doi:10.1016/j.jmb.2004.08.041 J. Mol. Biol. (2004) 343, 601–613

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

High Resolution Crystal Structure of a Key Editosome Enzyme from *Trypanosoma brucei*: RNA Editing Ligase 1

Junpeng Deng[1,2], Achim Schnaufer[3,4], Reza Salavati[3,4] Kenneth D. Stuart[3,4] and Wim G. J. Hol[1,2]*

[1]*Howard Hughes Medical Institute, University of Washington, Seattle WA 98195 USA*

[2]*Department of Biochemistry and Biological Structure Biomolecular Structure Center University of Washington P.O. Box 357742, Seattle WA 98195, USA*

[3]*Seattle Biomedical Research Institute, 307 Westlake Avenue N, Suite 500, Seattle, WA 98109 USA*

[4]*Department of Pathobiology University of Washington Seattle, WA 98195, USA*

**Corresponding author*

Trypanosomatids are causative agents of several devastating tropical diseases such as African sleeping sickness, Chagas' disease and leishmaniasis. There are no effective vaccines available to date for treatment of these protozoan diseases, while current drugs have limited efficacy, significant toxicity and suffer from increasing resistance. Trypanosomatids have several remarkable and unique metabolic and structural features that are of great interest for developing new anti-protozoan therapeutics. One such feature is "RNA editing", an essential process in these pathogenic protozoa. Transcripts for key trypanosomatid mitochondrial proteins undergo extensive post-transcriptional RNA editing by specifically inserting or deleting uridylates from pre-mature mRNA in order to create mature mRNAs that encode functional proteins. The RNA editing process is carried out in a ~1.6 MDa multi-protein complex, the editosome. In *Trypanosoma brucei*, one of the editosome's core enzymes, the RNA editing ligase 1 (TbREL1), has been shown to be essential for survival of both insect and bloodstream forms of the parasite. We report here the crystal structure of the catalytic domain of TbREL1 at 1.2 Å resolution, in complex with ATP and magnesium. The magnesium ion interacts with the β and γ-phosphate groups and is almost perfectly octahedrally coordinated by six phosphate and water oxygen atoms. ATP makes extensive direct and indirect interactions with the ligase *via* essentially all its atoms while extending its base into a deep pocket. In addition, the ATP makes numerous interactions with residues that are conserved in the editing ligases only. Further away from the active site, TbREL1 contains a unique loop containing several hydrophobic residues that are highly conserved among trypanosomatid RNA editing ligases which may play a role in protein–protein interactions in the editosome. The distinct characteristics of the adenine-binding pocket, and the absence of any close homolog in the human genome, bode well for the design of selective inhibitors that will block the essential RNA ligase function in a number of major protozoan pathogens.

Keywords: RNA ligase; editosome; trypanosomatids; nucleotidyl transferases; RNA editing

Introduction

Suitable drugs are lacking for several devastating human and animal diseases caused by trypanosomatid pathogens. *Trypanosoma brucei*, which causes African sleeping sickness, *T. cruzi*, which causes Chagas' disease, and several *Leishmania* species, which cause a spectrum of leishmanial diseases, aggregate an annual incidence of about 30 million and cause about one million deaths a year.[1]† All trypanosomatids use a remarkable post-transcriptional process, termed RNA editing, to create most of the mature mitochondrial mRNAs

Abbreviation used: MAD, multiple-wavelength anomalous dispersion.

E-mail address of the corresponding author: wghol@u.washington.edu

† http://www.who.int/csr/resources/publications/CSR_ISR_2000_1tryps/en/index2.html

that encode functional proteins.[2–5] This editing occurs by the specific insertion and deletion of uridylates (Us) in pre-mature mRNAs (pre-mRNA) and, in some cases, is so extensive that the pre-mRNA is almost doubled in size by the end of the process. The edited sequence information is encoded in numerous small *trans*-acting guide RNAs (gRNAs). Each gRNA consists of three regions: (i) a 5′ anchor region that can duplex with pre-mRNA by base-pairing; (ii) an adjacent central region that specifies the precise Us to be inserted and deleted; and (iii) a 3′ oligo(U) tail that might provide for interactions between gRNA, pre-mRNA and the editing enzyme complex. Multiple gRNAs are used in succession to edit an mRNA with each gRNA specifying the editing of several sites and creating new sequences that are complementary to the next gRNA that is used. The editing of each site occurs by a coordinated series of catalytic steps that entails endonucleolytic cleavage, U-addition or U-deletion as specified by the gRNA, and religation of the processed fragments. Although editing endonucleases are not yet identified, two editing 3′-terminal-uridylyl-transferases, candidates for U-specific 3′-exonuclease,[6,7] and two editing RNA ligases have been reported. The RNA-editing enzymes in *T. brucei* are organized into a ~1.6 MDa multi-protein complex, the editosome containing ~20 different proteins.[8–14] Crucial functions are performed by two related RNA editing ligases TbREL1 and TbREL2 of 52 kDa and 48 kDa, respectively, which show 41% amino acid sequence identity.[11,15–21] One of these key enzymes, TbREL1, is the focus of this work.

RNA ligation generally occurs by a three-step process. The ligase first autoadenylylates a critical lysine residue, using ATP to form a covalent protein–AMP intermediate while releasing pyrophosphate. The AMP moiety is then transferred to the 5′ phosphate group of the donor (3′) RNA fragment, generating a 5′–5′ pyrophosphate linkage. In the final step, the 3′ hydroxyl group of the acceptor RNA displaces this 5′ AMP, thereby forming a new phosphodiester bond. Within the editosome, TbREL1 and TbREL2 are primarily associated with subcomplexes for U-deletion and U-insertion RNA editing, respectively, although TbREL1 appears able to functionally replace TbREL2, but not *vice versa*.[15–17] Importantly, TbREL1 is essential for survival of both the disease-causing bloodstream form and the insect form of *T. brucei*.[15,17,18]

