Structural studies of trypanosoma brucei RNA editing ligases and their binding partner proteins

Shaneh, Alireza; Purisima, Enrico O.; Salavati, Reza; Sulea, Traian

Publisher's version / Version de l'éditeur:
https://doi.org/10.1021/acs.biochem.5b01257
Biochemistry, 55, 16, pp. 2319-2331, 2016-03-31

BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Structural Studies of *Trypanosoma brucei* RNA Editing Ligases and Their Binding Partner Proteins

Alireza Shaneh,[†,∥] Enrico O. Purisima,[∥,‡,§] Reza Salavati,[†,∥,§] and Traian Sulea*[,†,‡]

[†]Institute of Parasitology, McGill University, 21111 Lakeshore Road, Sainte-Anne-de-Bellevue, H9X 3V9 Quebec, Canada

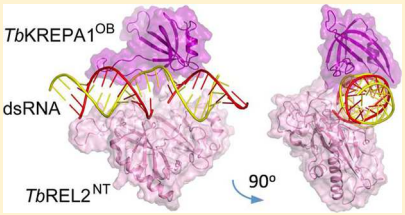
[‡]Human Health Therapeutics, National Research Council Canada, 6100 Royalmount Avenue, Montreal, H4P 2R2 Quebec, Canada

[§]Department of Biochemistry, McGill University, McIntyre Medical Building, 3655 Promenade Sir William Osler, Montreal, H3G 1Y6 Quebec, Canada

[∥]McGill Centre for Bioinformatics, McGill University, Bellini Building, 3649 Promenade Sir William Osler, Montreal, H3G 0B1 Quebec, Canada

Supporting Information

ABSTRACT: To study the mechanism of ligating nicked RNA strands, we conducted molecular dynamics simulations of *Trypanosoma brucei* RNA editing ligases L1 and L2 complexed with double-stranded RNA (dsRNA) fragments. In each resulting model, a Mg^{2+} ion coordinates the 5′-PO_4 of the nicked nucleotide and the 3′-OH of the terminal nucleotide for a nucleophilic reaction consistent with the postulated step 3 chemistry of the ligation mechanism. Moreover, coordination of the 3′-OH to the Mg^{2+} ion may lower its pK_a, thereby rendering it a more effective nucleophile as an oxyanion. Thus, Mg^{2+} may play a twofold role: bringing the reactants into the proximity of each other and activating the nucleophile. We also conducted solvated interaction energy calculations to explore whether ligation specificities can be correlated to ligase−dsRNA binding affinity changes. The calculated dsRNA binding affinities are stronger for both L1 and L2 when the terminal nucleotide is changed from cytosine to guanine, in line with their experimentally measured ligation specificities. Because the ligation mechanism is also influenced by interactions of the ligase with partner proteins from the editosome subcomplex, we also modeled the structure of the RNA-bound L2 in complex with the oligonucleotide binding (OB) domain of largest editosome interacting protein A1. The resulting L2−dsRNA−A1 model, which is consistent with mutagenesis and binding data recorded to date, provides the first atomic-level glimpse of plausible interactions around the RNA ligation site in the presence of an OB domain presented *in-trans* to a nucleic acid ligase.



Polynucleotide ligases join the nicked strands of DNA or RNA, thereby playing an important role in nucleic acid repair. Two classes of RNA ligases, NAD^+- and ATP-dependent ligases, seal RNA fragments by catalyzing the formation of a phosphodiester bond between the 3′-hydroxyl group of the upstream fragment and the 5′-phosphate group of the downstream RNA fragment.

Trypanosoma brucei RNA editing ligases 1 and 2 (*Tb*REL1 and *Tb*REL2, termed L1 and L2 in this paper, respectively) exemplify two families of ATP-dependent ligases that share a strong degree of homology in their sequences.[1−3] L1 and L2 catalyze the ligation step of RNA editing, a post-transcriptional modification process in which uridylate is either inserted into (U-insertion) or deleted from (U-deletion) cleaved mitochondrial mRNA transcripts, resulting in mature mRNAs that are translated to functional proteins.

L1 and L2 belong to the covalent nucleotidyltransferase superfamily. The superfamily sequences share five signature motifs that are important for catalytic activities of the enzymes. While L1 and L2 sequences are 41% identical and 61% similar, it has been proposed that L1 plays a role in U-deletion and L2 in U-insertion.[4−6] Deng and colleagues have determined the crystal structure of the L1 N-terminal catalytic domain in complex with ATP at an atomic resolution of 1.2 Å.[7] Furthermore, on the basis of the strong degree of similarity between L1 and L2 N-terminal domains (49% identical and 66% homologous sequence), we have modeled the structure of the L2 N-terminal catalytic domain and shown that the two catalytic domains differ in their electrostatic properties.[8]

The mechanism of ligation of ATP-dependent ligases entails three chemical steps (reviewed in ref 9). In the first step, ATP adenylates the enzyme by covalently bonding to a conserved lysine residue in the catalytic cleft, releasing pyrophospate and forming an enzyme−AMP intermediate complex. In the second step, the AMP is transferred to the 5′-PO_4 group of the downstream fragment and makes an AMP−5′-RNA complex. Finally, the 3′-OH of the upstream fragment acts as a nucleophile and attacks the phosphate group of the AMP−5′-RNA complex, forming a phosphodiester bond and releasing AMP. A series of crystal structure snapshots capturing the three steps of nucleotidyl transfer chemistry in T4 RNA ligase 2 (T4-Rnl2) have been determined.[9] These include the enzyme−AMP intermediate, the AMP−5′-RNA intermediate representing the

Received: November 20, 2015
Revised: March 22, 2016
Published: March 31, 2016

DOI: 10.1021/acs.biochem.5b01257
Biochemistry 2016, 55, 2319−2331

step 2 product, and a conformational transition of the AMP−5′-RNA intermediate representing the step 3 substrate.

DNA and RNA ligases have different specificities for their polynucleotide substrates. DNA ligases require DNA as the 3′-fragment, but they catalyze the ligation with either DNA or RNA 5′-fragment.[10−12] In contrast, RNA ligases are indifferent to whether the downstream fragment is DNA or RNA; however, to catalyze ligation, RNA ligases need the upstream fragment to be RNA.[13] In addition, the specificity for the RNA substrates arises from different bases in the last two nucleotides of the upstream RNA fragment. In particular, Palazzo and colleagues[14] have shown that the ligation activities of L1 and L2 vary as the terminal base of the 5′-fragment is changed to another base. They have also shown that recombinant L1 and L2 attain maximal ligation activity when the terminal base of the upstream fragment is guanine, though their ligation activity levels differ from each other with higher activity for L1.

Molecular dynamics (MD) simulations are powerful tools for conducting structural analyses of proteins, nucleic acids, and their complexes.[15] Hence, Amaro and colleagues studied the dynamics of L1 in the presence and absence of ATP.[16] During the simulations of the apoenzyme, they observed a significant rearrangement of the ATP binding pocket leading to widening of the cavity, suggesting that L1 has structural flexibility around the catalytic pocket. Another MD-based study of L1 in complex with the double-stranded RNA (dsRNA) substrate described the L1−RNA interactions and predicted water-mediated coordination of a Mg^{2+} ion in the active site.[17]

The basal RNA ligation activities of L1 and L2 are further modulated by specific interactions with editosomal protein components.[6] Using biochemical and structural analysis, Park and colleagues have delineated the domains of L2 and its editosomal binding protein KREPA1 (termed A1 in this paper) that interact with each other.[18] In this model (Figure 1), the C-terminal segment of L2 is recruited by A1 via a region encompassing its first Zn-finger domain. In turn, the proximity of A1 allows an *in-trans* interaction of its oligonucleotide binding (OB)-folded C-terminal domain with the L2−dsRNA complex to assist ligation. Because of the high degree of homology between the L1 and L2 ligases and their interaction partners, it is plausible that this interaction mode will be prototypical for the L1 ligase interacting with its binding partner KREPA2 (termed A2 in this paper). However, the L1−A2 and L2−A1 interactions are highly specific as the L1−A1 and L2−A2 cross-interactions have not been established.[6] By mutating several residues on the surface of the A1 OB-fold domain, Park and co-workers identified three arginine residues that are important for the binding specificity to RNA-bound L2.[18] However, the structural details underlying the molecular interactions among all these domains remain largely elusive. Also poorly understood are the mechanistic implications that these putative *in-trans* interactions may have on the RNA ligation activity vis-à-vis the activation mechanisms proposed for *in-cis* interactions surrounding DNA ligation.[19]

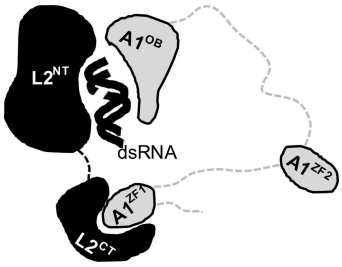


Figure 1. Cartoon representation of putative contact surfaces between the largest editosome interacting protein A1 (comprising the OB-folded domain $A1^{OB}$ and two Zn-finger domains $A1^{ZF1}$ and $A1^{ZF2}$) and the RNA editing ligase L2 (comprising an N-terminal catalytic domain $L2^{NT}$ and a C-terminal domain $L2^{CT}$).

In this paper, we first conducted MD simulations of the L1 and L2 N-terminal catalytic domains in the presence of a dsRNA substrate. The main objectives of these simulations were to shed some light on the structural roles of the metal ion in mediating the ligation reaction and to investigate possible differences between the interactions of L1 and L2 with RNA. Simulated ligase−RNA complexes were also used to investigate whether nucleotide specificity at the ligation site can be explained by relative substrate binding affinities. To go a step further into the modulation of ligation activity, we then used the predicted structure of the L2−dsRNA complex for building a structural model that includes an interaction with the OB-fold domain of binding protein A1. We will show that the L2−dsRNA−A1 ternary model is consistent with the experimental evidence recorded to date and highlights novel interactions that are useful for guiding further mutagenesis experiments.

