (Supplemental Table 1). The structure of Rai1 alone, with rms distance of 0.5 Å for equivalent Cα atoms to those in the complex with Rat1, shows that the β8-αE loop is stabilized by crystal packing and assumes a different conformation, and the αE helix is disordered (Supplemental Fig. 8), suggesting that this region of Rai1 may assume a stable conformation only in the Rat1-Rai1 complex. In the DOM3Z structure, with rms distance of 1.1 Å to that of Rai1, the αE helix is absent (Supplemental Fig. 8), due to the 7-residue deletion in this region (Supplemental Fig. 7), and the loop following β7 assumes a different conformation. The structure therefore confirms that DOM3Z is unlikely to interact with XRN2 or Rat1.

**Biochemical and functional characterization of the Rat1-Rai1 interface**

The structure of the Rat1-Rai1 complex shows that residues 845-856 in the C-terminal segment of Rat1 are located in the center of the interface (Supplemental Fig. 1). This is supported by our observations on the C-terminal deletion mutants of Rat1. Rat1 mutants containing the N-terminal 780, 815 or 850 residues were unable to interact with
Rai1, while those containing 885, 920, or 950 residues and the full-length Rat1 did associate with Rai1 (data not shown). These data demonstrate that the integrity of the Rat1 C-terminal loop is crucial for the formation of this complex.

Rat1 alone produced major intermediates with all three RNA substrates, but the stalling occurred at different positions in the RNA (Fig. 2c). This suggests that the stalling is not likely some inherent property of the Rat1 enzyme itself, but rather due to secondary structure features in the RNA. This is consistent with earlier published data showing that the presence of an oligo(G) tract or stem loop can create stalls in Rat1 activity \textsuperscript{19, 20}, and secondary structure prediction, with the program Mfold \textsuperscript{21}, showed the presence of many stem loop structures in the three RNA substrates. It could also be possible that Rat1 alone is not completely processive, at least under our assay conditions. Importantly, this stalling is alleviated partly or entirely by Rai1 with all three substrates, highlighting the importance of Rai1 to Rat1 activity. Earlier studies demonstrating processive activity for Rat1 actually used the Rat1-Rai1 complex as the enzyme \textsuperscript{19, 20}. In any event, our data explain why Rai1 is required along with Rat1 for efficient Pol II termination in yeast \textsuperscript{16}. 
## Table 1
Summary of crystallographic information

| Structure | Rat1-Rai1 complex, free enzyme | Rat1-Rai1 complex, soaked with 50 mM Mn$^{2+}$ | Rai1, free enzyme | Mouse DOM3Z, free enzyme | DOM3Z in complex with GDP |
|-----------|-------------------------------|-----------------------------------------------|-------------------|--------------------------|---------------------------|
| Resolution range (Å) | 30–2.2 | 30–2.5 | 30-2.0 | 30-2.0 | 30-2.6 |
| Space Group | $P_2_1_2_1$ | $P_2_1_2_1$ | $C_2$ | $P_2_1$ | $P_2_1_2_1$ |
| Unit cell parameters ($a$, $b$, $c$, $\beta$) | 96.7, 190.9, 84.6 | 96.6, 190.8, 84.3 | 101.6, 60.8, 73.1, 103.1 | 46.0, 88.4, 50.1, 114.2 | 58.3, 73.6, 108.9 |
| Number of observations | 310,922 | 379,250 | 91,235 | 72,257 | 48,083 |
| Number of reflections$^1$ | 77,326 | 102,603 | 29,187 | 23,258 | 14,119 |
| Completeness (%) | 92 (60) | 99 (98) | 98(91) | 98(100) | 99(98) |
| $R_{merge}^2$ (%) | 7.3 (35.6) | 8.9 (42.7) | 7.9(37.6) | 7.7(19.0) | 13.1(48.9) |
| $I/\sigma I$ | 17.5 (1.9) | 15.4 (2.9) | 16.8(2.6) | 17.1(6.0) | 11.1(2.7) |
| $R$ factor$^3$ (%) | 21.8 (28.7) | 21.5 (26.6) | 19.5(24.3) | 22.6(22.8) | 20.4(31.1) |
| Free $R$ factor$^3$ (%) | 25.9 (33.0) | 25.6 (30.1) | 23.0(31.6) | 28.3(30.5) | 26.7(37.5) |
| rms deviation in bond lengths (Å) | 0.008 | 0.008 | 0.009 | 0.008 | 0.007 |
| rms deviation in bond angles (°) | 1.0 | 1.0 | 1.2 | 1.2 | 1.1 |