Sequence comparisons show that REL1 and REL2 belong to the superfamily of covalent nucleotidyl transferases,[11,15] which share five characteristic signature motifs. The RNA editing ligases from trypanosomatids are all closely related[21] (Figure 1). In contrast, TbREL1 has less than 15% amino acid sequence identity with DNA ligases and only about 20% with bacteriophage T4 RNA ligase 2 (T4Rnl2), the closest non-trypanosomatid homolog identified to date[22–24] (Figure 1). Distinctly different from all the DNA ligases and mRNA capping enzymes, which have been shown to consist of at least an adenylation domain and an OB fold domain, TbREL1 has a N-terminal adenylation domain but an entirely different C-terminal domain[25] (not shown in Figure 1). The latter has been predicted to be involved in interactions with other proteins within the editosome that have putative OB fold domains.[21,25] The five signature motifs of the superfamily, which are crucial for catalysis of ATP-dependent RNA and DNA strand joining,[26] are present in the N-terminal domain of TbREL1. Our activity tests show that the N-terminal domain of TbREL1 alone is capable of overall *in vitro* RNA strand joining, although with reduced activity and altered kinetics (Figure 2). This is contrary to the adenylylation domain of T4Rnl2, which, under comparable conditions, was reported to be inactive for RNA strand joining.[24] Here, we report the crystal structure at 1.2 Å resolution of the N-terminal catalytic domain of TbREL1 in complex with ATP and Mg^{2+}, providing a first and very high-resolution view of an RNA editing ligase that catalyzes 3′–5′ RNA strand joining. Unveiling the structure of the editing RNA ligase from *T. brucei* revealed unique characteristics that not only assist in further understanding of its catalytic mechanism, but also may provide the basis for structure-based drug design.

Results and Discussions

Overall structure

The TbREL1 catalytic domain has been solved by the SeMet multiple-wavelength anomalous dispersion (MAD) method and refined to near-atomic resolution, which is by far the highest resolution of any member of the superfamily to date.[27–31] The domain consists of 11 α-helices and 13 β-strands, which constitute two subdomains, both formed mainly by highly curved anti-parallel β-sheets flanked by α-helices (Figure 3(a)). The contiguous subdomain 1 contains helices 3–6 and strands 2–10; the non-contiguous subdomain 2 comprises helices 1, 2, 7–11 and strands 1 and 11–13. The electron density of a very well defined ATP molecule together with a Mg^{2+} (Figures 3 and 4) is located in a deep pocket between the two curved anti-parallel β-sheets, surrounded by conserved residues from all five signature motifs of the superfamily, which provide exquisite interactions with essentially every atom of the ATP molecule.

The Mg^{2+}:ATP binding site

All the three phosphate groups of the ATP molecule are involved in tight interactions with the ligase (Figure 4(b)). The γ-phosphate group is extensively linked to the protein molecule, both directly *via* a hydrogen bond between its O3G and $N^{\eta 2}$ of R111, and indirectly through a water-mediated hydrogen-bonding network that