MATERIALS AND METHODS

Assembly and Preparation of L1−RNA and L2−RNA Complexes. Each assembled complex consisted of protein, nicked dsRNA substrate, and a magnesium ion. We utilized the available crystal structure of the L1 N-terminal domain (termed $L1^{NT}$ in this paper) in complex with ATP [Protein Data Bank (PDB) entry 1XDN][7] as a starting structure for protein modeling. We have reported the procedure of L2 N-terminal domain ($L2^{NT}$) homology modeling elsewhere.[8] We used the crystal structure of T4-Rnl2 in complex with a DNA−RNA hybrid (PDB entry 2HVR) as a starting point for protein−RNA modeling. Using SYBYL version 8.1.1 (Tripos, Inc., St. Louis, MO), we built a 27-mer A-form structural model of a dsRNA fragment the with sequence

```
          ++++++++++++  -------------
g(5′-UUUUUCCGAUAGUGGGGGUCGCAAUUG-3′
  3′-AAAAAGGCUAUCACCCCCAGCGUUAAC-5′)
     987654321012345678 9
```

The length of the dsRNA ensured that both ends of the strands extended a few nucleotides beyond the point of contact with the protein. The motivation was to mitigate edge effects in the protein−RNA interactions. The adenylated nicked base (underlined, position 0) consists of an adenosine moiety and a cytosine base. We used the AMBER software package[20] to build the model of the nicked base from the RA3 and RC nucleotide residues of the AMBER library and bridged them with an oxygen atom. We then superimposed the resulting dsRNA model onto the DNA−RNA hybrid from the crystal structure of T4-Rnl2. Because of the bent geometry of the template DNA−RNA helix relative to the A-form dsRNA geometry, the modeled dsRNA was clipped at the editing site and the resulting halves were fitted independently to the template, followed by the joining of the gRNA strand and manual reorientation of the neighboring bases to relieve steric clashes. We separately superimposed the $L1^{NT}$ and $L2^{NT}$ structures onto the catalytic domain from the crystal structure of T4-Rnl2. We then merged the dsRNA model with each of the $L1^{NT}$ and $L2^{NT}$ structures to yield initial models of the $L1^{NT}$−dsRNA and $L2^{NT}$−dsRNA complexes.

DOI: 10.1021/acs.biochem.5b01257
Biochemistry 2016, 55, 2319−2331

Force field parameters for the protein were taken from the AMBER FF99SB force field.[21] For the RNA substrate, we used the AMBER parmbsc0 force field[22] supplemented with additional parameters from Meagher et al.[23]

The initial position of a Mg^{2+} ion in the active site was modeled after the crystal structure of $L1^{NT}$ bound to ATP (PDB entry 1XDN), and we have preserved the four coordinated water molecules. In the crystal structure, the other two coordination positions are occupied: one of the γ-phosphoryl oxygen atoms of ATP and an oxygen atom from the β-phosphoryl group. In our model, the pyrophosphoryl group of the nick overlays with the α- and β-phosphoryl groups of ATP in the crystal structure, maintaining the coordination of Mg^{2+} by the corresponding phosphoryl oxygen. The 3′-OH of the $C(-1)^{\alpha}$ nucleotide occupies the coordination site previously occupied by the γ-phosphate of ATP. The positioning in L2 of the Mg^{2+} ion and its four coordinated water molecules was done by overlaying the $L2^{NT}$ modeled structure with $L1^{NT}$ and merging the Mg^{2+} and water molecules into L2.

We employed the cationic dummy atom method[24−26] to incorporate a magnesium ion within each complex. In the cationic dummy atom scheme, instead of a divalent ion being characterized as a spherical ion with a formal positive charge at the center, the formal charge is distributed equally among interaction points arranged in an octahedral or a tetrahedral geometry around the center. Operationally, these interaction points are constructed as dummy atoms linked by dummy bonds to the metal center. Various studies[24−26] have shown that such a charge distribution over cationic dummy atoms helps maintain the coordination geometry during the MD simulation relative to having the single-point total formal charge at the center of the spherical ion. Because the crystal structure of $L1^{NT}$ shows the Mg^{2+} ion in octahedral geometry, we built an octahedral ion model with six dummy atoms as nodes each charged with a partial charge of 0.33 at a distance of 0.9 Å from the ion center. We used the XLEAP module of AMBER to build the Mg^{2+} cationic dummy model and adopted the force field parameters from Oelschlaeger et al.[24]

Modeling the Unbound OB Domain of Interaction Partner A1. The structure of the C-terminal OB-folded domain of KREPA1 (termed $A1^{OB}$, residues 627−762) was recently determined at 2.65 Å resolution (PDB entry 4DK6) as a homodimer in complex with a nanobody.[18] For molecular docking purposes, we retained one $A1^{OB}$ monomer and reconstructed several missing regions (loops L23 and L56, helix α1) using SYBYL. A 38-residue loop (L23, residues 658−695), which was substituted with a four-residue linker for the purpose of crystallization, was not built here as studies have shown that L23 is not important for RNA binding.[18] The repaired $A1^{OB}$ structure was refined by a 5 ns molecular dynamics simulation in explicit solvent. Because room-temperature MD simulation is not an efficient conformational sampling method, to address the likelihood of induced-fit docking to dsRNA we conducted further conformational sampling of the solvent-exposed and potentially flexible loop L45 of $A1^{OB}$ (residues 729−747), whose conformation could also be affected by crystal packing in the nanobody complex. This search was done in the unbound state of $A1^{OB}$ by full reconstruction of the L45 backbone using the kinematic closure (KIC) method[27] with Monte Carlo-Metropolis sampling as implemented in the Rosetta software.[28,29] The number of loop conformations to be generated was 10000.

Modeling of the $L2^{NT}$−dsRNA−$A1^{OB}$ Complex. A knowledge-based docking approach was taken to model this complex, as our initial protein−nucleic acid *de novo* docking attempts failed to generate results consistent with experimental mutagenesis data. The crystal structure of the human DNA ligase I (LIG1) in complex with dsDNA (PDB entry 1X9N)[10] was used as a template for merging the models of the $A1^{OB}$ domain and the $L2^{NT}$−dsRNA complex. Specifically, we used SwissPDBViewer version 4.0[30] for root-mean-square (rms) fitting of the Cα coordinates of homologous residues between (i) the $L2^{NT}$ domain and the NT domain of LIG1 and (ii) the $A1^{OB}$ domain and the OB domain of LIG1. The resulting structure of the $L2^{NT}$−dsRNA−$A1^{OB}$ complex was submitted to refinement by a 40 ns MD simulation.

Molecular Dynamics Simulations. All MD simulations were conducted with the AMBER version 11.0[20] and the AMBER FF99SB force field.[21,31] The TIP3P[32] explicit water model was used to solvate the solute in a truncated octahedron extending 12 Å around the solute. Counterions were added to neutralize the system to a final salt concentration of 0.1 M. Our molecular dynamics simulation workflow started with a minimization routine. In the first round of minimization, we conducted 1300 steps of the steepest descent along with 700 steps of conjugate gradient algorithm. A harmonic restraint was added on the solute (k = 500 kcal mol^{-1} Å^{-2}) while the solvent and ion molecules were free to move. In the next round of minimization, the system was relaxed with 3000 steps of steepest descent and 2000 steps of conjugate gradient. In the case of $L1^{NT}$−dsRNA and $L2^{NT}$−dsRNA complexes, restrained constant pressure and temperature (*NPT*) molecular dynamics was conducted at 50 K and 1 atm for 40 ps with 10 kcal mol^{-1} Å^{-2} restraints on the protein and RNA residues. Subsequently, the systems were heated starting from 50 to 300 K (room temperature) with a weak harmonic restraint of 10 kcal mol^{-1} Å^{-2} on the solute atoms. We continued heating for 25 ps. A Berendsen heat bath was used for the simulations of $L1^{NT}$−dsRNA and $L2^{NT}$−dsRNA complexes. Pressure and temperature coupling parameters for those simulations were set to 2.0. The next step was restrained heating from 50 to 300 K in 40 ps with the same restraints and bath coupling parameters. The restraints were gradually released in three 40 ps MD runs at 300 K with decreasing restraints of 5.0, 2.0, and 1.0 kcal mol^{-1} Å^{-2}. Finally, for $L1^{NT}$−dsRNA and $L2^{NT}$−dsRNA complexes, we ran 40 ns of unrestrained MD with snapshots saved after every 500 steps (0.8 ps), and in the case of the $L2^{NT}$−dsRNA−$A1^{OB}$ complex, we sampled the production phase for 40 ns. To constrain the bond length of hydrogen atoms, we used SHAKE.[33] For long-range electrostatic treatment, particle mesh Ewald (PME)[34] was used. In case of $L2^{NT}$−dsRNA−$A1^{OB}$ complex, we used Langevin dynamics,[35] and we set the particle collision frequency to 1 ps^{-1} as we maintained the temperature of production phase at 300 K. The time step was set to 1.6 fs (a larger time step was not possible because of the dummy point charges on the Mg^{2+} ion), and we saved the trajectories every 0.8 ps. In the case of the MD simulation of the $L2^{NT}$−dsRNA−$A1^{OB}$ complex, we used a restraint-releasing strategy during the production phase. Thus, we restrained the $L2^{NT}$ domain as well as the nicked nucleotide and Mg^{2+} to their initial coordinates with harmonic force constants of 100 kcal mol^{-1} Å^{-2}. We then gradually released the restraints by 25 kcal mol^{-1} Å^{-2} per each nanosecond of simulation time; that is, no restraint was applied during the fifth nanosecond and the rest of the production run. Trajectories were analyzed with the PTRAJ module in AMBER.[20] An energy-

