THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 6, pp. 3617–3624, February 5, 2010
© 2010 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in the U.S.A.

The Structure of the Mammalian RNase H2 Complex Provides Insight into RNA·DNA Hybrid Processing to Prevent Immune Dysfunction*[S]

Received for publication, August 24, 2009, and in revised form, September 21, 2009 Published, JBC Papers in Press, November 18, 2009, DOI 10.1074/jbc.M109.059048

Nadine M. Shaban, Scott Harvey, Fred W. Perrino, and Thomas Hollis[1]

From the Department of Biochemistry, Center for Structural Biology, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157

The mammalian RNase H2 ribonuclease complex has a critical function in nucleic acid metabolism to prevent immune activation with likely roles in processing of RNA primers in Okazaki fragments during DNA replication, in removing ribonucleotides misinserted by DNA polymerases, and in eliminating RNA·DNA hybrids during cell death. Mammalian RNase H2 is a heterotrimeric complex of the RNase H2A, RNase H2B, and RNase H2C proteins that are all required for proper function and activity. Mutations in the human RNase H2 genes cause Aicardi-Goutières syndrome. We have determined the crystal structure of the three-protein mouse RNase H2 enzyme complex to better understand the molecular basis of RNase H2 dysfunction in human autoimmunity. The structure reveals the intimately interwoven architecture of RNase H2B and RNase H2C that interface with RNase H2A in a complex ideally suited for nucleic acid binding and hydrolysis coupled to protein-protein interaction motifs that could allow for efficient participation in multiple cellular functions. We have identified four conserved acidic residues in the active site that are necessary for activity and suggest a two-metal ion mechanism of catalysis for RNase H2. An Okazaki fragment has been modeled into the RNase H2 nucleic acid binding site providing insight into the recognition of RNA·DNA junctions by the RNase H2. Further structural and biochemical analyses show that some RNase H2 disease-causing mutations likely result in aberrant protein-protein interactions while the RNase H2A subunit-G37S mutation appears to distort the active site accounting for the demonstrated substrate specificity modification.

RNA·DNA hybrids play essential roles in biology and are ubiquitously present in all cells. The RNA primers in Okazaki fragments are required for lagging strand synthesis during genomic DNA replication, and R-loops are generated in mitochondrial DNA replication (1–5). RNA·DNA intermediates are also formed during transcription and are part of class I retrotransposon element amplification, as well as a necessary intermediate in retrovirus replication (6–8). Ribonucleotides within RNA·DNA hybrids are recognized and hydrolyzed by the RNase H enzymes, which are divided into two major families, Type 1 and Type 2, based on amino acid sequence similarities and distinct biochemical properties (9–11). Enzymes in the RNase H Type 1 family (RNase H1)[2] recognize a minimum of four ribonucleotides within an RNA·DNA duplex with hydrolysis occurring near the 5′-end of the RNA sequence. In contrast, RNase H2 enzymes are able to recognize even a single ribonucleotide within a DNA duplex as a substrate and cleave the RNA-containing strand one nucleotide away from the RNA·DNA junction (for a review see Ref. 12).

The mammalian RNase H2 enzyme complex is composed of three separate proteins encoded by the *RNaseH2A*, *RNaseH2B*, and *RNaseH2C* genes, all of which are required for activity. The RNaseH2A protein has endoribonuclease activity, and the mammalian *RNaseH2A* gene has been recognized for several years by homology with the bacterial enzyme (13, 14). However, only recently were the genes for the RNase H2B and RNase H2C proteins identified to establish the components of the RNase H2 complex (15). In contrast to the mammalian RNase H2, the prokaryotic RNase H2 enzyme functions as a single protein and shares ~30% amino acid sequence identity with the human RNase H2A protein. Previous structural studies of the single subunit prokaryote RNase H in complex with RNA·DNA have identified residues involved in two-metal ion coordination and catalysis, as well as positively charged and polar residues that are presumably involved in substrate binding and recognition (16–20). However, it has been unclear if the multisubunit mammalian RNase H2 would function with a similar mechanism and what role the multiple subunits might have in substrate recognition and catalytic activity.

RNase H2 is the major RNase H activity in mammalian cells (14) with possible roles in various aspects of nucleic acid metabolism. RNase H2 cooperates with the endonuclease, FEN-1 (human) or RAD27 (yeast) in cell extracts to excise ribonucleotide primers from an RNA·DNA duplex, indicating a role in Okazaki fragment processing during DNA replication (21–23).

* This work was supported, in whole or in part, by National Institutes of Health Grant R01-GM069962 (to F. W. P.). This work was also supported by the Alliance for Lupus Research (Grant 67692 to F. W. P.) and the American Cancer Society (Grant RSG-04-187-01-GMC to T. H.).

The atomic coordinates and structure factors (code 3KIO) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1 and S2.

[1] To whom correspondence should be addressed: Dept. of Biochemistry, Center for Structural Biology, Wake Forest University School of Medicine, Center for Structural Biology, Medical Center Blvd., Winston-Salem, NC 27157. Tel.: 336-716-0768; Fax: 336-777-3242; E-mail: thollis@wfubmc.edu.

[2] The abbreviations used are: RNase H1 and H2, RNase H Type 1 and 2 families; AGS, Aicardi-Goutières syndrome; TFIIF, transcription factor IIF; PIC, polymerase initiation complex; FAM, carboxyfluorescein.

The RNase H2 forms a complex with proliferating cell nuclear antigen (24) suggesting it is tethered to the DNA replication apparatus, perhaps to remove ribonucleotides that might be misincorporated by DNA polymerases during DNA replication or repair (25–27). It is also likely that the human RNase H2 enzyme functions to degrade RNA·DNA hybrids during cell death to prevent nucleic acid-mediated immune activation (15).

Mutations in the three genes coding for the human RNase H2 subunits and in the gene encoding the TREX1 3′-5′ deoxyribonuclease cause the neurological disorder, Aicardi-Goutières syndrome (AGS) (15, 28, 29). The underlying pathogenesis of AGS is thought to involve the accumulation of non-processed DNA and RNA intermediates leading to activation of the innate immune response (30–32). *TREX1*$^{-/-}$ mice have elevated levels of interferon-α and develop inflammatory myocarditis (33). Furthermore, cells lacking TREX1 activity accumulate short single-stranded DNA (34) and DNA derived from endogenous retroelements (35). The biological mechanism for disease mediated through mutations in the three human RNase H2 subunit genes remains unclear. We and others have demonstrated that RNase H2 complexes containing the mutations identified in AGS patients retain ribonuclease H activity (24, 36), highlighting the complex nuclease dysfunction mechanisms in AGS and pointing toward multiple mechanisms, including catalytic inactivation and alterations in protein-protein or protein-nucleic acid interactions.