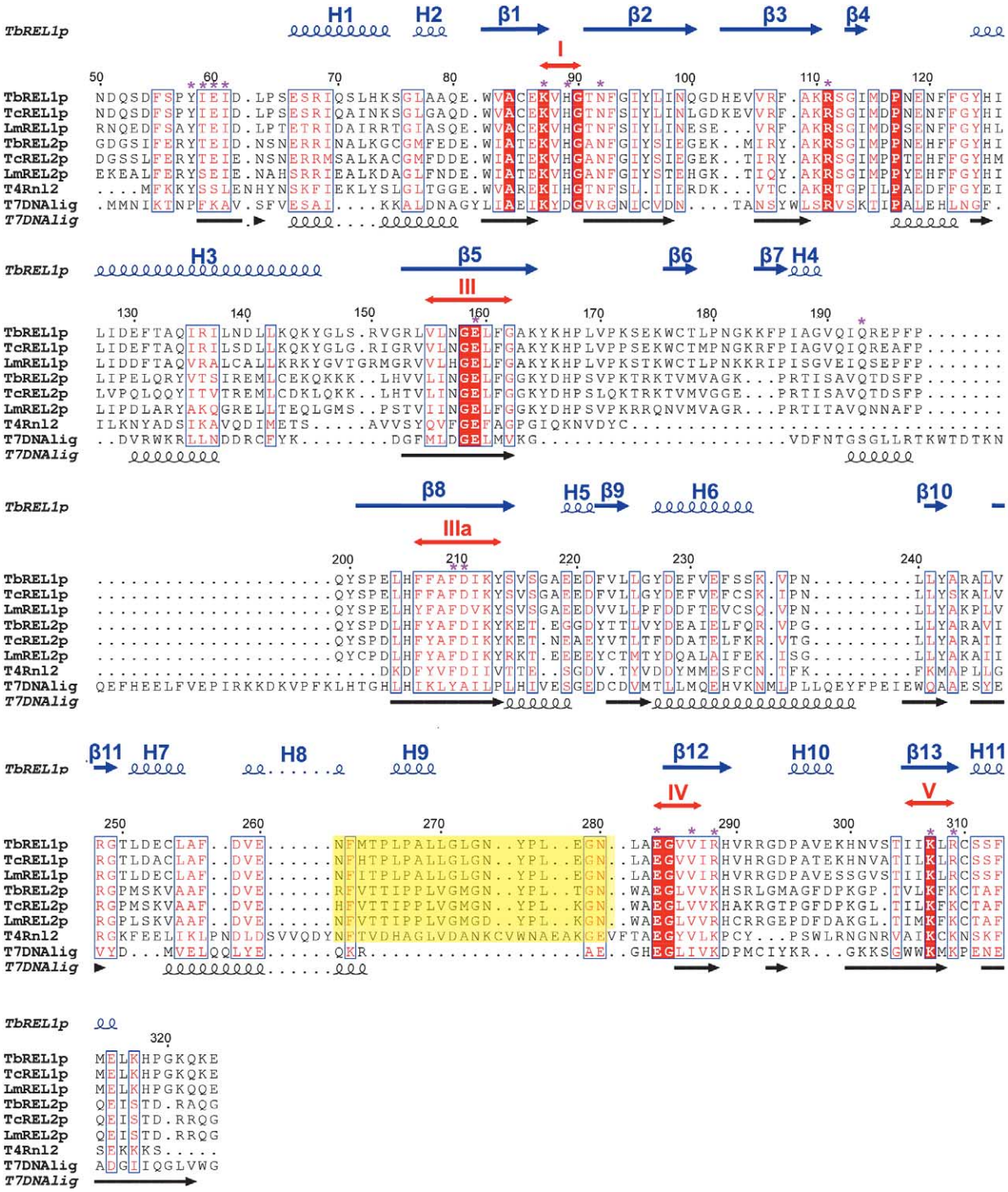

Figure 1. Multiple sequence alignment of adenylylation domains of selected RNA and DNA ligases. Depicted are REL1 and REL2 ligases from *T. brucei*, *T. cruzi* and *L. major*, together with Enterobacteriophage T4 RNA ligases 2 (T4Rnl2, NP_049790) and phage T7 DNA ligase (PDB 1a0i). The blue secondary structure symbols on top represent those observed in the TbREL1 structure. Those on the bottom in black are from the T7 DNA ligase structure. The five conserved ligase motifs, I, III, IIIa, IV and V are shown and conserved residues are indicated by red on white (similar) or white on red letters (identical). The hydrophobic helical loop region characteristic for trypanosomatids is indicated by a yellow box. The residues involved in ATP binding are shown with purple asterisks.

originates from the three hydrated oxygen atoms (Figure 4(c)). The β-phosphate group is directly hydrogen-bonded with the catalytic residue K87 (motif I) and with residue R309 (motif V) through the non-bridging oxygen atoms O2B and O1B, respectively. O1B is additionally linked to the guanidinium group of R309 *via* a water molecule. The α-phosphate group of ATP is stabilized by extensive hydrogen bonding between the non-bridging oxygen atoms and

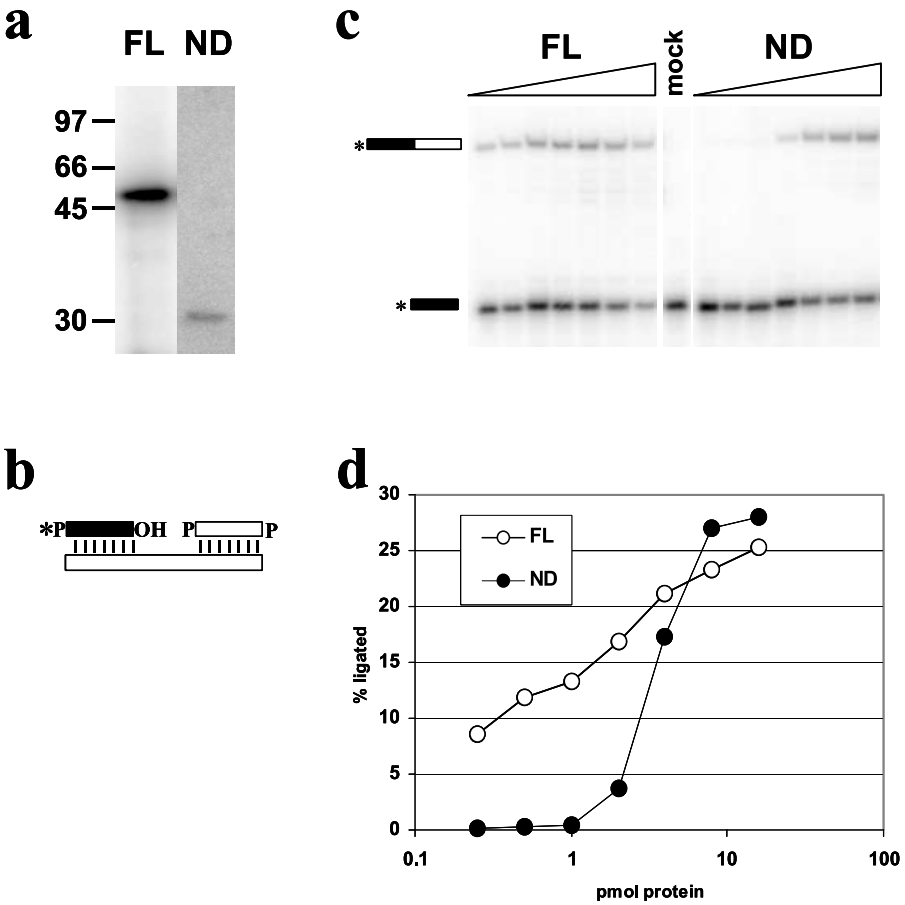

Figure 2. Activity assays. a, Autoadenylation assay.[12] FL, full-length protein; ND, N-terminal domain. Note that the phosphoimager software was used to increase the intensity for the ND lane tenfold compared to the FL lane. b, Cartoon of the ligation assay with nicked, double-stranded RNA substrate. The 5′ phosphate group of the acceptor fragment is labeled with ^{32}P. c, Ligation assay.[37] Increasing amounts (0.25, 0.5, 1, 2, 4, 8, 16 pmol) of full-length protein (FL) or N-terminal domain (ND) were incubated with pre-annealed RNA substrate in the presence of 10 μM ATP. The mock reaction contained no protein. See Methods and Materials for details. d, Quantification of data shown in c.

the protein. O1A forms bifurcated hydrogen bonds with the side-chains of K307 and R309 from motif V, respectively. O2A is hydrogen-bonded with the backbone nitrogen atom of residue I61 as well.

The experimental electron density map reveals that a Mg^{2+} interacts intimately with the β and γ-phosphate groups of the ATP molecule. The divalent magnesium ion is observed to be liganded with nearly perfect octahedral geometry (Figure 4(a)) involving six direct ligands: two non-bridging oxygen atoms, one each from the β and γ-phosphate groups, respectively, and four water molecules. Through three of the directly bound water ligands, the magnesium ion is further linked to conserved residues E236 of motif IV, E159 of motif III, and residues H89 and G90 from motif I.