1. Friedel pairs were treated as independent reflections for the Mn$^{2+}$ soak (SeMet crystal).

2. The numbers in parentheses are for the highest resolution shell.

3. 5% of the reflections were selected for free $R$ calculation.
References:
1. Page, A. M., Davis, K., Molino, C., Kolodner, R. D. & Johnson, A. W. Mutational analysis of exoribonuclease I from Saccharomyces cerevisiae. Nucl. Acid Res. 26, 3707-3716 (1998).
2. Luo, W., Johnson, A. W. & Bentley, D. L. The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. Genes Develop. 20, 954-965 (2006).
3. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Cryst. D60, 2256-2268 (2004).
4. Hwang, K. Y., Baek, K., Kim, H.-Y. & Cho, Y. The crystal structure of flap endonuclease-1 from Methanococcus jannaschii. Nat. Struct. Biol. 5, 707-713 (1998).
5. Sayers, J. R. & Artymiuk, P. J. Flexible loops and helical arches. Nat. Struct. Biol. 5, 668-670 (1998).
6. Chapados, B. R. et al. Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. Cell 116, 39-50 (2004).
7. Sakurai, S. et al. Structural basis for recruitment of human flap endonuclease 1 to PCNA. EMBO J. 24, 683-693 (2005).
8. Ceska, T. A., Sayers, J. R., Stier, G. & Suck, D. A helical arch allowing single-stranded DNA to thread through T5 5'-exonuclease. Nature 382, 90-93 (1996).
9. Mueser, T. C., Nossal, N. G. & Hyde, C. C. 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 (1996).
10. Devos, J. M., Tomanicek, S. J., Jones, C. E., Nossal, N. G. & Mueser, T. C. 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 (2007).
11. Kim, Y. et al. Crystal structure of Thermus aquaticus DNA polymerase. Nature 376, 612-616 (1995).
12. Murali, R., Sharkey, D. J., Daiss, J. L. & Murthy, H. M. K. Crystal structure of Taq DNA polymerase in complex with an inhibitory Fab: The Fab is directed against an intermediate in the helix-coil dynamics of the enzyme. Proc. Natl. Acad. Sci. USA 95, 12562-12567 (1998).
13. Glavan, F., Behm-Ansmant, I., Izaurrelde, E. & Conti, E. Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. EMBO J. 25, 5117-5125 (2006).
14. Clissold, P. M. & Ponting, C. P. PIN domains in nonsense-mediated mRNA decay and RNAi. Curr. Biol. 10, R888-R890 (2000).
15. Solinger, J. A., Pascolini, D. & Heyer, W.-D. Active-site mutations in the Xrn1p exoribonuclease of Saccharomyces cerevisiae reveal a specific role in meiosis. Mol. Cell. Biol. 19, 5930-5942 (1999).
16. Kim, M. et al. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature 432, 517-522 (2004).
17. Yang, W., Lee, J. Y. & Nowotny, M. Making and breaking nucleic acids: Two-Mg2+-ion catalysis and substrate specificity. Mol. Cell 22, 5-13 (2006).
18. Xue, Y. et al. Saccharomyces cerevisiae RAI1 (YGL246c) is homologous to human DOM3Z and encodes a protein that binds the nuclear exoribonuclease Rat1p. Mol. Cell. Biol. 20, 4006-4015 (2000).

19. Poole, T. L. & Stevens, A. Structural modifications of RNA influence the 5' exoribonucleolytic hydrolysis by XRN1 and HKE1 of Saccharomyces cerevisiae. Biochem. Biophys. Res. Commun. 235, 799-805 (1997).

20. Stevens, A. & Poole, T. L. 5'-exonuclease-2 of Saccharomyces cerevisiae. Purification and features of ribonuclease activity with comparison to 5'-exonuclease-1. J. Biol. Chem. 270, 16063-16069 (1995).

21. Zucker, M. Mfold web server for nucleic acid folding and hybridization prediction. Nucl. Acid Res. 31, 3406-3415 (2003).