The relevance of RNase H2 to proper nucleic acid maintenance for the prevention of immune activation and disease in humans prompted our studies to determine the x-ray structure of the heterotrimeric RNase H2 complex, particularly in light of the fact that the RNase H2B and H2C proteins have no known counterparts in prokaryotic organisms. To elucidate the interactions of the individual proteins that bear on nucleic acid substrate recognition and cellular activity regulation, we have determined the crystal structure of the mouse RNase H2 enzyme complex. Our results reveal an intricately interwoven complex that is an ideal platform for nucleic acid binding and hydrolysis coupled with the structural scaffold to accommodate multiple protein-protein interactions. The structure provides a framework to model RNA-DNA interactions that provide insight into RNase H2 recognition of the likely cellular substrates such as Okazaki fragments and singly incorporated ribonucleotides.

EXPERIMENTAL PROCEDURES

Enzyme Preparation and Purification—Mouse RNase H2 enzyme complex was expressed and purified as described previously (36). Briefly, the genes encoding a polyhistidine-tagged RNase H2B protein and RNase H2A and RNase H2C proteins were co-expressed in *Escherichia coli* BL21 (DE3) cells. The RNase H2 complex was isolated intact from cell extracts using nickel-nitrilotriacetic acid affinity chromatography followed by ion-exchange chromatography. The purified protein was concentrated to 10 mg/ml and flash-frozen at −80 °C until needed. Mutations in the RNase H2 genes were introduced using the QuikChange Site-Directed Mutagenesis^{TM} method (Stratagene) and verified by nucleotide sequencing. All mutant proteins were expressed and purified as the wild-type enzyme.

Ribonuclease H Assays—The RNase H2 reactions contained 20 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 2 mM dithiothreitol, 50 μg/ml bovine serum albumin, 200 nM 5′-FAM-RNA_{20}-DNA_{10} or 5′-FAM-DNA_{19}-RNA_1-DNA_{10} annealed to the complementary 30-mer DNA, and RNase H2 enzyme as indicated. Samples were incubated at 25 °C and quenched by addition of 3 volumes of cold ethanol and dried *in vacuo*. Pellets were resuspended in 6 μl of formamide, heated to 95 °C for 5 min, and separated on 23% denaturing polyacrylamide gel. Fluorescently labeled bands were visualized and quantified using a Typhoon phosphorimaging device (Amersham Biosciences) (36, 37).

Protein Crystallization—The RNase H2 enzyme was crystallized using the sitting drop, vapor-diffusion technique. Purified protein (10 mg/ml) was mixed with equal volumes of reservoir solution and placed on a bridge above 100 μl of reservoir containing 20% polyethylene glycol 3350, 400 mM sodium chloride, 200 mM potassium thiocyanate, 5 mM tris(2-carboxyethyl)phosphine. Crystallization experiments were carried out at 20 °C, and crystals grew within 7–10 days. Prior to data collection, crystals were soaked in reservoir solution containing 15% 2-methyl-2,4-pentanediol for 30 min in preparation for freezing. Crystals were then mounted on a nylon loop and flash frozen in liquid nitrogen. Crystals of the RNase H2 complex belong to space group $P2_12_12$ with unit cell dimensions a = 285.7 Å, b = 44.6 Å, c = 67.6 Å, and $\alpha = \beta = \gamma = 90°$ (Table 1). One RNase H2 complex occupies the asymmetric unit (V_M = 2.5 Å^3/Da).

Phasing and Refinement—Multiwavelength anomalous diffraction x-ray data were collected at beamline X25 of the National Synchrotron Light Source, Brookhaven National Labs. Intensity data were processed with the program HKL2000 (38). Selenomethionyl-derivatized protein was purified as wild type and crystallized in similar conditions (39). We were able to identify the positions of 8 of 12 possible selenium sites, and after initial phasing with MLPHARE (40) and density modification using the program RESOLVE (41) electron density for most of the RNase H2A subunit could be deciphered. However, only poor density was present for the rest of the complex. To overcome this, three additional methionines were introduced into the complex by site-directed mutagenesis of conserved leucine residues (L32M and L155M in RNase H2C and L97M in RNase H2B). Multiwavelength anomalous diffraction phasing using the crystals of the mutant protein followed by density modification allowed for calculation of electron density that showed the RNase H2B and RNase H2C proteins. The model for each of the proteins was built manually using the program COOT (42) and refined with the program REFMAC5 (43) using the experimental phase information. The validity of the refinement was monitored at each step by change in the free *R*-factor (44) as well as the inspection of stereochemical parameters with the program MOLPROBITY (45). The refinement of the model converged with a final *R*-factor of 27.5% (R_{free} = 31.6%) using all observed x-ray data measurements in the resolution range 30-3.1 Å. A Ramachandran plot shows >90% of all residues in the most favored region. Coordinates for the three subunit

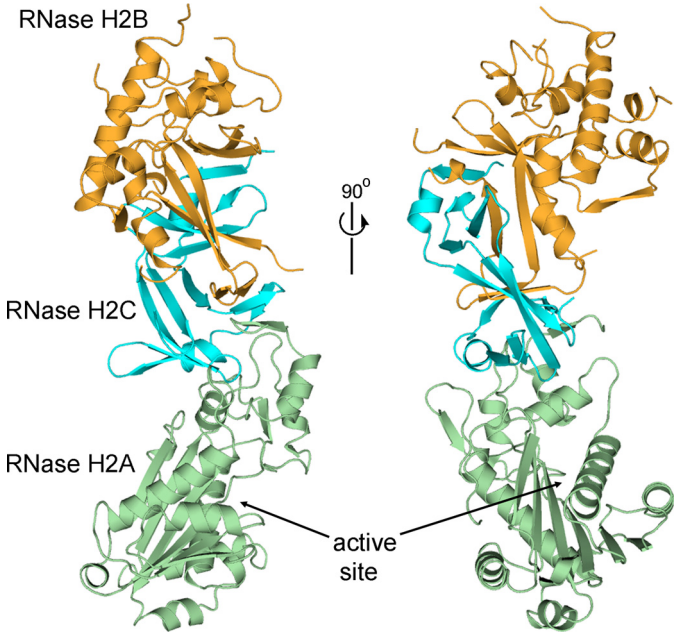

FIGURE 1. **Ribbon structure of the mouse RNase H2 enzyme complex consisting of the RNase H2A (*green*), RNase H2B (*gold*), and RNase H2C (*blue*) proteins.** The elongated complex is stabilized by the weaving together of the H2B and H2C proteins to form a triple-barrel motif that provides a platform for binding the H2A protein. The active site of the catalytic H2A protein is located near one end of the complex with no physical contributions from the H2B or H2C protein.

mouse RNase H2 complex have been deposited in the Protein Data Bank (PDB id: 3KIO).

RESULTS AND DISCUSSION

Structure of the RNase H2 Complex—We have determined the structure of the mouse RNase H2 consisting of the 301-amino acid RNase H2A, 308-amino acid RNase H2B, and 166-amino acid RNase H2C proteins to 3.1-Å resolution (Fig. 1). The structure was determined using multiwavelength anomalous diffraction with selenomethionine-substituted protein (Table 1). A single heterotrimeric complex consisting of the H2A, H2B, and H2C proteins occupies the asymmetric unit.