The ribose moiety of ATP is engaged in extensive hydrogen bonding interactions with the ligase. Bifurcated hydrogen bonds connect both hydroxyl groups on the sugar ring with side-chain atoms of E159, N92, R111 and the main-chain carbonyl group of I59 (Figure 4(b)). Atom O4′ on the sugar ring is also hydrogen-bonded to the amino group of the catalytic residue K87, which has been identified as the lysine residue that forms a phosphoramide bond with AMP during the RNA ligation reaction.[15,18] Contrary to what is observed in the *Chlorella* virus DNA ligase–AMP (1fvi) structure,[30] the adenine base of ATP bound to TbREL1 displays

Figure 3. TbREL1 and its closest distant relatives. (a) Overall structure of the adenylylation domain of TbREL1. The two subdomains of TbREL1 are colored in green and yellow, respectively. Helices are numbered as H1–H11 and β-strands are numbered as b1–b13. (b) Superimpositions of three distantly related nucleotidyl transferase structures. TbREL1 is colored in green and yellow. Purple, T7 DNA ligase (PDB1a0i) with 16% sequence identity to TbREL1; cyan, *Chlorella* virus DNA ligase (PDB1fvi) with 15% sequence identity to TbREL1. Note the TbREL1 H9 region circled in red broken oval. (c) Stereo view of the conserved hydrophobic helical-loop region. The solvent-exposed hydrophobic residues are shown in red.

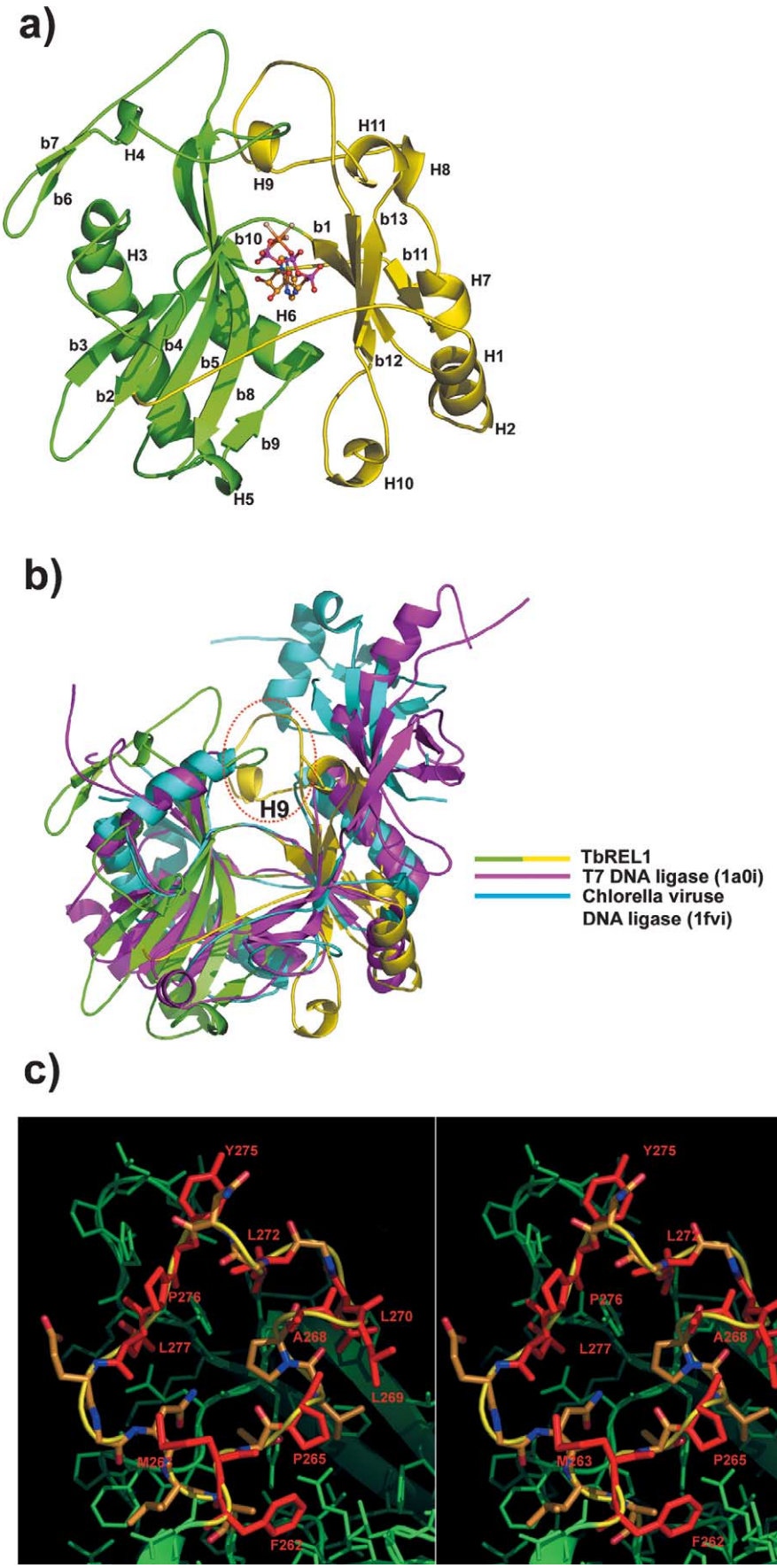
a)
b7
H4
H11
H8
H9
b6
b1
b13
b10
H3
b11
H7
H6
b3
b4
b5
b12
H1
b2
b8
H2
b9
H5
H10
b)
H9
TbREL1
T7 DNA ligase (1a0i)
Chlorella viruse
DNA ligase (1fvi)
c)
Y275
L272
P276
L270
A268
L277
L269
M263
P265
F262
Y275
L272
P276
L270
A268
L277
L269
M263
P265
F262

Figure 3.