DOI: 10.1021/acs.biochem.5b01257
Biochemistry 2016, 55, 2319−2331

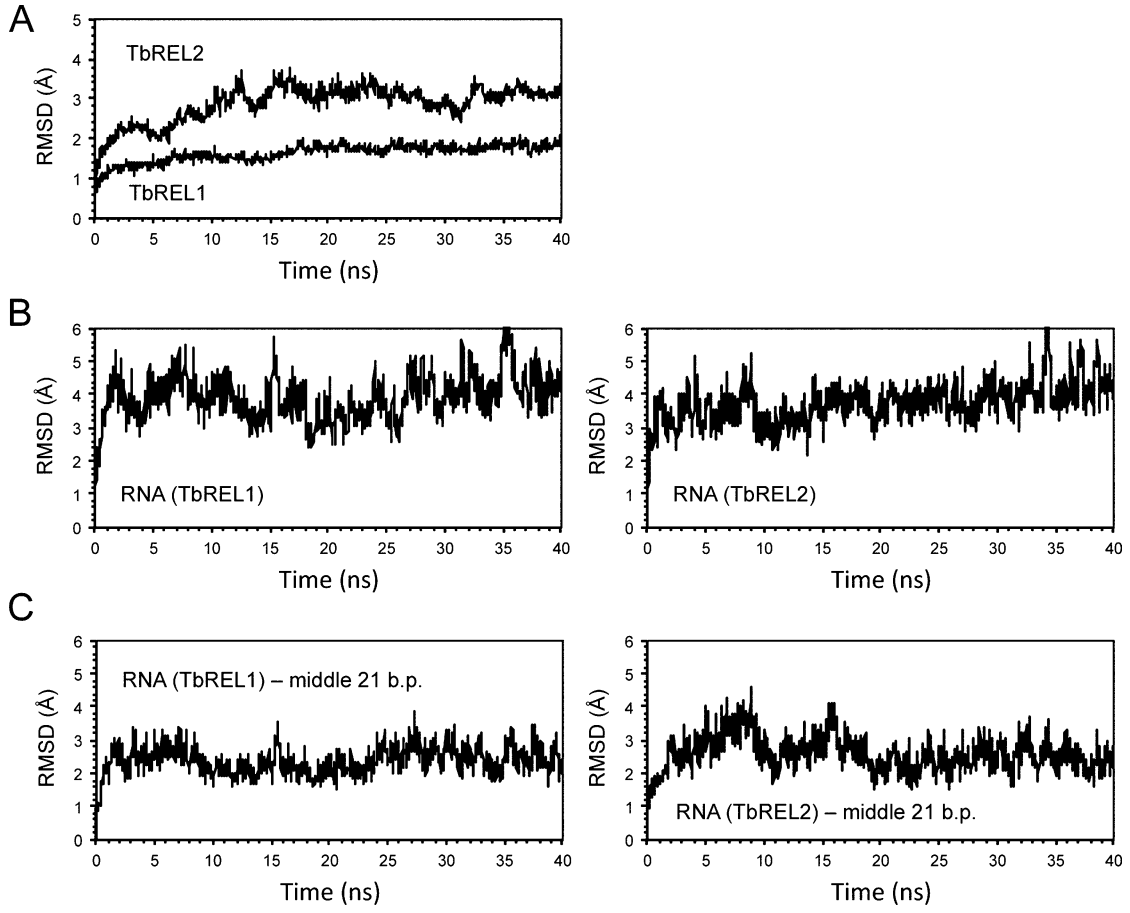


Figure 2. Equilibration of $L1^{NT}$–dsRNA and $L2^{NT}$–dsRNA complexes during the MD trajectory. (A) rmsds from the starting structure of Cα atoms of $L1^{NT}$ and $L2^{NT}$ domains. (B) rmsds from the starting structure of the dsRNA from the L1 and L2 complexes. (C) rmsds from the starting structure of the central 21 bp of dsRNA from the L1 and L2 complexes. Three base pairs from each end of the dsRNA strands were excluded.

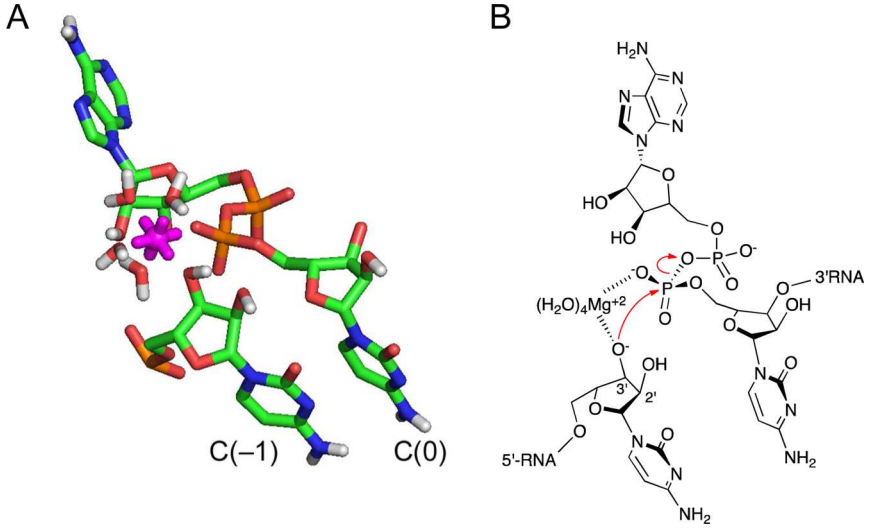


Figure 3. Participation of magnesium ion in the ligation mechanism. (A) Configuration of the nick and C(−1) nucleotides in relation to the coordinated Mg^{2+}. The Mg^{2+} ion is colored magenta with the octahedral distribution of its charged dummy atoms shown as short bonds. (B) Schematic depicting the role of Mg^{2+} in activating the 3′-OH for nucleophilic attack.

minimized average structure for the last nanosecond of simulation was typically generated for structural interpretation.

Binding Free Energy Calculation. To calculate the binding free energy between ligases and the dsRNA substrate, we applied the solvated interaction energy (SIE) binding affinity function:[36]

$$\Delta G_{\text{bind}}^{\text{calc}} \approx \text{SIE} = \alpha(E_{\text{inter}}^{\text{C}} + \Delta G_{\text{bind}}^{\text{R}} + E_{\text{inter}}^{\text{vdw}} + \Delta G_{\text{bind}}^{\text{npsol}}) + C$$

where $E_{\text{inter}}^{\text{C}}$ is the contribution of Coulomb interaction energy, $\Delta G_{\text{bind}}^{\text{R}}$ is the difference in reaction field energy between bound

DOI: 10.1021/acs.biochem.5b01257
Biochemistry 2016, 55, 2319−2331

and free states, E_{inter}^{vdw} is the intermolecular van der Waals energy, and ΔG_{bind}^{npsol} is the change in nonpolar solvation energy upon binding. We used the default values of $\alpha = 0.1048$ kcal/mol and $C = -2.89$ kcal/mol. The validity of SIE has been proven by over a wide variety of applications.[37,38] We used SIETRAJ,[36,39] a collection of scripts (http://bri.nrc.ca/ccb/pub) for calculating SIE free energies from AMBER-generated trajectories. We analyzed the last 23 ns of the simulations using snapshots at 40 ps intervals. We used the PTRAJ module of AMBER to remove water and counterions from the trajectories prior to the SIE calculation.

RESULTS AND DISCUSSION

RNA Complexes of L1 and L2. *Stability of MD-Simulated Complexes.* The root-mean-square deviation (rmsd) of $L1^{NT}$ and $L2^{NT}$ Cα atoms from their initial structures postequilibration reaches stable values of ∼1.8 and ∼3.1 Å after ∼16 ns for L1 and L2, respectively (Figure 2A). The larger deviation observed from L2 reflects the fact that the starting point for the simulation was a modeled structure of the protein rather than a crystal structure as in the case of L1. Compared to the protein atoms, there is more deviation in the RNA strands from their initial structures, 3.9 and 4.0 Å for L1 and L2, respectively (Figure 2B). Part of this is because the initial dsRNA structure was a model, just like in the case of L2 (deviations for the RNA strands and L2 are comparable). Also contributing is the fact that the ends of the strands extending from the protein are quite mobile. If we exclude three nucleotides from each end of the strand from the rmsd calculation, we obtain significantly smaller rmsds of 2.4 Å for bound dsRNA (Figure 2C). For the core groups involved in the reaction [the C(−1) and nicked nucleotide, and Mg^{2+}], rmsds from initial structures are 1.2 and 1.0 Å for the L1 and L2 complexes, respectively. These data indicate that the MD simulations reached steady states characterized by well-equilibrated ligase−RNA solvated systems.

Role of Mg^{2+} Ion. Figure 3 shows how the C(−1) and nick nucleotides and four water molecule are predicted to coordinate to the Mg^{2+} ion in the simulated L1 complex. This coordination pattern is maintained throughout the 40 ns MD simulation. The Mg^{2+} is coordinated to the C(−1) O3′ hydroxyl and the nick α-phosphoryl groups. The average distances of Mg^{2+} to the C(−1) O3′ and nick O2P atoms are 2.1 and 1.9 Å, respectively. This coordination locks the relative orientation of the C(−1) O3′ hydroxyl and phosphoryl groups to be ligated. The average distance between the nucleophilic C(−1) O3′ and its target phosphorus atom is 3.6 Å. The average angle formed by the C(−1) O3′, P, and the ester oxygen in the leaving AMP is 144.9°. Hence, the O3′ atom is well positioned for a nucleophilic attack. Moreover, coordination of O3′ may also reduce the pK_a of this hydroxyl group, facilitating the formation of an oxyanion, rendering it an even more effective nucleophile. Hence, the proposed mechanism implicates the nucleophilic O3′ atom already in its activated oxyanion state (Figure 3B). This is reminiscent of the proposed role of Mg^{2+} in the splicing of introns by metalloenzymes.[40]

This mechanism is different from that proposed by Swift and co-workers.[17] In that study, the Mg^{2+} ion was more loosely bound to the complex, being surrounded by five water molecules bridging the ion to the protein−RNA complex, with only one direct contact observed between Mg^{2+} and the β-phosphate group of the nick. Such a less engaged Mg^{2+} ion appears to be due to a different binding mode of the 3′-end of the dsRNA showing no contact with the protein, thus allowing the solvent to penetrate between the dsRNA and L1 and hydrate the Mg^{2+} ion. In our MD models of L1−dsRNA and L2−dsRNA complexes, there is an intimate contact of the 3′-end of dsRNA with the enzyme, resembling more closely the binding mode seen in the crystal structure of the T4-Rnl2−substrate complex.[9] For example, the H-bond established between the PO_4 group of the 3′-terminal nucleotide (−1) and the backbone NH amide of Thr39 in the T4-Rnl2−substrate crystal structure is mimicked in our models of the complexes by hydrogen bonds to the Asn92 NH amide and/or the Thr91 side chain of L1. The corresponding interaction in L2 is with the Asn46 amide. In contrast, the previous model[17] deviates significantly from this binding mode, with the distance to the Thr91 side chain hydroxyl hydrogen being >8 Å. Clearly, our models imply a role for the Mg^{2+} ion in the RNA ligation mechanism more active than that of Swift et al.

The organization of Mg^{2+}, C(−1), and the nick is essentially mirrored in the $L2^{NT}$−dsRNA complex. The average distances of Mg^{2+} to C(−1) O3′ and O2P of the nick are 2.1 and 1.9 Å, respectively. The average distance between the nucleophilic C(−1) O3′ and its target phosphorus atom is 3.6 Å. The average angle formed by the C(−1) O3′, P, and the ester oxygen in the leaving AMP is 143.5°. Hence, we expect L1 and L2 to share the same RNA ligation mechanism.