22. DeLano, W. L. (DeLano Scientific, San Carlos, CA, 2002).
Fig. 1. Sequence alignment of *S. pombe* Rat1, human XRN2, and human XRN1. Residues that are conserved among the three sequences are colored in magenta. Residues in the linker between the two conserved regions are indicated with the gray bar, and those not observed in our structure of Rat1 are shown in italic. The colored dots highlight those residues in the active site (red) or in the interface with Rai1 (blue).
Fig. 2. (Top). Schematic drawing of the structure of *S. pombe* Rat1. Residues are colored according to Fig. 1a. The side chain of Tyr643, equivalent to the site of the *rat1-1* temperature-sensitive mutation in *S. cerevisiae* *RAT1* ², and its hydrogen-bonding partner Asp710 are shown as stick models with carbon atoms in light blue. (Bottom). Structure of *S. pombe* Rat1, viewed after roughly an 80° rotation around the horizontal axis, to show the tower domain. All the structure figures were produced with Pymol ²².
Fig. 3. (A). Schematic drawing of the structure of *A. fulgidus* FEN-1 nuclease in complex with DNA. The phosphate backbone of the DNA is shown in black. Secondary structure elements that are equivalent to those in Rat1 are labeled. (B). Schematic drawing of the structure of T4 RNase H in complex with a fork DNA. The metal ions observed in the structure of the free enzyme are shown as spheres in gray for reference. Overall molecular surface of FEN-1 (C), T4 RNase H (D), and Rat1 (E). The active site is indicated with the red star.
**Fig. 4.** Schematic drawing of the active site of *S. pombe* Rat1. Residues in the first conserved region are colored in cyan for the main chain and light blue for the side chain, and the second region in magenta and pink. The C-terminal segment (in yellow) does not make any direct contribution to the active site.
**Fig. 5.** Overlay of the conformation of the cluster of acidic residues in the active site of Rat1 (in cyan) and T4 RNase H (in gray). The bound positions of the two metal ions in RNase H are shown as gray spheres. Residue numbers in parenthesis are for RNase H.
Fig. 6. Detailed interactions between Rat1 and Rai1. The side chains of Rat1 are shown in light blue (in the first conserved region, with main chain in cyan) and orange (in the C-terminal segment, with main chain in yellow), and those of Rai1 are shown in black (with main chain in green).
Fig. 7. Sequence alignment of *S. pombe* Rai1, *S. cerevisiae* Rai1, and their mouse homolog DOM3Z. Residues that are conserved among the three sequences are colored in magenta. The colored dots highlight those residues in the large, active site pocket (red) or in the interface with Rat1 (blue).
Fig. 8. Overlays of the structure of Rai1 in the complex with Rat1 (in color) and the structure of Rai1 alone (Top) or mouse DOM3Z (Bottom) (in gray). The arrow points to the region of interaction with Rat1, which has large differences among the three structures.
Fig. 9. Overlays of the metal ion binding site in the structure of Rai1 in complex with Rat1 with that in Rai1 alone (Top) or mouse DOM3Z (Bottom, residues numbers in parenthesis are for DOM3Z). The structure in the Rat1-Rai1 complex is shown in black, and the other structures in gray. The interaction between Glu192 and the cation is mediated by a water molecule in DOM3Z. Other than this, the metal ion binding modes are highly similar.
Fig. 10. Cleavage of 5'-end phosphorylated, 3'-end labeled RNA substrate by Rat1 and the Rat1-Rai1 complex. The E199A/D201A mutant can still activate the ribonuclease activity of Rat1 towards RNA substrate with 5’ monophosphate.
Fig. 11. Rai1 pyrophosphatase activity is dependent on RNA body. $^{32}$P labeled pppGp or GTPs were used as substrates incubated without or with 50 nM recombinant Rai1 and Rat1 at 37 °C for 45 minutes. Reaction products were resolved on PEI TLC developed in 1.5 M KH$_2$PO$_4$ (pH 3.4). The labeled phosphate is denoted by *. The marker migrations were indicated on the right.
Fig. 12. Molecular insight into substrate binding to DOM3Z. (a). Schematic drawing of the structure of mouse DOM3Z in complex with GDP (in light gray). The divalent cation is shown as a gray sphere. (b). Final omit $F_o-F_c$ electron density at 2.6 Å resolution for the GDP molecule, contoured at 2σ. (c). Overlay of the bound position of GDP (in light gray) and Hepes (in black) in the active site of DOM3Z.