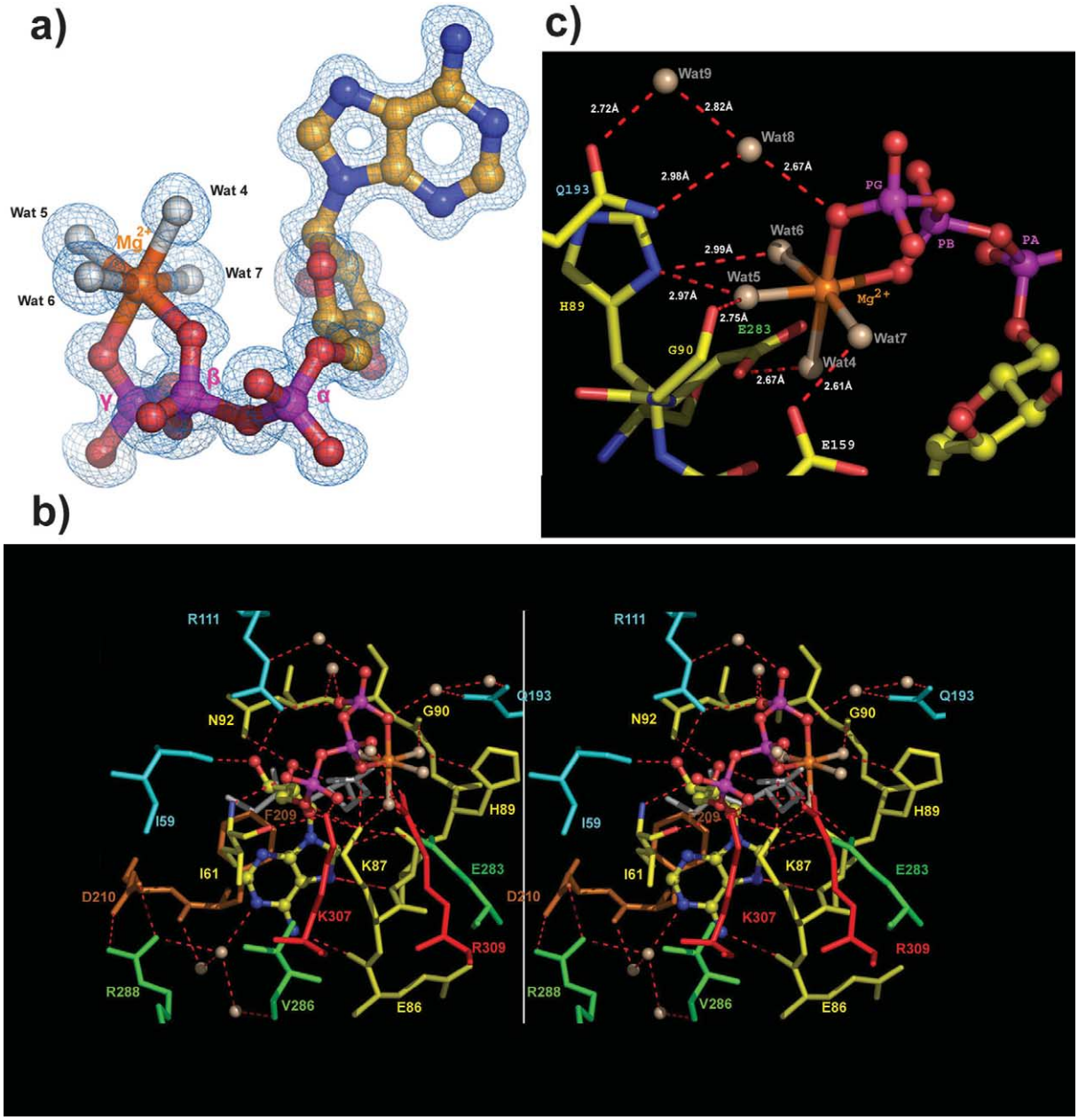

Figure 4. Interactions of Mg^{2+}-ATP with TbREL1. (a) The $2mF_o-DF_c$ electron density map[49] covering the ATP molecule and Mg^{2+}. The Mg^{2+} is shown as an orange sphere. The direct water ligands of Mg^{2+} are shown as white spheres. (b) ATP-binding pocket in stereo. The ATP molecule is shown as ball-and-stick. Mg^{2+} is purple. Water molecules are white. Residues are colored according to their signature sequences: yellow for motif I, white for motif III, brown for motif IIIa, green for motif IV and red for motif V. Residues in cyan are outside the five signature motifs. Hydrogen bonds are shown in red broken lines. (c) The exquisite interactions of the Mg^{2+} with the β and γ-phosphate groups of ATP, TbREL1 and four water molecules.

a *syn* conformation and the sugar ring adopts a C3′-*endo* puckered geometry as observed in the other two DNA ligases[27,31] (1a0i, 1dgs), the mRNA capping enzyme[28] (1ckm) and T4Rnl2.[24]

The adenine base is deeply buried in a pocket formed by conserved residues provided by all the five signature motifs (Figures 4 and 5), which are distributed along the entire polypeptide chain (Figure 1). The six-membered ring of the adenine base is tightly sandwiched by the aromatic ring of F209 (motif IIIa) and the side-chain of V286 (motif IV) (Figures 4(b) and 5(b)). The five-membered ring of the adenine base stacks onto the side-chain of the catalytic K87 (motif I). The N6 atom of the adenine base forms hydrogen bonds with the carbonyl oxygen atom of E86 (motif I). N7 of the adenine base is hydrogen-bonded to the backbone nitrogen atom of V88 (Figure 4(b)). CH···O interactions are observed between C8 and the carbonyl oxygen atom of V88 with a C8 to O distance of 3.16 Å. Important additional interactions of the base occur near the N1 atom on the adenine ring. Three hydrogen-bonded water molecules (Wat1, Wat2, and Wat3) are observed at the "deep end" of the