The coordinated water molecules do not appear to have a direct role in the nucleophilic attack. However, they do stabilize the Mg^{2+} position by providing bridging interactions with surrounding protein residues that include the side chains of Glu159 and Glu283 in L1. The corresponding amino acids in L2 are Glu126 and Glu246.

Three replicate MD simulations using different random number seeds were run for both $L1^{NT}$−dsRNA and $L2^{NT}$−dsRNA complexes. The configuration of Mg^{2+}, C(−1), and the nick is maintained across replicates. Time series for Mg^{2+} coordination by the RNA ends to be ligated (two distances) and nucleophilic attack (a distance and an angle) are shown in Figure S1 from all three 40 ns MD replicates for the two enzymes, with the corresponding average and standard deviation values listed in Table 1. It is noteworthy to point out that these mean geometry values differ significantly, and by as much as 2.8 Å and 51°, relative to those of the initial geometries after energy minimization of complexes assembled on the basis of available crystallographic data. This underscores the instrumental role

Table 1. Sample Averages and Standard Deviations (SDs) for Key Geometry Parameters Defining the Mg^{2+} Coordination by the RNA Ends To Be Ligated and the Nucleophilic Attack for Ligation[a]

complex	parameter	average ± SD
$L1^{NT}$−dsRNA	distance (Mg−$O^{3'\text{-OH}}$)	2.14 ± 0.08 Å
	distance (Mg−$O^{5'\text{-PO}_4}$)	1.95 ± 0.05 Å
	distance ($O^{3'\text{-OH}}$−$P^{5'\text{-PO}_4}$)	3.43 ± 0.26 Å
	angle[b] ($O^{3'\text{-OH}}$−$P^{5'\text{-PO}_4}$−$O^{\text{leaving-AMP}}$)	154.90 ± 10.77°
$L2^{NT}$−dsRNA	distance (Mg−$O^{3'\text{-OH}}$)	2.13 ± 0.07 Å
	distance (Mg−$O^{5'\text{-PO}_4}$)	1.95 ± 0.04 Å
	distance ($O^{3'\text{-OH}}$−$P^{5'\text{-PO}_4}$)	3.49 ± 0.16 Å
	angle[b] ($O^{3'\text{-OH}}$−$P^{5'\text{-PO}_4}$−$O^{\text{leaving-AMP}}$)	150.60 ± 10.51°

[a]Data from three MD replicates each of 40 ns for $L1^{NT}$−dsRNA and $L2^{NT}$−dsRNA complexes. See also Figure 3B for a schematic of geometric parameters. [b]Angle corresponds to the atoms indicated by red arrows in Figure 3B.

DOI: 10.1021/acs.biochem.5b01257
Biochemistry 2016, 55, 2319−2331

played by the MD simulations in refining the geometry of RNA ligation reaction catalyzed by these enzymes.

Protein–RNA Contacts. The binding site for the nick is surrounded with basic amino acids. In particular, Lys87 and Lys307 of L1 help stabilize the binding of the α- and β-phosphoryl groups of the nick and maintain hydrogen bonding interactions throughout the simulation (Figure 4A). The corresponding residues in L2 are Lys57 and Lys269, and they form similar interactions with the nicked nucleotide (Figure 4B).

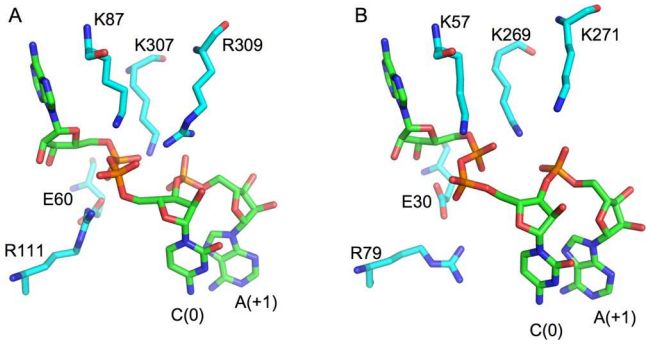


Figure 4. Interactions of the nicked nucleotide with basic residues around the ligase active site in the (A) L1–RNA complex and (B) L2–RNA complex. Ligase residues are shown with C atoms colored cyan and RNA nucleotides with C atoms colored green.

Along with the proposed mechanism shared by L1 and L2, our MD simulations predict quite similar interactions between the protein and RNA around the ligation site for these two enzymes (Figure 4). Approaching the α- and β-phosphoryl groups from an almost diametrically opposite direction relative to the lysine side chains mentioned above is Arg79 of L2. It maintains a hydrogen bond with the pyrophosphoryl group for most of the simulation, while also being in the proximity of the Glu30 side chain. The corresponding Arg111 side chain of L1 remains hydrogen-bonded to the corresponding Glu60 carboxylate for most of the simulation, but it also approaches the pyrophosphate of the nick. The side chain of Arg309 of L1 is mobile but spends part of the simulation with a hydrogen bond to the phosphate of A(+1) or the O2′ hydroxyl of the nick, while the shorter side chain of the corresponding Lys271 of L2 is mostly hydrogen bonded to the phosphate of A(+1).

There are also a number of interactions established by $L1^{NT}$ and $L2^{NT}$ with the dsRNA substrate, which are remote from the ligation site. Some of these interactions differ between the two homologous ligases. Table 2 summarizes all residue–nucleotide contacts with >80% of contact frequency for the last 10 ns of simulations. These lists of residue–nucleotide contacts include the catalytic residues that are conserved in both ligases as well as different residues in other regions of these ligase catalytic domains. Structural details for most contacts are shown in the Supporting Information in Figures S2–S8 for the L1–RNA complex and in Figures S9–S12 for the L2–RNA complex.

The structures of the dsRNA complexed with $L1^{NT}$ and $L2^{NT}$ show a tilt of the helical axis of the bound nucleic acid of ~25° toward the gRNA 5′-end when compared with that of the nucleic acid (hybrid DNA–RNA) complexed to the T4-Rnl2 template structure.[9] This shift allows the ends of the model dsRNA substrate to interact intimately with the L1 and L2 catalytic domains, via structural elements that are different from those in T4-Rnl2 (Figure S13).

Table 2. List of the Protein–RNA Interactions in the Simulated $L1^{NT}$–dsRNA and $L2^{NT}$–dsRNA Complexes[a]

$L1^{NT}$–dsRNA complex		$L2^{NT}$–dsRNA complex	
residue	nucleotide	residue	nucleotide
Tyr58	AB	Arg27	gU(+4)
Ser67	gU(+9)	Ile31	AB
Arg68	gC(+7), gC(+8)	Arg38	gC(+6), gG(+7)
Ser71	gC(+8)	Ala41	gC(+7)
Lys87	AB	Thr55	AB
Val88	AB	Glu56	AB
Thr91	C(−1), C(−2)	Lys57	AB
Asn92	C(−1)	Val58	AB
Lys110	C(−3)	Gly60	AB
Arg111	AB, C(−1), C(−2)	Ala61	C(−2)
Ser112	C(−2), C(−3)	Asn62	C(−1)
Phe121	C(−2), C(−3)	Lys78	C(−3)
Phe122	gG(−2), C(−3)	Arg79	AB, C(−2)
Glu159	AB, C(−1)	Ser80	C(−3)
Gln193	C(−1), gG(−2)	Phe89	C(−2), C(−3)
Arg194	gG(−1), gG(−2)	Phe90	C(−2), gG(−4), gG(−4)
Phe207	AB	Glu126	AB, C(−1)
Phe209	AB	Gln157	C(−1)
Glu283	AB	Thr158	gG(−2), gG(−3)
Val286	AB	Tyr171	AB
Asn301	gC(+7)	Phe173	AB
Ser303	gC(+7)	Thr266	gG(+6)
Lys307	AB, A(+1)	Lys269	AB, A(+1)
Arg309	AB, A(+1)	Lys271	AB, A(+1)

[a]AB, adenylated base, C(0).

Substrate Binding Affinity and Specificity Calculations. Cruz-Reyes and colleagues[5] have reported that L1 and L2 differ in their affinities for ATP and have shown that, in step 1 chemistry, L2 has a higher affinity for ATP. To explore whether L1 and L2 preserve their differential affinities for their substrates in step 3 catalysis, we examined their binding free energies for the nicked dsRNA substrates using SIE calculation.[39] For these calculations, we considered Mg^{2+} and its four coordinated water molecules as part of the protein and the dsRNA as the ligand. We calculated the SIE averages from the last 20 ns trajectory of each complex (Table S1). The average SIE values were different for the two structures. The calculated average binding free energies are −17.54 ± 2.48 and −20.15 ± 1.02 kcal/mol for $L1^{NT}$ and $L2^{NT}$, respectively. The more favorable interaction of the dsRNA model substrate for L2 may be partly due to different electrostatic properties of the catalytic domains of these ligases,[8] including a net charge of 9 for L2 versus −5 for L1.