The overall structure reveals an elongated arrangement of the subunits with the H2C protein in the middle flanked by the H2A and H2B proteins on the ends. The H2C protein consists of 11 β-strands and 2 α-helices and appears to be largely a structural domain that facilitates cohesion of the complex. It weaves together with the N-terminal region (amino acids 1–92) of the H2B protein to form three β-barrels in a motif known as a "triple barrel" (Fig. 2) (46). The triple barrel is formed from a total of 18 parallel and anti-parallel β-strands and produces a pseudo-2-fold axis of symmetry along the central barrel. The intricate interweaving of the β-strands suggests a co-folding of the H2B and H2C proteins rather than a docking of the two and may explain the necessity for all three subunits to generate a stable, active enzyme.

The triple-barrel motif formed by the H2B·H2C subcomplex interactions creates an interface that securely docks the H2A protein. The side and end of barrel 1 form a tight interface with amino acids 197–258 in the C-terminal region of H2A (Fig. 1).

TABLE 1

Crystallographic data and refinement statistics

Space group: $P2_12_12$; unit cell dimensions (Å): a = 285.7, b = 44.6, c = 67.6, $\alpha = \beta = \gamma = 90°$.

	Peak	Remote
Wavelength (Å)	0.9792	0.9780
Resolution (Å)	30-3.1	30-3.1
Unique reflections	17,266	17,409
Completeness (%)	99.1 (92.6)	99.8 (97.4)
R_{merge} (%)[a]	8.3 (48.1)	7.8 (47.6)
Mean I/σ	41.3 (3.1)	43.6 (3.0)
Average redundancy	10.1 (7.2)	10.1 (7.3)
Phasing power	1.13	
R_{cullis}	0.89	
Initial FOM[b]	0.34	
FOM after density modification	0.59	
R-factor (%)[c]	27.5	
R_{free} (%)	31.6	
No. of protein atoms	4,150	
Average B-factor (Å^2)	57	
Root mean square deviation bond lengths (Å)	0.15	
Root mean square deviation bond angles (°)	1.74	

[a] $R_{merge} = \Sigma|I - \langle I\rangle|/\Sigma I$, where I is the observed intensity and $\langle I\rangle$ is the average intensity.
[b] FOM, figure of merit.
[c] $R_{factor} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$. R_{free} is the same as R but calculated with 5% of the reflections that were never used in crystallographic refinement.

The H2A-H2C protein interface buries a total surface area of 2658 Å^2 that is composed mainly of hydrophobic residues, further facilitated by three salt bridges.

The triple-barrel arrangement of the H2B·H2C subcomplex in the mammalian RNase H2 is structurally similar to that of the RAP34-RAP70 protein subdomains of the general eukaryotic transcription factor IIF (TFIIF) complex (Fig. 2) (46). The TFIIF protein complex affects RNA polymerase II transcription initiation and elongation by stabilizing a short RNA·DNA hybrid in the RNA polymerase II active site, inducing promoter DNA wrapping around the polymerase initiation complex (PIC), and mediating protein-protein interactions within the PIC (47–49). The smaller RAP34 (240 amino acids) forms the triple-barrel interaction with the N-terminal region of the larger RAP70 (517 amino acids) and the mostly α-helical C-terminal domain of RAP70 functions in protein-protein interactions (50). By analogy, the smaller mammalian RNase H2C protein interacts with the N-terminal region of H2B, leaving the mostly α-helical C-terminal region of H2B available for potential interactions with other proteins or to facilitate bending of longer polynucleotide molecules. Moreover, the structural similarity between the H2B·H2C proteins and TFIIF coupled with recognition and cleavage of RNA within an RNA·DNA duplex make it tempting to speculate on a role for RNase H2 in transcription, perhaps directly interacting with the RNA polymerase complex through a mechanism similar to that of TFIIF. Consistent with this notion is the finding that overexpression of RNase H in *E. coli* is required for efficient full-length RNA synthesis in the absence of topoisomerase I (51, 52).

The Catalytic H2A Protein—The core of the H2A protein contains an RNase H-like fold (SCOP data base (53)) and has significant structural homology to the archaeal RNase HII enzyme (18–20), as well as to the bacterial endonuclease V protein (54) (Z scores >6, Dali protein structure data base search (55)). Additionally, the mammalian H2A protein contains approximate 30-amino acid N-terminal and 45-amino acid C-terminal extensions that are not found in the bacterial or