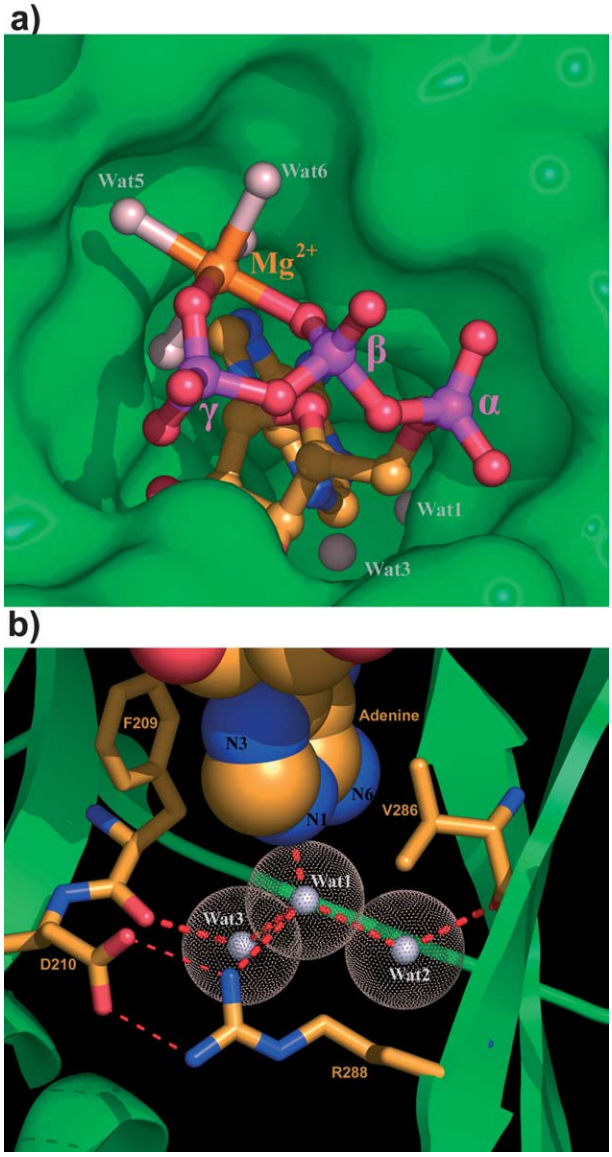

Figure 5. The ATP pocket. (a) Mg^{2+}-ATP in its deep pocket with the protein surface in green. (b) The three water molecules around the deep end of the adenine pocket. Water molecules are shown as white spheres with dotted surface. The adenine base is shown as space-filling model.

adenine pocket (Figures 4(b) and 5). Water molecule Wat1 is directly hydrogen-bonded to N1 as well as to $N^{\eta 1}$ of R288 (motif IV). *Via* water molecules Wat2 and Wat3, water molecule Wat1 is further connected to the backbone carbonyl oxygen atoms of V286 and F209 residues. This water-mediated hydrogen-bonding network is a unique feature of the adenine pocket of TbREL1 in the set of the superfamily structures known to date.[24,27–31]

Comparison with other nucleotidyl transferase structures

A Dali search[32] revealed that the overall fold of TbREL1 resembles closest the adenylylation domains of T4Rnl2 (PDB 1s68), DNA ligases (PDB 1dgs, 1a0i and 1fvi) and an mRNA capping enzyme (PDB 1ckm), though the shared amino acid sequence identity is less than 20% in each pairwise comparison. The r.m.s. deviation of 210 equivalent C^{α} atoms between the current structure and the top *Z*-scoring T4Rnl2 is about 2.3 Å. While the adenine base in T4Rnl2, DNA ligases, and the mRNA capping enzyme occupies a nearly identical position as in TbREL1, the sugar and phosphate group positions differ greatly between TbREL1 and their distant relatives.

TbREL1 displays a new binding mode of the ATP γ-phosphate group and magnesium ion coordination that has not been observed before in the superfamily. In both T7 DNA ligase[27] and the mRNA capping enzyme open-form structures[28], the γ-phosphate group of the non-covalently bound ATP molecule protrudes away from the protein surface, having few interactions with the protein. In contrast, the γ-phosphate group of ATP in TbREL1 is positioned near the protein surface by the hydrated Mg^{2+} and tightly bound to the enzyme through hydrogen bonding interactions (Figure 4). It has been proposed for the mRNA capping enzyme that conformational changes between the two protein domains are required to stabilize the GTP γ-phosphate group in order to switch the conformation of the GTP phosphate backbone so that the α-phosphorus atom is brought close to the catalytic lysine residue (equivalent to K87 in TbREL1) in an attacking position.[28] Furthermore, cooperation of divalent metal ions in the vicinity of the α-phosphate group has been proposed to trigger the guanylylation of the catalytic lysine residue in the capping enzyme.[28] Interestingly, although a divalent magnesium ion is present, the catalytic residue K87 in the TbREL1–ATP-Mg^{2+} structure is not adenylylated (Figure 4), and the magnesium ion appears to stabilize both the β and γ-phosphate groups rather than the α-phosphate group. As observed in the T7 DNA ligase–ATP (1a0i), the open-form mRNA capping enzyme–GTP (1ckm), and the T4Rnl2–AMP (1s68) structures, the ATP molecule in TbREL1 is not covalently attached to the catalytic residue K87.

Combining the observation of the Mg^{2+} coordination of our catalytically active TbREL1 domain with the suggested role of divalent metal ions in the adenylylation step,[28,30] leads to the possibility that Mg^{2+} might play multiple roles in RNA strand joining of TbREL1. First, a Mg^{2+} facilitates the precise positioning of the β and γ-phosphate groups in the initial Mg^{2+}-ATP–TbREL1 complex. Next, possibly after conformational changes induced by the C-terminal domain of the full-length protein, or by protein partners from the editosome complex, another Mg^{2+} binds to the α-phosphate group of ATP and promotes the formation of the covalent bond with residue K87. Indeed, kinetic analysis has shown the two metal-ion mechanism to occur in this class of ligases.[33]

The critical residue K87 of TbREL1 is involved in numerous interactions with ATP: (i) its N^{ζ} interacts with O4′ on the sugar ring, as well as O5′ and O2B of the phosphate group of ATP; (ii) a large part of its side-chain stacks against the adenine base (Figure 4(b)). The α-phosphorus to K87 amino nitrogen distance is 3.74 Å, with the amino group not in an apical position with respect to the leaving pyrophosphate group for nucleophilic attack (Figure 4(b)). Thus, the observed structure is likely to represent an initial stage along the catalytic pathway. Our activity tests show that, when compared to full-length protein, the N-terminal domain of TbREL1 has a reduced overall *in vitro* RNA strand joining activity and altered kinetics (Figure 2c and d). When protein is added in 20-fold or more excess to substrate, activity of the N-terminal domain of TbREL1 is comparable to that of the full-length protein. Auto-adenylylation activity is also much reduced (Figure 2a).