Palazzo and colleagues[14] have studied the preference of recombinant L1 and L2 for the terminal nucleotides of the upstream RNA fragment and have shown a 9-fold increase in ligation activity for L1 when the terminal residue of the upstream RNA fragment is changed from cytosine to guanine. Our dsRNA template contained cytosine as the terminal nucleotide. Consequently, we have explored computationally whether the C → G substitution at the terminal position of the upstream RNA fragment would affect binding affinity. We conducted 40 ns MD simulations for the C → G variants for each of the $L1^{NT}$–dsRNA and $L2^{NT}$–dsRNA complexes. We report the average SIE values for the $L1^{NT}$–dsRNA complex (C → G variant) as −23.54 ± 0.96 kcal/mol and for the $L2^{NT}$–dsRNA complex as −24.49 ± 1.16 kcal/mol. These results predict a 6.0 kcal/mol strengthening of binding affinity for the L1–RNA complex and 4.3 kcal/mol for

DOI: 10.1021/acs.biochem.5b01257
Biochemistry 2016, 55, 2319–2331

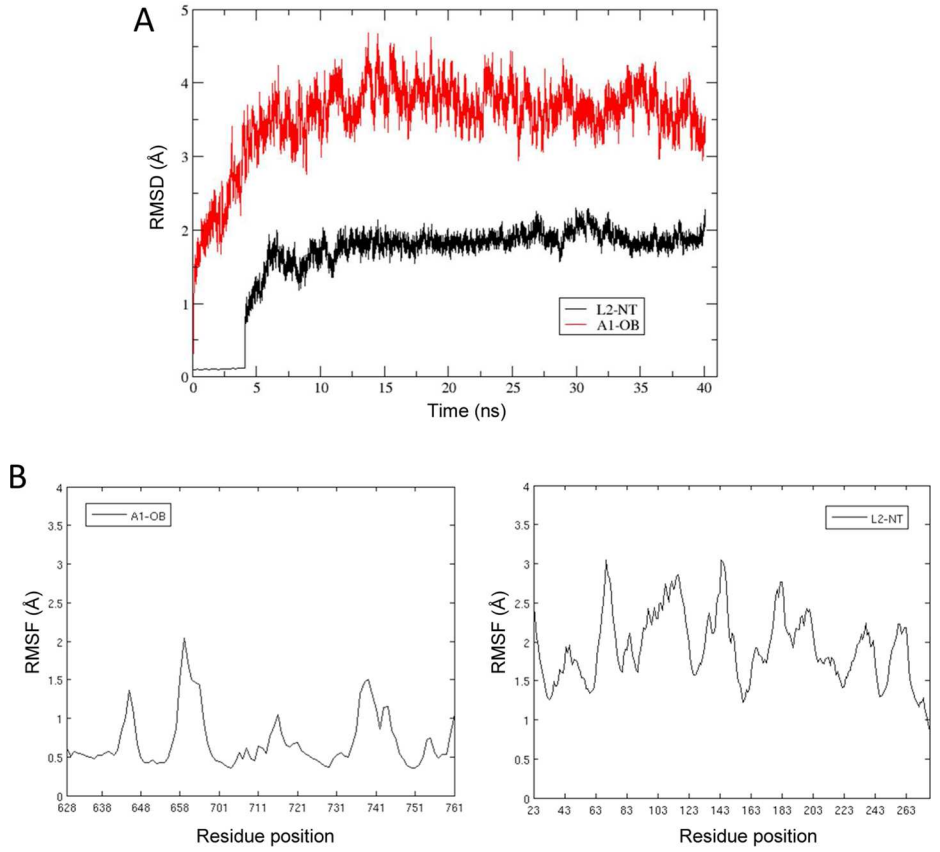


Figure 5. Structural dynamics of the $L2^{NT}$–dsRNA–$A1^{OB}$ complex. (A) rms deviations of backbone atoms of bound $A1^{OB}$ and $L1^{NT}$ domains in the complex with respect to the initial structure during the course of MD simulation. (B) Per-residue rms fluctuations of Cα atoms of $A1^{OB}$ and $L2^{NT}$ around the average structure calculated over the 40th (last) ns of the MD trajectory.

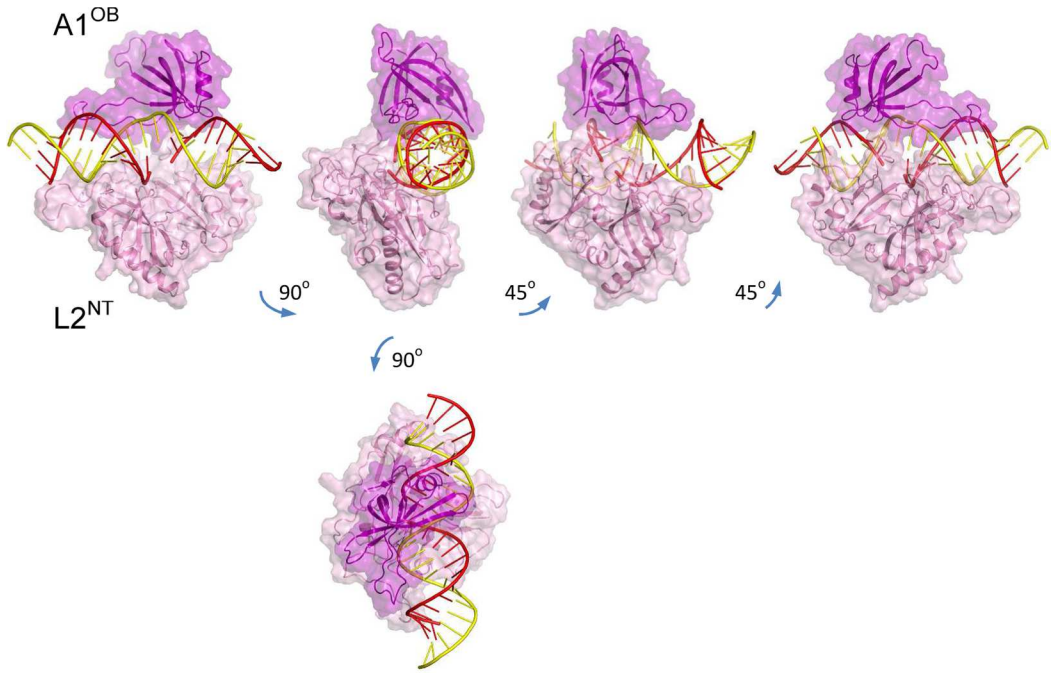


Figure 6. Overall structure of the modeled $L2^{NT}$–dsRNA–$A1^{OB}$ complex. The energy-mimimized average structure from the last nanosecond of the 40 ns MD simulation is shown from various angles as indicated by rotation arrows. The $A1^{OB}$ domain is colored purple, and the $L2^{NT}$ domain is colored pink. The dsRNA is shown in cartoon representation with the mRNA strand colored red and the gRNA strand colored yellow. The mRNA strand is interrupted at the nicked nucleotide intentionally left not rendered to highlight the location of the ligation site.

DOI: 10.1021/acs.biochem.5b01257
Biochemistry 2016, 55, 2319–2331

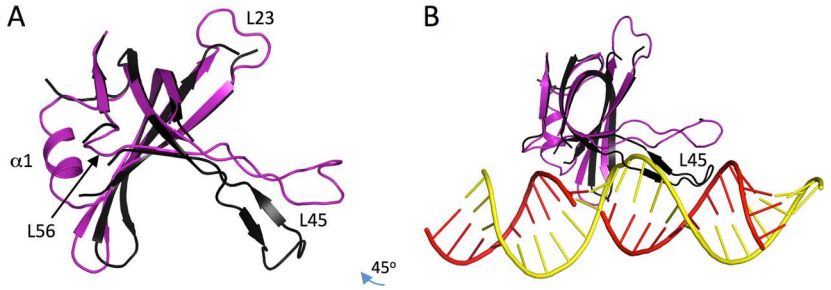


Figure 7. Conformational change in the $A1^{OB}$ domain predicted to occur upon binding to dsRNA. (A) Comparison of the bound and free states. (B) Steric clash of the L45 loop in the free-state conformation with the dsRNA in the modeled docking pose. The crystal structure of the free state (PDB entry 4DK6, chain C) is colored black, and the modeled structure of the bound form is colored purple; dsRNA is colored red and yellow for the mRNA and gRNA strands, respectively.

the L2−RNA complex when the C → G variant is used. Structural changes incurred in the C → G variants of the L1−RNA and L2−RNA complexes are shown in Figures S14 and S15, respectively. Our observation is in qualitative agreement with the improvement seen by Palazzo and co-workers, although our calculated values overestimate the absolute magnitude of the improvement. In their experiment, the C → G substitution in recombinant L1 increased the ligation activity by 9 times, and in L2 by 4 times.

Interaction with Binding Partner Protein. The previous section described the interactions predicted for the step 3 chemistry mechanism of RNA ligation for L1 and L2 based on MD simulations of their complexes with RNA substrates. OB-fold domains of the largest editosome interacting proteins KREPA1 and KREPA2 binding proteins (termed $A1^{OB}$ and $A2^{OB}$ domains, respectively) play an important role by remodeling the catalytic sites of L2 and L1, respectively. Park and colleagues have recently provided experimental evidence of the interaction of $L2^{NT}$ and $A1^{OB}$ domains with dsRNA in a ternary complex.[18] We further explored this complex here by computational simulations at atomic resolution to gain additional insights into the molecular mechanisms that modulate RNA editing.

Overall Structure Predicted for the $L2^{NT}$−dsRNA−$A1^{OB}$ Complex. The structural model of the $L2^{NT}$−dsRNA−$A1^{OB}$ complex represents an *in-trans* interaction of an OB-folded domain with an oligonucleotide-bound ligase, which has not been available from experimental structure determination methods thus far. To arrive at this model, an initial placement of the $A1^{OB}$ domain onto the $L2^{NT}$−dsRNA complex was informed by the structure of the DNA-bound human DNA ligase I (LIG1) that presents the OB domain *in-cis*. The modeled structure was refined by all-atom MD simulation in explicit solvent and at physiological salt concentration. A stable structure was obtained after simulation for ∼5 ns, characterized by rmsd values from the initial placement of the $A1^{OB}$ and $L2^{NT}$ domains that fluctuate around mean values of 3.5 and 2 Å, respectively (Figure 5A). To highlight regions of higher flexibility, rms fluctuations (rmsf) per residue are shown in Figure 5B. We see that most regions of the $A1^{OB}$ domain typically fluctuate by 0.5 Å, with only the tips of loops L23 (residues 657−662, the location of a long insertion loop shown not to be functionally important) and L45 (residues 737−740) reaching fluctuations higher than 1.5 Å and at most 2 Å. The $L2^{NT}$ domain fluctuates mainly in the 1−3 Å range around the average structure. These data indicate an equilibrated and relatively stable dynamical model of the complex.

The overall structure of the $L2^{NT}$−dsRNA−$A1^{OB}$ complex, represented by the energy-minimized average over the last nanosecond of MD simulation, is shown from various angles in Figure 6. In this model, the $A1^{OB}$ domain interacts with dsRNA only and makes no direct contact with $L2^{NT}$ amino acids. We can see that the $A1^{OB}$ domain binds mostly into the minor groove of the dsRNA. The location of the $A1^{OB}$ is such that it contacts mainly the template gRNA strand right across the nicked nucleotide, in a region spanning nucleotides −3 to +3 relative to the nucleotide complementary to the nicked nucleotide. Indeed, the template strand spans the entire concave length of the β-barrel from L12 all the way to L45. In addition, important contacts are also made with the mRNA strand directly at the level of the nicked nucleotide. On the $A1^{OB}$ side, the interface with dsRNA corresponds to the concave face of the β-sheet of the OB fold, with direct contributions from strands β3 and β5, as well as extensive implications for several loops, notably the long loop L45, the shorter loop L12, and to a lesser extent the L3α1 and L56 loops.