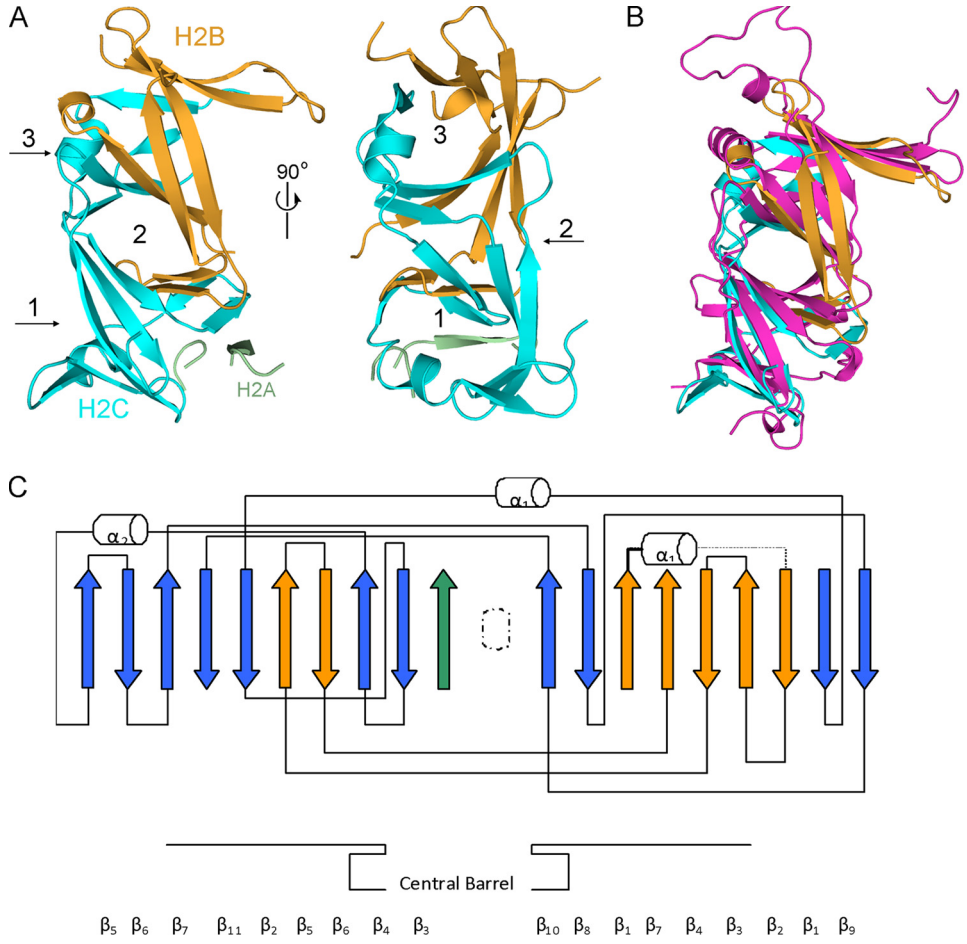

FIGURE 2. **Interactions within RNase H2 triple-barrel motif.** *A*, eighteen strands contributed from the H2C (*blue*) and N-terminal H2B (*gold*) proteins interact in parallel and antiparallel fashion to form three perpendicular barrels (*numbered*). The end and side of barrel 1 form a platform for the stable interaction with the H2A protein (*green*). *B*, a superposition of the triple-barrel motifs from the H2B and H2C proteins on the eukaryotic transcription factor IIF (TFIIF) (*magenta*) reveals a similar architecture. The structural similarity between the two complexes suggests a role for DNA binding and bending as well as protein interaction for H2B/H2C as seen in TFIIF. *C*, schematic of triple-barrel topology formed in the mammalian RNase H2. Strands from the H2C protein are in *blue*, H2B in *gold*, and H2A in *green*.

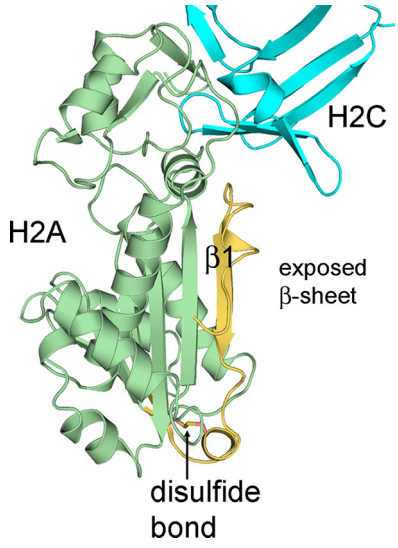

FIGURE 3. **The mammalian RNase H2A protein has N- and C-terminal domains not present in bacterial RNase H2 proteins.** The N-terminal domain (*yellow*) of mouse RNase H2A forms the first strand of the central β-sheet that forms the core of the protein. A disulfide bond (Cys-24 to Cys-29) helps anchor the strand in place. The resulting exposed edge of the β-sheet strongly suggests an additional protein-protein interface for the RNase H2 complex.

archaeal RNase HII proteins. The majority of the C-terminal extension (amino acids 258- 301) is disordered in our structure and appears to extend into solution. Perhaps it functions as a protein interaction motif and becomes ordered upon binding to its cognate partner. The N-terminal extension, however, is completely ordered in the structure and forms a β-strand (β1) that begins the central β-sheet that extends through the core of the H2A protein (Fig. 3). At the C-terminal end of the strand there is a disulfide bond formed between Cys-24 and Cys-29 that appears to be rigidly holding it in place. The result of this architectural arrangement is that the edge of the central β-sheet is extended away from the center of the protein and exposed to solution. Because interactions between the edges of β-sheets occur widely in the formation of protein quaternary structures and protein-protein interactions (56), this exposed β-sheet edge strongly suggests an additional protein-protein interface for the H2A enzyme with an unknown partner. The RNase HIII protein from *Bacillus stearothermophilus* is also structurally similar to the mammalian RNase H2A protein and contains an extended N-terminal domain. However, the Bst-RNase HIII N terminus contains ~90 amino acids that structurally resemble the TATA-box binding proteins (57). This domain is also positioned away from the core of the catalytic domain, unlike the N terminus of the RNase H2A protein.

The mammalian RNase H2 active site geometry suggests a two-metal ion-dependent catalytic mechanism. A two-metal ion-mediated catalysis of phosphodiester hydrolysis has been proposed for many phosphoryl hydrolases, including RNase H enzymes, and involves two metal ions jointly coordinated by the nucleic acid substrate and catalytic residues of the enzyme (58). This mechanism requires four acidic residues for the full coordination of the two divalent metal ions. The active site of the catalytic H2A protein is located in a cleft near one end of the complex (Fig. 1) formed by strands β2 and β5, and helices α4 and α5. It is structurally very similar to the archaeal RNase H2 active site (20) and contains four positionally conserved acidic residues (Asp-34, Glu-35, Asp-142, and Asp-170). Since all four of these amino acids are strictly conserved in RNase H2 enzymes across all species, we tested the contributions to catalysis of these residues through site-directed mutagenesis using the mouse RNase H2 and ribonuclease activity assays. The four acidic residues were mutated individually to asparagines or ala-

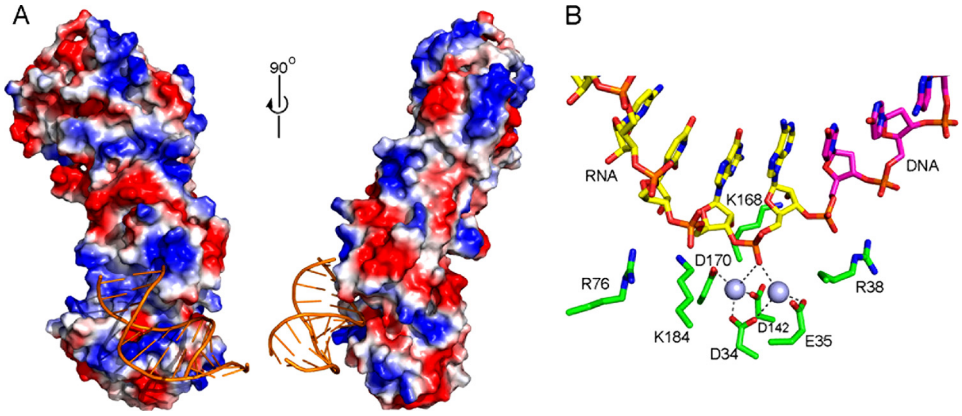

FIGURE 4. **Model of RNA·DNA hybrid binding to the RNase H2 complex.** *A*, a model Okazaki fragment was docked into the active site of the mammalian RNase H2 by a superposition of the bacterial RNase H in complex with a double strand RNA·DNA duplex (PDB ID: 1ZBI; see text for details). The three-protein RNase H2 complex is shown as an electrostatic surface representation. The RNA·DNA duplex was modified to contain an RNA·DNA junction on the ribonucleotide-containing strand. The minor groove of the double strand oligonucleotide straddles the $\beta6-\alpha6$ loop of the protein, which may play a role in substrate recognition. *B*, the model is positioned in the active site to simulate hydrolysis between the last two 3′ ribonucleotides. The four conserved acidic residues (Asp-35, Glu-35, Asp-142, and Asp-170) likely coordinate two divalent metal ions (*blue spheres*) for catalysis. Several positively charged amino acids are positioned to make favorable electrostatic interactions with the bound oligonucleotide, including Lys-168, which may serve as a sensor of the RNA·DNA junction by interacting with the 2′-hydroxyl group of the ribonucleotides. To illustrate the position of the RNA·DNA junction, ribonucleotides are shown in *yellow* and deoxyribonucleotides are shown in *magenta*.

nine (D34N, E35A, D142N, and D170N), and the mutant heterotrimeric RNase H2 complexes were purified (supplemental Fig. S1). The ribonuclease H activity was measured using a model Okazaki fragment substrate containing an RNA_{20}-DNA_{10} annealed to the complementary DNA_{30} (36). The wild-type mouse RNase H2 exhibits catalytic activity similar to that of the human RNase H2, and the four mutant complexes exhibit no detectible activity indicating a >200-fold loss in catalytic function. These data confirm the likely role of these four carboxylate residues in coordination of divalent metal ion and support a two-metal ion mediated mechanism for RNase H2 catalysis (supplemental Fig. S1).