The highly conserved residue K307 from motif V is closer to the α-phosphorus atom of ATP than the critical K87, with a K307 N^{ζ} to α-phosphorus distance of 3.55 Å. The K307 amino group is hydrogen-bonded to the O1A atom (Figure 4(b)). This configuration suggests an important role for K307 in transfer of the α-phosphate group to the canonical catalytic residue K87 in the first step of the adenylylation reaction. This is consistent with the finding that alanine mutation of the corresponding K238 in T7 DNA ligase,[34] which also contacts the α-phosphate group in the T7 DNA ligase–ATP crystal structure,[27] abolished autoadenylylation activity as well as joining of pre-adenylylated substrate. Similarly, the K225A mutation in T4Rnl2 abrogated the adenylylation of the enzyme.[22]

R309 is a unique feature occurring only in trypanosomatid REL1 ligases. The equivalent residue in all other RNA ligases (including trypanosomatid REL2, DNA ligases, and capping enzymes) is a lysine. A conservative replacement of this lysine residue with arginine in RNA ligase T4Rnl2 completely abolished adenylylation and strand-joining of preadenylylated substrates.[22] R309 positions its guanidinium group close to the α and β phosphate groups of ATP. The residue equivalent to R309 in T7 DNA ligase has been implicated in nick recognition.[34] It seems reasonable to assume that the presence of R309 in motif V in TbREL1 is at least partly responsible for its distinct catalytic properties in comparison to TbREL2 (see below).

The closest homologue to TbREL1 discovered to-date is T4Rnl2.[22–24] The non-covalently bound AMP molecule located in the T4Rnl2 structure adopts a conformation nearly identical with that of the AMP moiety from the ATP molecule in the TbREL1 structure. Most of the conserved residues from the five signature motifs display similar geometry in both of the structures (Figure 6). However, superposition of the two RNA ligase structures shows significant displacements of the two basic motif V residues (K307 and R309 in TbREL1, correspondingly K225 and K227 in T4Rnl2). In addition, the backbone atoms of I59 and I61 in TbREL1, which are recognizing the sugar and α-phosphate moieties of the ATP molecule *via* hydrogen bonding interactions, are significantly displaced with respect to the cofactors as well as to the corresponding residues in the T4Rnl2:AMP complex. As a result, the equivalent residues S6 to L8 of T4Rnl2 are not engaged in any direct interactions with the AMP moiety bound (Figure 6).

Unique loop regions

A potentially important structural feature of REL1 of *T. brucei* is formed by residues 262–282, comprising helix 9 followed by a loop. This helix-loop region is highly conserved among all trypanosomatids, very different in length and sequence in T4Rn12, and absent from all other DNA ligases and the mRNA capping enzyme (Figure 1). This

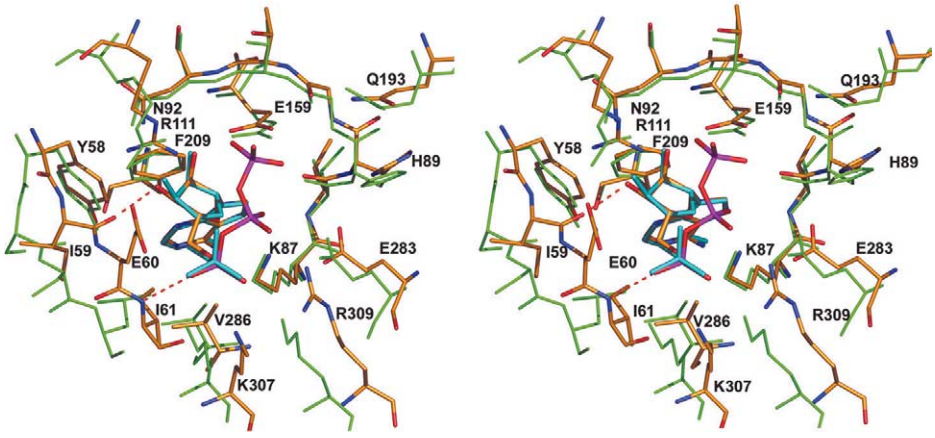

Figure 6. Superposition of TbREL1–ATP and T4Rnl2–AMP in stereo view. The two structures are superimposed on each other based on AMP. TbREL1 and ATP are shown in atom color and the TbREL1 residues near the ATP binding site as labeled. T4Rnl2 (PDB id 1s68) is shown in green and the AMP is shown in cyan. Notice the hydrogen bonding between the backbone atoms of residues I59 and I61 in TbREL1 and the sugar moiety and α-phosphate group of the ATP, respectively, as shown in red broken lines; these hydrogen bonds are absent from the T4Rnl2–AMP complex.

helix-loop region is located behind the ATP-binding site, in between the two halves of the molecule (Figure 3(b)). Superimposition shows that the helix-loop region of TbREL1 is located in the groove between the two domains of the T7 DNA ligase structures (Figure 3(b)), which has been proposed to bind DNA substrate.[27,34] The loop exposes a number of hydrophobic residues such as F262, M263, P265, A268, L269, L270, L272, Y275, P276 and L277 to the solvent (Figure 3(c)). These residues are conserved in all REL1 and REL2 editing ligases in trypanosomatids which suggests a critical role in the editosome, possibly promoting protein–protein interactions that are crucial for forming editosome subcomplexes.[25]

Two additional unique loops extend into this inter-domain area and might be involved in RNA recognition, protein interactions, or both. The first is formed by residues 165 to 175 in between β-strands 5 and 6. The second by residues 190 to 200, in between helix 4 and β-strand 8. Both loops feature positively charged residues (K166 and R194) that extend into the area thought to be occupied by the double-stranded substrate in the superfamily.[34] Thus, these residues appear to be well situated to contact the phosphate backbone of the RNA substrate. In fact, the first loop region of TbREL1 (165–175) is absent from T4Rnl2 full-length protein. The truncated N-terminal adenylylation domain of T4Rnl2 is dysfunctional in RNA strand joining,[24] while the TbREL1 N-terminal domain retains activity albeit at reduced levels (Figure 2). Therefore, unlike TbREL1, T4Rnl2 may need its C-terminal portion to compensate for the loss of this putative RNA-binding loop region that is unique to TbREL1.