It is important to compare the bound and free conformations of the $A1^{OB}$ domain. Apart from a few regions modeled here that are not present in the crystal structure (loops L23 and L56, and helix α1), a large conformational change in loop L45 is predicted between the free and bound structures (Figure 7A). In the free state, the L45 loop adopts a conformation that increases the concavity of the β-sheet, whereas in the bound state, the loop direction relative to the β-sheet is changed at its anchors, resulting in a shallower concavity of the domain. This is important because in the "free-state" conformation, the L45 loop would collide with the dsRNA structure in this binding mode (Figure 7B). We note that the current knowledge-based docking mode was based on the structure of LIG1 as a template. The OB-folded domain in the latter structure[10] (PDB entry 1X9N) has a similarly long L45 loop that is also oriented like the bound conformation in the present $A1^{OB}$ domain model. It is important to note that the dsRNA-bound conformation of the L45 loop of $A1^{OB}$ was not modeled on the basis of the DNA-bound conformation of the L45 loop of LIG1. Rather, it was generated as the lowest-energy conformation based on conformational sampling during which complete reconstruction of the loop backbone was sampled via a loop closure approach (see Materials and Methods). The conformational sampling was prompted by the length of the loop and the possibility of crystal packing artifacts in the crystal structure[18] (PDB entry 4DK6). Hence, it appears that according to the structural model presented here, an induced-fit mechanism would take place during the interaction of the $A1^{OB}$ domain with the L2-bound dsRNA, which implicates a conformational change in the L45 loop of $A1^{OB}$ (residues 729−747).

DOI: 10.1021/acs.biochem.5b01257
Biochemistry 2016, 55, 2319−2331

Specific Interactions Predicted at the dsRNA–A1^{OB} Interface. As mentioned earlier, the A1^{OB} domain is predicted to establish a number of specific interactions with the L2-bound dsRNA strands. These interactions are listed in Table 3 and include both hydrogen-bonded and nonpolar interactions.

Table 3. Detailed Interactions between the A1^{OB} Domain and Nicked dsRNA Bound to L2^{NT}

A1^{OB} structural element	A1^{OB} residue	type of interaction	location on dsRNA[a]
L12 loop	mc L643	packing	mc gG(+2)
	mc G644	packing	mc gA(+3)
	mc A645	packing	mc gA(+3)
	sc M646	packing	minor groove gA(+3)
β2 strand	sc V648	packing	minor groove gU(+1)
β3 strand	sc R703	H-bonding	mc AB
L3α1 loop	sc M705	packing	minor groove gU(+1)
	mc G706	packing	mc A(+4)
β4 strand	sc T729	packing	mc C(+2)
L45 loop	sc R731	H-bonding	sc AB
	mc M732	H-bonding	mc gG(−1)
	sc M732	packing	minor groove gG(−2)
	mc/sc R734	H-bonding	mc gG(−2),gG(−3)
	mc H735	H-bonding	mc gG(−3)
	sc R742	H-bonding	mc gG(−3)
β5 strand	sc F748	packing	mc gG(−1)
	sc Q750	H-bonding	sc gG(0)
L56 loop	sc V752	packing	minor groove U(+3)
	sc P754	packing	minor groove U(+3)

[a]Abbreviations: mc, main chain; sc, side chain; AB, adenylated base, C(0).

Perhaps the most important of those are the interactions established by two arginines around the nicked nucleotide. As shown in Figure 8A, the side chain of Arg731 from the N-terminal end of the L45 loop interacts by direct hydrogen bonding with the 2′-OH and the cytosine carbonyl group of the nicked nucleotide. On the other hand, the side chain of Arg703 from the β3 strand interacts also by H-bonds with the main chain 5′-phosphodiester group of the nucleotide complementary to the nicked nucleotide. Hence, according to this model, it is possible that both of these Arg residues play important structural roles in the catalysis to affect favorably the inherent basal ligation activity of the L2^{NT} domain.

In addition, there are other interactions established by A1^{OB} domain residues with dsRNA nucleotides flanking the nicked position (Table 3). Hydrogen bond interactions implicate mainly the main chain phosphodiester groups of the template gRNA strand, which are engaged by several A1^{OB} residues. Thus, side chain guanidinium protons from the L45 loop residues Arg734 and Arg742, as well as main chain amide protons from Arg734 and His735, bind to phosphate groups at template backbone positions gG(−2) and gG(−3). Additional H-bonding interactions are established between the main chain carbonyl of Met732 and the 2′-OH group at position gG(−1), and between the side chain of Gln750 from the β5 strand and the 2′-OH and the base of the nucleotide complementary to the nick. Moreover, a number of nonpolar residues are located in the minor groove and interact with nucleotides from both the broken and template strands. These include the side chains of Met646, Val752, and Pro754 (modeled in a *cis* conformation) at the level of U(+3) and gA(+3) nucleotides, Val648 and Met705 with the gU(+1) nucleotide, and Met732 interacting with the gG(−2) nucleotide (Figure 8B). Other packing interactions involve the side chain of Phe748 from the β5 strand against the gRNA main chain at position gG(−1), L12 loop residues Leu643, Gly644, and Ala645 against the gRNA main chain at positions gG(+2) and gA(+3), and Gly706 and Thr729 against the broken strand main chain at positions A(+4) and C(+2).

Some of these interacting residues are conserved in the OB-folded domain of the homologous A2 protein (see the sequence alignment in Figure S16). However, it should be noted that conservation is not a requirement for the validity of an interaction site, given the tremendous sequence variability that can be accommodated by the OB fold and the diversity of specific interactions that can be established by OB-folded domains with nucleic acid duplexes while maintaining the same overall binding mode relative to the position of the ligation site.[19]

The L2 interactions around the nicked nucleotide, including the Mg^{2+}-mediated geometry of the RNA ends to be ligated, seem not to be affected by the presence of the A1^{OB} domain based on the simulated structures of the binary and ternary complexes. Therefore, this model suggests that the molecular mechanism by which the A1^{OB} domain enhances the inherent basal ligation activity of L2 may be to simply provide additional stabilization of the L2^{NT}–dsRNA complex due to the added affinity afforded by the extensive complementary interactions at the A1^{OB}–dsRNA interface. This may further stabilize the alignment of the RNA ends for the nucleophilic substitution in the step 3 chemistry of ligation described earlier. It is natural to interrogate whether the A1^{OB} interactions also lead to a realignment of the nicked nucleotide and the 3′-end of the broken strand for enhanced ligation activity. This has been repeatedly observed in the cases of DNA ligases that present an OB-folded domain *in-cis*, like the human DNA ligase I (LIG1), the *E. coli* DNA ligase LigA, or the *Chlorella* virus DNA ligase.[19] In each of these cases, the interactions are established by different residues and in different ways, but the net result seems to always consist of a local distortion around the nick. This includes a widened minor groove immediately upstream and a B-form conformation downstream of the nick, which positions or aligns the ends of the nick for ligation. Hence, this DNA duplex distortion resulting in an A- to B-form transition across the nick ultimately leads to an RNA-like structure that seems to favor ligation. This may explain why the RNA ligase studied here does have basal ligation activity even without assistance from an OB-folded domain; however, addition of the OB domain *in-trans* (as provided by A1) leads to enhanced dsRNA stabilization and improved ligation efficiency. Whether other structural factors in addition to the extra stabilization like a better alignment of the RNA ends to be ligated would be promoted by the A1^{OB} domain interactions is an intriguing hypothesis. Addressing this would require further structural studies as the magnitude of such conformational transitions is outside of the time scale of the MD simulation performed here.

Consistency of the Complex Model with Experimental Evidence. Park and colleagues have recently probed the A1^{OB} interacting surface with L2-bound dsRNA by site-directed mutagenesis.[18] Substituting arginine with glutamate independently at seven solvent-exposed positions on the surface of the A1^{OB} domain, they found that three residues, Arg703, Arg731, and Arg734, were essential for dsRNA binding. However, the mechanism by which those three arginine residues interact with dsRNA was unclear. Our simulation revealed why those residues

DOI: 10.1021/acs.biochem.5b01257
Biochemistry 2016, 55, 2319–2331

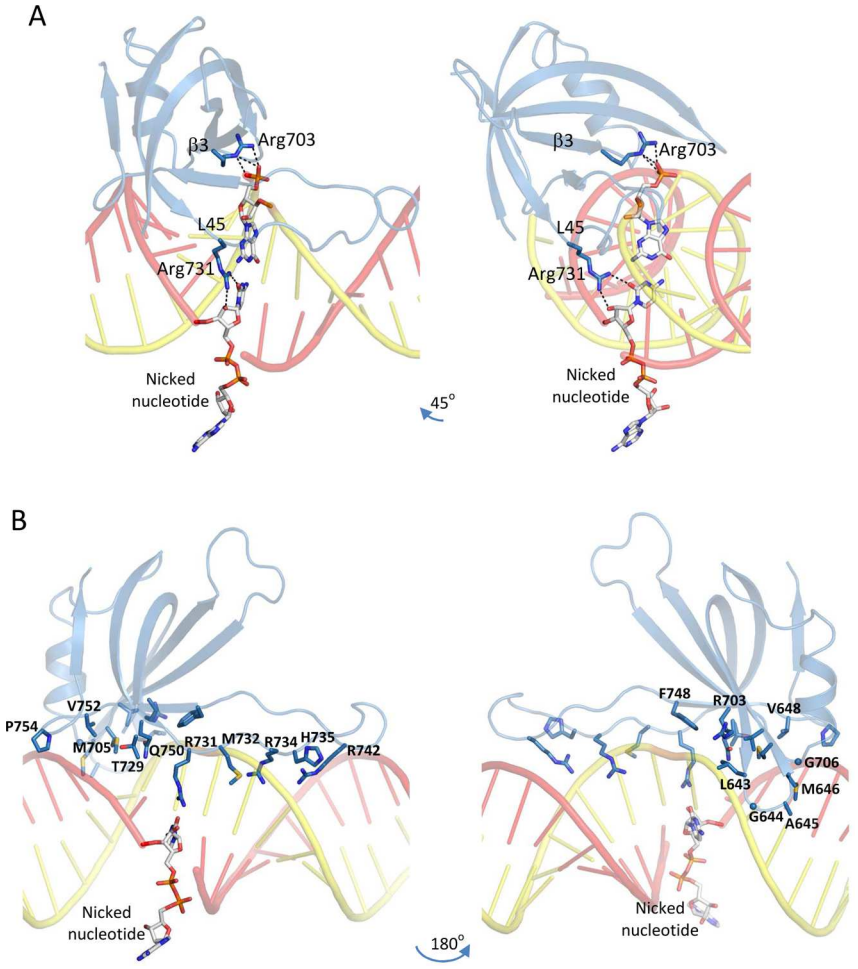


Figure 8. Detailed interactions established by the A1^{OB} domain with dsRNA. (A) Interactions of the A1^{OB} domain (purple rendering) with the nicked nucleotide on the broken RNA strand (red rendering) and with its complementary nucleotide from the template RNA strand (yellow rendering). Interacting Arg residues from the A1^{OB} domain are shown as sticks with purple C atoms, and interacting nucleotides are shown as sticks with white C atoms. Hydrogen bonds are indicated by black dashed lines. (B) Structural overview of all A1^{OB} domain residues predicted to establish direct contacts with nicked dsRNA. Interacting partners or locations on the dsRNA for each of the shown A1^{OB} residues are listed in Table 3.