Model for Nucleic Acid Interaction—We have created a model for an Okazaki fragment binding to the mouse RNase H2 complex. Because there are currently no structures of an RNase H2 enzyme in complex with nucleic acids, we used the *Bacillus halodurans* RNase H1·RNA·DNA complex (PDBID: 1ZBL) as a starting model. Although the two enzymes are structurally similar, a superposition of the *Bh*·RNase H1 complex onto the RNase H2A protein based on the overall protein structures reveals several unique features in the RNase H2 enzyme. First, the RNA·DNA hybrid makes significant steric clashes with the RNase H2A protein indicating that substrate binding is considerably different than is observed in the Bh·RNase H enzyme. Secondly, the positioning of the four catalytic acidic residues within the two active sites is not sufficiently well conserved to support an RNA·DNA binding model in RNase H2 that would support hydrolysis of any phosphodiester bonds. These issues are resolved by considering the relative geometry of the acidic residues in the active site. The precise positioning and separation of divalent metal ions is essential in two-metal ion catalysis (58); therefore, we reasoned that the relative positioning of the coordinating amino acid residues should also be conserved. By considering the geometric and distance relationship of the four metal chelating, acidic residues in multiple enzymes that utilize a two-metal ion mechanism of catalysis we identified a conserved trigonal pyramidal arrangement of the residues with relatively similar distances and geometry among all the enzymes (supplemental Fig. S2). The *Bh*·RNase H1·DNA complex was superimposed onto the RNase H2A protein using only the trigonal pyramidal arrangement of the catalytic residues (Fig. 4). The double-stranded RNA·DNA oligonucleotide of the Bh·RNase H1 complex was then modified to contain an RNA·DNA junction on the substrate strand that mimics an Okazaki fragment.

The model has several appealing features that help explain the substrate specificity of RNase H2 and highlights important differences about how the prokaryotic and eukaryotic RNase H2 enzymes likely interact with nucleic acids. The double-stranded RNA·DNA molecule runs through the active site cleft and is positioned to make several favorable electrostatic interactions and no significant steric clashes with the protein. The RNA·DNA hybrid is situated so that the target phosphodiester bond is in the proper orientation for nucleophile attack initiated by a two-metal ion chemistry (Fig. 4). The minor groove of the RNA·DNA hybrid straddles the $\beta6-\alpha6$ loop that contains Lys-168, which might act as a sensor for the RNA·DNA junction by forming an interaction with the 2′-hydroxyl group of the ribose in the 3′ nucleotide of the RNA primer. The absence of a 2′-hydroxyl group in DNA would preclude this interaction insuring that only ribonucleotides are positioned in the active site.

Our nucleic acid binding model positions the non-substrate DNA strand such that there is relatively little interaction with the protein that would provide the biological flexibility to recognize multiple nucleic acid duplex conformations (Fig. 4). In contrast to the Bh·RNase H, the mammalian RNase H2A protein contains no apparent second groove on the surface for recognition of the RNA·DNA hybrid substrate. This is consistent with the enzymatic activity of the protein that indicates recognition of single ribonucleotides as well as longer stretches of RNA within RNA·DNA hybrids. Duplex DNA containing a single ribonucleotide retains mainly B-form structure (10), whereas longer RNA·DNA hybrids adopt intermediate A/B-form structure with a minor groove width of ~9.8 Å compared with ~7.5 Å for B-form DNA (59, 60). Therefore, the presence of a fixed surface channel capable of fitting only a single helical conformation of double-stranded nucleic acid would be counterproductive to the RNase H2 enzyme that apparently needs to bind both conformations.

Disease-causing Mutations in RNase H2—Recently, mutations in each of the genes encoding the three proteins of the human RNase H2 enzyme complex have been identified in patients with Aicardi-Goutières syndrome (AGS) (15). AGS

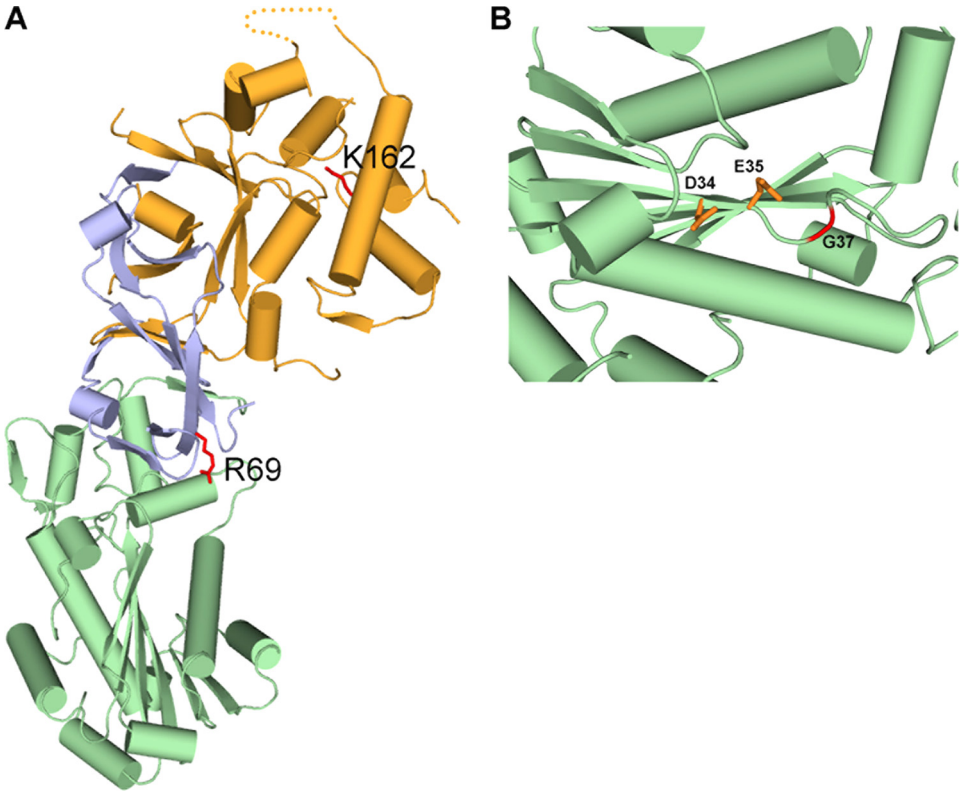

FIGURE 5. **AGS mutations in RNase H2.** *A*, mutation of residue Lys-162 (*red*) to threonine in the H2B protein or residue Arg-69 (*red*) to tryptophan in the H2C protein results in catalytically active enzyme. The structure indicates mutation of K162T may disrupt packing of the helix against the core of the protein that could destabilize potential protein-protein interactions with the distal end of the complex. Likewise, mutation of R69W could potentially disrupt protein interactions with any partner binding the exposed β-sheet of the H2A protein. Residues Ala-177, Val-185, and Tyr-219 are in disordered region shown at the *top* of the complex. *B*, residue Gly-37 (*red*) of the H2A protein lies within the active site near the catalytic Asp-34 and Glu-35 amino acids.