might be important. In Figure 9, we identify on our structural model of the complex all seven basic residues that were subjected to charge-reversal mutagenesis. As discussed earlier, Arg703 and Arg731 make direct hydrogen bond contacts with the nicked nucleotide and its complementary nucleotide, respectively (Figure 8A), while Arg734 is predicted to hydrogen bond to the backbone phosphodiester at positions gG(−2) and gG(−3) on the template strand (Figure 8B and Table 3). These interactions cannot be sustained upon Arg-to-Glu mutations, and eventually, these mutations cause unstable binding of the OB fold to the dsRNA. We also see that three of the four basic residues that were found not to be critical to complex formation, Lys715, Lys719, and Lys741, point away from dsRNA and into the solvent in our model of the complex. Hence, the only apparent disagreement between this structural model and experimental mutagenesis data of Park and colleagues is at Arg742, which tolerated glutamate mutation, whereas in our model, it is H-bonded to the backbone phosphodiester at template strand position gG(−3). However, in contrast to the functionally important Arg734 that is predicted to H-bond at the same location, Arg742 is positioned further toward the more flexible tip of the L45 loop (see also Figure 8B) where its side chain has ample conformational freedom. Hence, the R742E mutant could easily circumvent the electrostatic repulsion with

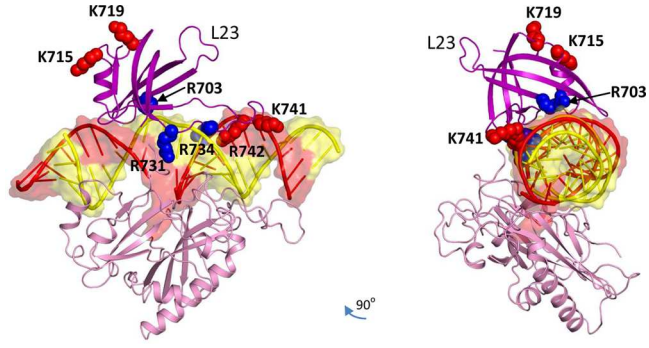


Figure 9. Consistency of the structural model with experimental mutagenesis data.[18] Seven A1^{OB} domain basic residues independently mutated to glutamate are shown in sphere representation and labeled. Their colors encode the functional outcome of these mutations: blue indicates loss of binding, while red represents retained binding activity to L2-dsRNA. Color rendering of the complex is as in Figure 6. The relative location of the L23 loop, where a long insertion present in the A1^{OB} sequence has been shown not to contribute to L2−dsRNA binding, is also labeled.

DOI: 10.1021/acs.biochem.5b01257
Biochemistry 2016, 55, 2319−2331

the dsRNA main chain introduced upon mutation, something that could not be readily accomplished in the R734E mutant because of the steric constraint at this position.

There is additional experimental evidence that supports the present structural model. First, the $A1^{OB}$ domain sequence includes a long insertion (38 residues) in the L23 loop relative to the homologous $A2^{OB}$ (Figure S16), and such an insertion is atypical in the structural family of OB-folded domains. This likely flexible insertion was substituted with a four-residue linker to determine the $A1^{OB}$ crystal structure,[18] and consequently, this long insertion was also not included in the model presented here. Importantly, it was shown that deletion of the L23 insertion has no functional implications, including binding of A1 to the $L2^{NT}$–dsRNA complex.[18] Consistent with this finding, the L23 loop of $A1^{OB}$ points away from the L2-bound dsRNA in our structural model of the complex (Figure 9). Second, the crystal structure of $A1^{OB}$ was determined in the presence of a nanobody that facilitated crystallization. This raised an interesting question of whether the $A1^{OB}$ interface utilized for nanobody binding overlaps with the interface required for $L2^{NT}$–dsRNA binding. It was found that nanobody recruitment did not alter significantly the dsRNA binding affinity for the A1–L2 complex, hence suggesting that the interface used to bind the nanobody is not critical for dsRNA binding; i.e., distinct surfaces on the $A1^{OB}$ domain are used for nanobody and RNA binding.[18] In agreement with this experimental evidence, our structural model is compatible with nanobody binding to the $L2^{NT}$–dsRNA–$A1^{OB}$ complex (Figure 10).

CONCLUSION

We modeled the *T. brucei* RNA ligases L1 and L2 in complex with RNA substrate using the available crystal structure of T4-RNA ligase 2 (T4 Rnl2) in complex with a DNA–RNA hybrid substrate as a template. In our models, we retained the Mg^{2+} ion and its four coordinated water molecules observed in the crystal structure of ATP-bound L1. The Mg^{2+} ion is also found to coordinate an oxygen atom of the 5′-PO_4 of the nick and the 3′-OH of the C(−1) nucleotide to be ligated. In this configuration, the 3′-OH of the terminal nucleotide is poised for nucleophilic attack of the 5′-PO_4 of the nick, consistent with the putative step 3 chemistry of the ligation mechanism. Moreover, coordination of the 3′-OH to the Mg^{2+} ion may lower its pK_a, thereby rendering it a more effective nucleophile as an oxyanion. In our model, Mg^{2+} plays a twofold role: bringing the reactants into the proximity of each other and activating the nucleophile. In the absence of crystal structures, our simulations yielded plausible working models for the L1 and L2 complexes with RNA and provided insights into the possible role of Mg^{2+} in the ligation reaction.

The RNA ligase activities of these enzymes are sensitive to the nature of the terminal nucleotide at the 3′-end of the upstream RNA fragment. We explored whether binding free energies extracted from the molecular dynamics simulations are able to capture such a nucleotide preference. We ran MD simulations for C → G substitutions at the terminal position of the 3′-end of the upstream RNA fragment and used the trajectories to calculate relative binding affinities. We observed enhancements in the predicted binding affinities characterizing the Michaelis complexes of the step 3 ligation catalysis for both L1 and L2 ligases. This parallels experimentally observed enhanced ligation activities arising from the C → G substitution for these enzymes and suggests that binding affinity in step 3 is an important contributor to catalytic efficiency.

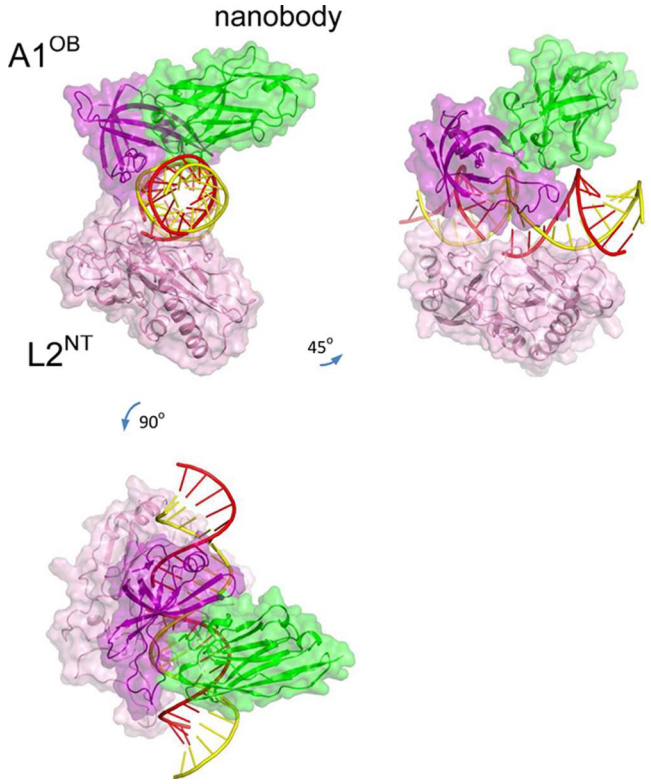


Figure 10. Consistency of the structural model with experimental evidence of a ternary complex with the anti-A1 antibody.[18] The nanobody structure is colored green; all other rendering is as in Figure 6. The complex was generated from a superposition of the crystal structure of the $A1^{OB}$–nanobody complex (PDB entry 4DK6) onto the MD model of the $L2^{NT}$–dsRNA–$A1^{OB}$ complex, by using the β-barrel Cα atoms of the $A1^{OB}$ domain from each complex for an rms fit.

Furthermore, this study predicts structural details at the atomic level that are responsible for the observed increase in binding affinity and stabilization of the L2–dsRNA complex by the OB-folded domain of the largest editosome binding protein A1. A structural model of the $L2^{NT}$–dsRNA–$A1^{OB}$ complex is described in terms of $A1^{OB}$ residues engaging in direct interactions with dsRNA nucleotides. Two Arg residues are predicted to make direct H-bond contacts with the nicked nucleotide and its complementary nucleotide from the template strand. Other interactions in the minor groove and with the template strand around the ligation site are also described, which offer a structural basis for further experimental studies aimed at better defining this interface. The model of the complex is shown to be consistent with experimental evidence recorded to date on the $A1^{OB}$ surface implicated in this complex.