is a rare genetic encephalopathy that mimics a congenital viral infection characterized by elevated interferon-α levels in cerebrospinal fluid. The underlying pathology is likely related to the accumulation of non-processed DNA and RNA intermediates that leads to chronic activation of innate and adaptive immune responses (31, 32). A single missense mutation (G37S) was identified in RNase H2A, two in RNase H2C (R69W and K143I), and at least nine mutations within RNase H2B with K162T, A177T, and V185G as the majority of the cases (Fig. 5). We have previously shown that these mutant proteins retain catalytic activity on RNA·DNA hybrid substrates that mimic an Okazaki fragment suggesting that the functional consequences of the genetic mutations is likely mediated through altered protein-protein interactions or function unrelated to ribonuclease catalytic activity (36).

The amino acids corresponding to three of the identified patient mutations are visible in the structure, and the rest are in disordered loops (Fig. 5). Residue Lys-162 of the H2B protein is located near the surface in the middle of a long helix in the C-terminal region of the protein and seems to play a role in stabilizing the packing of the helix to the core of the protein. The K162T mutation would potentially disrupt that interaction and perhaps allow a conformational change in the helix. In the H2C protein, residue Arg-69 is near the interface of the H2A and H2C proteins and exposed to solution. Modeling of the R69W mutation does not appear to disrupt the protein-protein interface or disturb other local interactions. This residue does however face the extended central β-sheet of the H2A protein. If that edge of the β-sheet is indeed a site for protein-protein interactions with another binding partner, Arg-69 of the H2C protein is in an ideal position to help facilitate that interaction. With both of these mutations a defect in protein-protein interaction rather than catalytic activity is consistent with our structural observations. Finally, residue Gly-37 of the H2A protein is located in very close proximity to the active site (Fig. 5). The flexibility of the glycine residue appears to aid in a turn at the end of the β2 strand. Mutation of Gly-37 to serine (G37S) would put strain on the turn due to the restrained torsion angle geometry and potentially disrupt the positioning of the strand that contains two of the catalytic residues, Asp-34 and Glu-35. This model might predict an alteration in catalytic activity or substrate specificity.

Our previous results using an RNA·DNA substrate mimicking an Okazaki fragment showed that the RNase H2 complex containing the H2A G37S protein exhibits catalytic activity similar to the wild-type enzyme at the preferred enzyme cleavage site near the RNA·DNA junction (36). An additional potential biological role for the RNase H2 complex is to remove single ribonucleotides from double-stranded DNA that might be introduced by DNA polymerases during DNA replication or repair. To investigate the impact of this specific mutation on this process, we tested the RNase H2 G37S mutant catalytic activity using a single ribonucleotide-containing oligonucleotide hybridized to the complementary DNA to generate the double-stranded polynucleotide substrate (Fig. 6). The RNase H2 G37S mutant protein exhibits about a 30-fold lower catalytic activity using the single ribonucleotide substrate relative to the wild-type RNase H2. Two issues have been considered to account for this dramatic reduction in catalytic efficiency. First, the mammalian RNase H2-preferred site of polynucleotide cleavage using a multiple ribonucleotide-containing RNA·DNA duplex is between the penultimate and junction ribonucleotides positioned at the RNA·DNA junction (24, 36). The single ribonucleotide substrate does not contain a ribonucleotide-ribonucleotide phosphodiester bond, and the wild-type RNase H2 cleaves the 5′-deoxyribonucleotide–ribonucleotide bond at the RNA·DNA junction with similar efficiency to that of the ribonucleotide–ribonucleotide phosphodiester bond. Thus, it is apparent that the G37S mutation differentially affects RNase

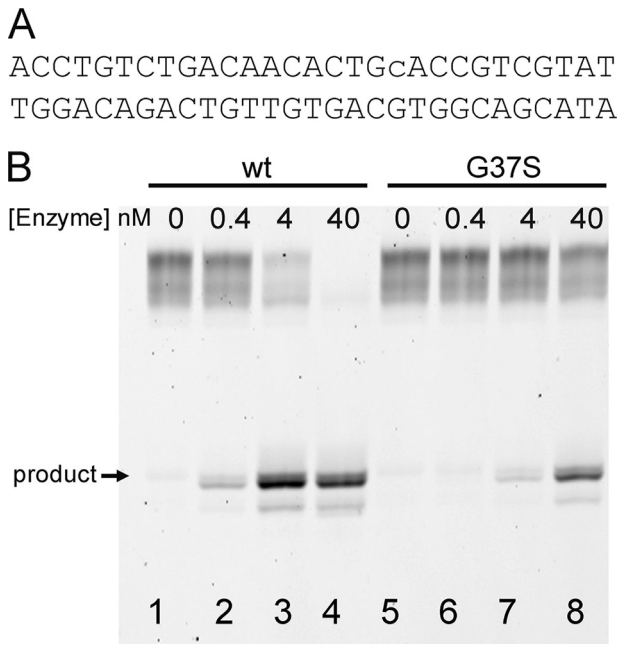

FIGURE 6. **Activity of RNase H2 on substrate containing a single ribonucleotide.** *A*, sequence of oligonucleotide substrate with ribonucleotide indicated in *lowercase* and deoxyribonucleotides in *uppercase*. *B*, RNase H2 WT (*lane 2*, 0.4 nM; *lane 3*, 4 nM; and *lane 4*, 40 nM) or G37S mutant (*lane 6*, 0.4 nM; *lane 7*, 4 nM; and *lane 8*, 40 nM) enzyme complex was incubated with oligonucleotide. *Lanes 1* and *5* contain no RNase H2. Quantification of the RNase H2A activity indicates that RNase H2 WT generates the products at 67 nM/nmol enzyme compared with 2.2 nM/nmol for the RNase H2 G37S mutant enzyme indicating a 30-fold reduction in activity.