Whether the $A1^{OB}$ domain binding has other structural implications beyond providing additional stabilization free energy to the dsRNA complex, e.g., to nucleic acid duplex distortion and a better alignment of the mRNA ends to be ligated, is tempting speculation that requires further structural studies. This model can also be regarded as a steppingstone toward a more complete understanding of the molecular assembly between A1 and L2. Of interest particular to this system will be computational structural biology studies directed toward the other contact point between L2 and A1, i.e., between the C-terminal domain of L2 ($L2^{CT}$) and the first Zn-finger domain of A1 ($A1^{ZF1}$) as well as toward elucidating exquisite partnering specificity between L1–A2 and L2–A1 homologous

DOI: 10.1021/acs.biochem.5b01257
Biochemistry 2016, 55, 2319–2331

complexes. On a more general understanding of molecular architectures characteristic of DNA versus RNA ligation, it will be interesting to probe a tentative direct assembly between the current $L2^{NT}$–dsRNA–$A1^{OB}$ model and a future $L2^{CT}$–$A1^{ZF1}$ model. This assembly may be used to interrogate whether the various domains provided *in-trans* by A1 and L2 may clamp around the RNA duplex, as molecular clamps have recently become apparent as the typical *in-cis* architecture for most DNA ligases from various organisms spanning vast evolutionary times.[19]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.5b01257.

Supplemental figures showing MD time series for Mg coordination geometry (Figure S1), additional contacts in the modeled protein–RNA complexes (Figures S2–S12), overall geometry of the simulated complex (Figure S13), structural implications of RNA substrate mutation (Figures S14 and S15), and a sequence alignment of partner protein OB-folded domains (Figure S16) and a table of calculated ligase–RNA binding free energies from replicate MD trajectories (Table S1) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*NRC Canada, 6100 Royalmount Ave., Montreal, QC H4P 2R2, Canada. E-mail: traian.sulea@cnrc-nrc.gc.ca. Phone: 514-496-1924. Fax: 514-496-5143.

Funding

This research is supported by National Sciences and Engineering Research Council of Canada (NSERC) Grant 328186 to R.S. This is the National Research Council of Canada (NRC) publication number 53312.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the members of Salavati's and Purisima's laboratories for critical reading and suggestions. We thank Dr. Qizhi Cui for the nick design as well as reading the manuscript.

■ ADDITIONAL NOTE

[a]The dsRNA nucleotides are numbered relative to the nicked mRNA strand. The nucleotides upstream to the nicked mRNA were numbered −1, −2, −3, etc., and the nucleotides downstream from the nick were numbered +1, +2, +3, etc. The adenylated base is numbered 0. The complementary nucleotides in the gRNA strand take the same numbers and the prefix "g" added to the nucleotide single-letter abbreviation.

■ REFERENCES

(1) McManus, M. T., Shimamura, M., Grams, J., and Hajduk, S. L. (2001) Identification of candidate mitochondrial RNA editing ligases from Trypanosoma brucei. *RNA 7*, 167–175.

(2) Panigrahi, A. K., Gygi, S. P., Ernst, N. L., Igo, R. P., Palazzo, S. S., Schnaufer, A., Weston, D. S., Carmean, N., Salavati, R., Aebersold, R., and Stuart, K. D. (2001) Association of two novel proteins, TbMP52 and TbMP48, with the Trypanosoma brucei RNA editing complex. *Mol. Cell. Biol. 21*, 380–389.

(3) Schnaufer, A., Panigrahi, A. K., Panicucci, B., Igo, R. P., Wirtz, E., Salavati, R., and Stuart, K. (2001) An RNA ligase essential for RNA editing and survival of the bloodstream form of Trypanosoma brucei. *Science 291*, 2159–2162.

(4) Huang, C. E., Cruz-Reyes, J., Zhelotkina, A. G., O'Hearn, S., Wirtz, E., and Sollner-Webb, B. (2001) Roles for ligases in the RNA editing complex of Trypanosoma brucei: band IV is needed for U-deletion and RNA repair. *EMBO J. 20*, 4694–4704.

(5) Cruz-Reyes, J., Zhelonkina, A. G., Huang, C. E., and Sollner-Webb, B. (2002) Distinct functions of two RNA ligases in active Trypanosoma brucei RNA editing complexes. *Mol. Cell. Biol. 22*, 4652–4660.

(6) Schnaufer, A., Ernst, N. L., Palazzo, S. S., O'Rear, J., Salavati, R., and Stuart, K. (2003) Separate insertion and deletion subcomplexes of the Trypanosoma brucei RNA editing complex. *Mol. Cell 12*, 307–319.

(7) Deng, J., Schnaufer, A., Salavati, R., Stuart, K. D., and Hol, W. G. J. (2004) High resolution crystal structure of a key editosome enzyme from Trypanosoma brucei: RNA editing ligase 1. *J. Mol. Biol. 343*, 601–613.

(8) Shaneh, A., and Salavati, R. (2010) Kinetoplastid RNA editing ligases 1 and 2 exhibit different electrostatic properties. *J. Mol. Model. 16*, 61–76.

(9) Nandakumar, J., Shuman, S., and Lima, C. (2006) RNA ligase structures reveal the basis for RNA specificity and conformational changes that drive ligation forward. *Cell 127*, 71–84.

(10) Pascal, J. M., O'Brien, P. J., Tomkinson, A. E., and Ellenberger, T. (2004) Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature 432*, 473–478.

(11) Sekiguchi, J., and Shuman, S. (1997) Ligation of RNA-containing duplexes by vaccinia DNA ligase. *Biochemistry 36*, 9073–9079.

(12) Sriskanda, V., and Shuman, S. (1998) Specificity and fidelity of strand joining by Chlorella virus DNA ligase. *Nucleic Acids Res. 26*, 3536–3541.

(13) Nandakumar, J., and Shuman, S. (2004) How an RNA ligase discriminates RNA versus DNA damage. *Mol. Cell 16*, 211–221.

(14) Palazzo, S. S., Panigrahi, A. K., Igo, R. P., Salavati, R., and Stuart, K. (2003) Kinetoplastid RNA editing ligases: complex association, characterization, and substrate requirements. *Mol. Biochem. Parasitol. 127*, 161–167.

(15) MacKerell, A. D., and Nilsson, L. (2008) Molecular dynamics simulations of nucleic acid-protein complexes. *Curr. Opin. Struct. Biol. 18*, 194–199.

(16) Amaro, R. E., Swift, R. V., and McCammon, J. A. (2007) Functional and structural insights revealed by molecular dynamics simulations of an essential RNA editing ligase in Trypanosoma brucei. *PLoS Neglected Trop. Dis. 1*, e68.

(17) Swift, R. V., Durrant, J., Amaro, R. E., and McCammon, J. A. (2009) Toward understanding the conformational dynamics of RNA ligation. *Biochemistry 48*, 709–719.

(18) Park, Y. J., Budiarto, T., Wu, M., Pardon, E., Steyaert, J., and Hol, W. G. J. (2012) The structure of the C-terminal domain of the largest editosome interaction protein and its role in promoting RNA binding by RNA-editing ligase L2. *Nucleic Acids Res. 40*, 6966–6977.

(19) Pascal, J. M. (2008) DNA and RNA ligases: structural variations and shared mechanisms. *Curr. Opin. Struct. Biol. 18*, 96–105.

(20) Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., III, Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B. P., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossvry, I., Wong, K. F., Paesani, F., Vanicek, J., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.-J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2010) *AMBER 11*, University of California, San Francisco.

(21) Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Genet. 65*, 712–725.

(22) Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., III, Laughton, C. A., and Orozco, M. (2007) Refinement of the AMBER

DOI: 10.1021/acs.biochem.5b01257
Biochemistry 2016, 55, 2319–2331

force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J. 92*, 3817−3829.
(23) Meagher, K. L., Redman, L. T., and Carlson, H. A. (2003) Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem. 24*, 1016−1025.
(24) Oelschlaeger, P., Klahn, M., Beard, W. A., Wilson, S. H., and Warshel, A. (2007) Magnesium-cationic dummy atom molecules enhance representation of DNA polymerase beta in molecular dynamics simulations: improved accuracy in studies of structural features and mutational effects. *J. Mol. Biol. 366*, 687−701.
(25) Pang, Y. P., Xu, K., Yazal, J. E., and Prendergas, F. G. (2000) Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. *Protein Sci. 9*, 1857−1865.
(26) Pang, Y. P. (2001) Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method. *Proteins: Struct., Funct., Genet. 45*, 183−189.
(27) Mandell, D. J., Coutsias, E. A., and Kortemme, T. (2009) Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat. Methods 6*, 551−552.
(28) Das, R., and Baker, D. (2008) Macromolecular modeling with Rosetta. *Annu. Rev. Biochem. 77*, 363−382.
(29) Chaudhury, S., Lyskov, S., and Gray, J. J. (2010) PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics 26*, 689−691.
(30) Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis 18*, 2714−2723.
(31) Cornell, W., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc. 117*, 5179−5197.
(32) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926−935.
(33) Ryckaert, J. P., Ciccotti, G., and Berendsen, J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys. 23*, 327−341.
(34) Darden, T. A., York, D., and Pedersen, L. (1993) Particle mesh Ewald: an N log(N) method for Ewald sums in large systems. *J. Chem. Phys. 98*, 10089−10092.
(35) Leimkuhler, B., and Matthews, C. (2013) Robust and efficient configurational molecular sampling via Langevin dynamics. *J. Chem. Phys. 138*, 174102.
(36) Naim, M., Bhat, S., Rankin, K. N., Dennis, S., Chowdhury, S. F., Siddiqi, I., Drabik, P., Sulea, T., Bayly, C. I., Jakalian, A., and Purisima, E. O. (2007) Solvated interaction energy (SIE) for scoring protein-ligand binding affinities. 1. Exploring the parameter space. *J. Chem. Inf. Model. 47*, 122−133.
(37) Sulea, T., and Purisima, E. O. (2012) The solvated interaction energy method for scoring binding affinities. *Methods Mol. Biol. 819*, 295−303.
(38) Hogues, H., Sulea, T., and Purisima, E. O. (2014) Exhaustive docking and solvated interaction energy scoring: lessons learned from the SAMPL4 challenge. *J. Comput.-Aided Mol. Des. 28*, 417−427.
(39) Cui, Q., Sulea, T., Schrag, J. D., Munger, C., Hung, M. N., Naim, M., Cygler, M., and Purisima, E. O. (2008) Molecular dynamics-solvated interaction energy studies of protein-protein interactions: the MP1-p14 scaffolding complex. *J. Mol. Biol. 379*, 787−802.
(40) Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J., and Strobel, S. A. (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature 430*, 45−50.

DOI: 10.1021/acs.biochem.5b01257
Biochemistry 2016, 55, 2319−2331