H2 activity using multiple ribonucleotide and single ribonucleotide substrates that might be attributed to the nature of the specific sugar (ribose or deoxyribose) contained at the position of the site of polynucleotide cleavage. Second, the dramatic reduction in catalytic activity of the G37S mutation observed using the substrate containing a single ribonucleotide relative to the substrate containing multiple ribonucleotides might be explained by considering the different conformations of the RNA·DNA and DNA·DNA nucleic acid helices present in the two different substrates. Our data suggest that the G37S mutation of the H2A protein distorts the active site in a way that only allows the binding of the wider minor groove present in an RNA·DNA hybrid with a longer stretch of ribonucleotides and not the smaller minor groove expected in a DNA·DNA duplex with only a single ribonucleotide. These data imply that one important biological role for the RNase H2 complex is to remove single ribonucleotides from double-stranded DNA and that failure to do so might be linked to activation of the immune response.

Conclusions—Our structural and biochemical analysis of the mammalian RNase H2 enzyme reveals the three-domain structure of the complex and the protein-protein interactions that stabilize its formation. The structure provides clues to how additional protein interactions might form that would facilitate multiple cellular functions. Four conserved acidic residues are identified that likely contribute to a two-metal ion mechanism of ribonucleotide hydrolysis and provide the basis for a model for double-stranded RNA·DNA polynucleotide substrate binding to the complex. The model offers insight into the mechanism for substrate specificity by the catalytic H2A protein, including recognition of the RNA·DNA junction. Finally, we can begin to understand the structural basis of biological dysfunction for some of the disease-causing mutations associated with RNase H2. The unique structural and functional aspects of the mammalian RNase H2 enzyme will provide continued avenues of biological investigation for this complex and predict new and exciting aspects of nucleic acid metabolism.

Acknowledgments—We gratefully acknowledge the help of Annie Héroux and the staff of the National Synchrotron Light Source, Brookhaven National Labs for help with x-ray data collection.

REFERENCES

1. Brown, T. A., Tkachuk, A. N., and Clayton, D. A. (2008) *J. Biol. Chem.* **283,** 36743–36751
2. Itoh, T., and Tomizawa, J. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77,** 2450–2454
3. Ogawa, T., and Okazaki, T. (1980) *Annu. Rev. Biochem.* **49,** 421–457
4. Okazaki, T., Kurosawa, Y., Ogawa, T., Seki, T., Shinozaki, K., Hirose, S., Fujiyama, A., Kohara, Y., Machida, Y., Tamanoid, F., and Hozumi, T. (1979) *Cold Spring Harb. Symp. Quant. Biol.* **43,** 203–219
5. Rossi, M. L., Purohit, V., Brandt, P. D., and Bambara, R. A. (2006) *Chem. Rev.* **106,** 453–473
6. Gilbert, W. (1986) *Nature* **319,** 618
7. Jern, P., and Coffin, J. M. (2008) *Annu. Rev. Genet.* **42,** 709–732
8. Temin, H. M. (1985) *Mol. Biol. Evol.* **2,** 455–468
9. Ohtani, N., Haruki, M., Morikawa, M., Crouch, R. J., Itaya, M., and Kanaya, S. (1999) *Biochemistry* **38,** 605–618
10. Eder, P. S., Walder, R. Y., and Walder, J. A. (1993) *Biochimie* **75,** 123–126
11. Haruki, M., Tsunaka, Y., Morikawa, M., and Kanaya, S. (2002) *FEBS Lett.* **531,** 204–208
12. Cerritelli, S. M., and Crouch, R. J. (2009) *FEBS J.* **276,** 1494–1505
13. Frank, P., Albert, S., Cazenave, C., and Toulmé, J. J. (1994) *Nucleic Acids Res.* **22,** 5247–5254
14. Frank, P., Braunshofer-Reiter, C., Wintersberger, U., Grimm R., and Büsen, W. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95,** 12872–12877
15. Crow, Y. J., Leitch, A., Hayward, B. E., Garner, A., Parmar, R., Griffith, E., Ali, M., Semple, C., Aicardi, J., Babul-Hirji, R., Baumann, C., Baxter, P., Bertini, E., Chandler, K. E., Chitayat, D., Cau, D., Déry, C., Fazzi, E., Goizet, C., King, M. D., Klepper, J., Lacombe, D., Lanzi, G., Lyall, H., Martínez-Frias, M. L., Mathieu, M., McKeown, C., Monier, A., Oade, Y., Quarrell, O. W., Rittey, C. D., Rogers, R. C., Sanchis, A., Stephenson, J. B., Tacke, U., Till, M., Tolmie, J. L., Tomlin, P., Voit, T., Weschke, B., Woods, C. G., Lebon, P., Bonthron, D. T., Ponting, C. P., and Jackson, A. P. (2006) *Nat. Genet.* **38,** 910–916
16. Pallan, P. S., and Egli, M. (2008) *Cell Cycle* **7,** 2562–2569
17. Nowotny, M., Gaidamakov, S. A., Crouch, R. J., and Yang, W. (2005) *Cell* **121,** 1005–1016
18. Muroya, A., Tsuchiya, D., Ishikawa, M., Haruki, M., Morikawa, M., Kanaya, S., and Morikawa, K. (2001) *Protein Sci.* **10,** 707–714
19. Lai, L., Yokota, H., Hung, L. W., Kim, R., and Kim, S. H. (2000) *Structure* **8,** 897–904
20. Chapados, B. R., Chai, Q., Hosfield, D. J., Qiu, J., Shen, B., and Tainer, J. A. (2001) *J. Mol. Biol.* **307,** 541–556
21. Qiu, J., Qian, Y., Frank, P., Wintersberger, U., and Shen, B. (1999) *Mol. Cell. Biol.* **19,** 8361–8371
22. Rydberg, B., and Game, J. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99,** 16654–16659
23. Ayyagari, R., Gomes, X. V., Gordenin, D. A., and Burgers, P. M. (2003) *J. Biol. Chem.* **278,** 1618–1625
24. Chon, H., Vassilev, A., Depamphilis, M. L., Zhao, Y., Zhang, J., Burgers, P. M., Crouch, R. J., and Cerritelli, S. M. (2009) *Nucleic Acids Res.* **37,** 96–110
25. Bergoglio, V., Ferrari, E., Hübscher, U., Cazaux, C., and Hoffmann, J. S.

(2003) *J. Mol. Biol.* **331,** 1017–1023
26. Ide, H., Yagi, R., Yamaoka, T., and Kimura, Y. (1993) *Nucleic Acids Symp. Ser.* **29,** 133–134
27. Nick McElhinny, S. A., and Ramsden, D. A. (2003) *Mol. Cell. Biol.* **23,** 2309–2315
28. Crow, Y. J., Hayward, B. E., Parmar, R., Robins, P., Leitch, A., Ali, M., Black, D. N., van Bokhoven, H., Brunner, H. G., Hamel, B. C., Corry, P. C., Cowan, F. M., Frints, S. G., Klepper, J., Livingston, J. H., Lynch, S. A., Massey, R. F., Meritet, J. F., Michaud, J. L., Ponsot, G., Voit, T., Lebon, P., Bonthron, D. T., Jackson, A. P., Barnes, D. E., and Lindahl, T. (2006) *Nat. Genet.* **38,** 917–920
29. Rice, G., Newman, W. G., Dean, J., Patrick, T., Parmar, R., Flintoff, K., Robins, P., Harvey, S., Hollis, T., O'Hara, A., Herrick, A. L., Bowden, A. P., Perrino, F. W., Lindahl, T., Barnes, D. E., and Crow, Y. J. (2007) *Am. J. Hum. Genet.* **80,** 811–815
30. Crow, Y. J., and Livingston, J. H. (2008) *Dev. Med. Child Neurol.* **50,** 410–416
31. Rice, G., Patrick, T., Parmar, R., Taylor, C. F., Aeby, A., Aicardi, J., Artuch, R., Montalto, S. A., Bacino, C. A., Barroso, B., Baxter, P., Benko, W. S., Bergmann, C., Bertini, E., Biancheri, R., Blair, E. M., Blau, N., Bonthron, D. T., Briggs, T., Brueton, L. A., Brunner, H. G., Burke, C. J., Carr, I. M., Carvalho, D. R., Chandler, K. E., Christen, H. J., Corry, P. C., Cowan, F. M., Cox, H., D'Arrigo, S., Dean, J., De Laet, C., De Praeter, C., Dery, C., Ferrie, C. D., Flintoff, K., Frints, S. G., Garcia-Cazorla, A., Gener, B., Goizet, C., Goutières, F., Green, A. J., Guet, A., Hamel, B. C., Hayward, B. E., Heiberg, A., Hennekam, R. C., Husson, M., Jackson, A. P., Jayatunga, R., Jiang, Y. H., Kant, S. G., Kao, A., King, M. D., Kingston, H. M., Klepper, J., van der Knaap, M. S., Kornberg, A. J., Kotzot, D., Kratzer, W., Lacombe, D., Lagae, L., Landrieu, P. G., Lanzi, G., Leitch, A., Lim, M. J., Livingston, J. H., Lourenco, C. M., Lyall, E. G., Lynch, S. A., Lyons, M. J., Marom, D., McClure, J. P., McWilliam, R., Melancon, S. B., Mewasingh, L. D., Moutard, M. L., Nischal, K. K., Ostergaard, J. R., Prendiville, J., Rasmussen, M., Rogers, R. C., Roland, D., Rosser, E. M., Rostasy, K., Roubertie, A., Sanchis, A., Schiffmann, R., Scholl-Burgi, S., Seal, S., Shalev, S. A., Corcoles, C. S., Sinha, G. P., Soler, D., Spiegel, R., Stephenson, J. B., Tacke, U., Tan, T. Y., Till, M., Tolmie, J. L., Tomlin, P., Vagnarelli, F., Valente, E. M., Van Coster, R. N., Van der Aa, N., Vanderver, A., Vles, J. S., Voit, T., Wassmer, E., Weschke, B., Whiteford, M. L., Willemsen, M. A., Zankl, A., Zuberi, S. M., Orcesi, S., Fazzi, E., Lebon, P., and Crow, Y. J. (2007) *Am. J. Hum. Genet.* **81,** 713–725
32. Rigby, R. E., Leitch, A., and Jackson, A. P. (2008) *BioEssays* **30,** 833–842
33. Morita, M., Stamp, G., Robins, P., Dulic, A., Rosewell, I., Hrivnak, G., Daly, G., Lindahl, T., and Barnes, D. E. (2004) *Mol. Cell. Biol.* **24,** 6719–6727
34. Yang, Y. G., Lindahl, T., and Barnes, D. E. (2007) *Cell* **131,** 873–886
35. Stetson, D. B., Ko, J. S., Heidmann, T., and Medzhitov, R. (2008) *Cell* **134,** 587–598
36. Perrino, F. W., Harvey, S., Shaban, N. M., and Hollis, T. (2009) *J. Mol. Med.* **87,** 25–30
37. de Silva, U., Choudhury, S., Bailey, S. L., Harvey, S., Perrino, F. W., and Hollis, T. (2007) *J. Biol. Chem.* **282,** 10537–10543
38. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
39. Doublié, S. (1997) *Methods Enzymol.* **276,** 523–530
40. CCP4 (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50,** 760–763
41. Terwilliger, T. C. (2000) *Acta Crystallogr. D Biol. Crystallogr.* **56,** 965–972
42. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
43. Skubák, P., Murshudov, G. N., and Pannu, N. S. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2196–2201
44. Brünger, A. T. (1992) *Nature* **355,** 472–475
45. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) *Nucleic Acids Res.* **35,** W375–W383, web server issue
46. Gaiser, F., Tan, S., and Richmond, T. J. (2000) *J. Mol. Biol.* **302,** 1119–1127
47. Khaperskyy, D. A., Ammerman, M. L., Majovski, R. C., and Ponticelli, A. S. (2008) *Mol. Cell. Biol.* **28,** 3757–3766
48. Robert, F., and Coulombe, B. (2001) *Methods Mol. Biol.* **148,** 383–393
49. Robert, F., Douziech, M., Forget, D., Egly, J. M., Greenblatt, J., Burton, Z. F., and Coulombe, B. (1998) *Mol. Cell* **2,** 341–351
50. Kamada, K., Roeder, R. G., and Burley, S. K. (2003) *Proc. Natl. Acad. Sci. U.S.A.* **100,** 2296–2299
51. Baaklini, I., Hraiky, C., Rallu, F., Tse-Dinh, Y. C., and Drolet, M. (2004) *Mol. Microbiol.* **54,** 198–211
52. Hraiky, C., Raymond, M. A., and Drolet, M. (2000) *J. Biol. Chem.* **275,** 11257–11263
53. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) *J. Mol. Biol.* **247,** 536–540
54. Dalhus, B., Arvai, A. S., Rosnes, I., Olsen, Ø. E., Backe, P. H., Alseth, I., Gao, H., Cao, W., Tainer, J. A., and Bjørås, M. (2009) *Nat. Struct. Mol. Biol.* **16,** 138–143
55. Holm, L., Kääriäinen, S., Rosenström, P., and Schenkel, A. (2008) *Bioinformatics* **24,** 2780–2781
56. Maitra, S., and Nowick, J. S. (2000) in *The Amide Linkage: Structural Significance in Chemistry, Biochemistry, and Material Science* (Greenberg, A., Breneman, C. M., and Liebman, J. F., eds) pp. 495–518, Wiley, New York
57. Chon, H., Matsumura, H., Koga, Y., Takano, K., and Kanaya, S. (2006) *J. Mol. Biol.* **356,** 165–178
58. Yang, W., Lee, J. Y., and Nowotny, M. (2006) *Mol. Cell* **22,** 5–13
59. Horton, N. C., and Finzel, B. C. (1996) *J. Mol. Biol.* **264,** 521–533
60. Kopka, M. L., Lavelle, L., Han, G. W., Ng, H. L., and Dickerson, R. E. (2003) *J. Mol. Biol.* **334,** 653–665