A signature for RNA editing ligases

The extensive interactions between conserved residues in the five ligase motifs and the ATP-Mg^{2+} ligand are consistent with their essential roles in catalysis, as has been confirmed by mutagenesis studies with RNA and DNA ligases[27–31] and T4Rnl2.[22,23] For example, T4Rnl2 mutagenesis studies showed that the conserved residues equivalent to K87 in motif I, E283 in motif IV, and K307 and R309 in motif V are required for overall RNA strand joining activity, in particular for the initial adenylylation step. The conserved histidine residue in motif I (corresponding to H89 in TbREL1) was shown not to be critical for adenylylation but crucial for strand joining.[22,23] In the current structure, this histidine residue appears to be located in the vicinity of the ATP γ-phosphate group and positioned at the surface of the TbREL1 protein: its $N^{\delta 1}$ is hydrogen-bonded with two direct Mg^{2+} water ligands, while $N^{\varepsilon 2}$ interacts with $O^{\varepsilon 1}$ of E195. Likely significant rearrangements of the active site after formation of the phosphoramide bond and removal of the pyrophosphate make it premature to predict a precise role for H89 on the basis of the current structure.

Outside of the five superfamily motifs, seven residues (Y58, E60, I61, E86, N92, R111 and Q193) in the active-site region are completely conserved among trypanosomatids (Figure 1), suggesting critical functions in RNA strand joining. Y58 is involved through its side-chain in hydrogen bonding and van der Waals interactions with D210 and F209 from motif IIIa, respectively. E60 is linked to the γ-phosphate group of ATP *via* water molecules. The backbone of I61 is hydrogen-bonded with the α-phosphate group of ATP and K307 from motif V. The side-chain of E86 is pointing away from the ATP-binding site and is engaged in interactions with the helix-loop region 262–282. N92 and R111 are recognizing the hydroxyl groups on the sugar moiety of ATP (Figure 4(b)). The side-chain of Q193 forms hydrogen bonds with the main chain of G90 in motif I and with the γ-phosphate group of ATP *via* a water molecule (Figure 4(b)). These residues are all located close to the ATP-binding site (Figure 7) and can be considered as a distinctive signature of RNA editing ligases in all trypanosomatids. Determining the precise role of these seven amino acid residues during the editing process will obviously require further investigations.

Examining the surface of TbREL1 around the ATP-binding site reveals three solvent-exposed aromatic residues: F121, F122 and F313 (Figure 7). These residues are completely conserved among the RNA ligases of the superfamily (Figure 1). Their aromatic rings are within 10 Å from the ATP site. They might be involved in non-specific interactions with RNAs by base-stacking interactions. Indeed, these types of interactions have been found in a variety of protein–DNA complexes.[35–38]

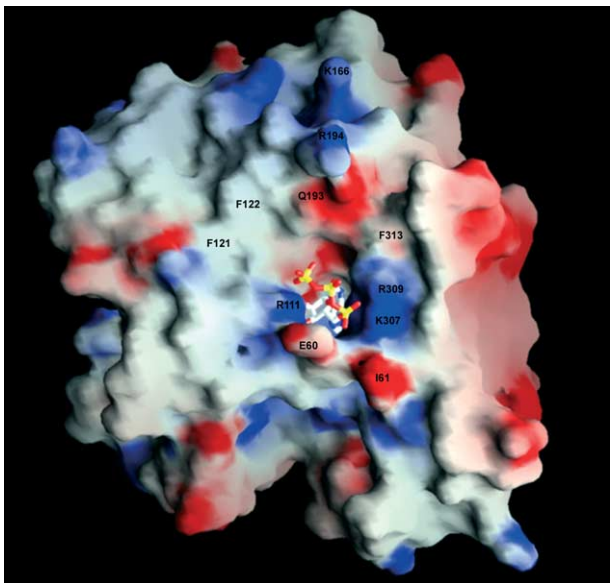

Figure 7. Electrostatic potential surface of TbREL1 generated by GRASP.[50] Blue: positive; red: negative. ATP is shown as ball-stick. Notice the negatively charged residue E60 near the ATP and the absence of extended positive potential surfaces. Seven conserved trypanosomatid residues are labeled.

REL1 *versus* REL2 ligases

As mentioned above, TbREL1 and TbREL2 ligases have been shown to have distinct RNA ligation activity.[25,39,40] TbREL2 appears to have a higher affinity for ATP, while TbREL1 has less stringent substrate requirements[20,39,40] and is able to substitute for REL2, but the reverse is not the case.[15–18] Analysis of the sequences of the two enzymes shows that they share almost all the key residues around the ATP-binding site. However, residue R309 is unique among all the REL1 orthologs (*T. brucei* REL1, *T. cruzi* REL1 and *Leishmania major* REL1), while the equivalent Lys residue in the REL2 paralogs is consistent with the canonical motif V sequence (Figure 1). This particular residue might contribute to different characteristics of TbREL1 and TbREL2, but exactly how TbREL1 can be involved in both U insertion and deletion, and TbREL2 only in U deletion, remains to be determined.

Quite remarkably, the RNA-binding site remains elusive, since the surface of TbREL1 does not show obvious positive electrostatic potential patches around the ATP-binding site (Figure 7) as is the case in DNA ligase structures.[27–31] One possibility is that binding partners of the ligase in the editosome, which contain zinc fingers and putative OB fold domains,[25] contribute significantly to RNA binding and that isolated RNA editing ligases bind RNA only with weak affinity. Thus, the distinct substrate preference of the two RNA ligases *in vivo* might arise mainly from their different partners in the editosome.

The ATP-binding pocket of TbREL1 in our binary complex appears to be a promising drug target. Of particular interest are the three water molecules at the "deep end" of the pocket that interact with the N1 atom of the adenine base as well as with backbone and side-chain atoms of key residues in the vicinity (Figure 5(b)). The characteristics of this adenine-binding pocket, in particular the conserved salt-bridge between D210 and R288 (Figures 1 and 5(b)), provide opportunities for the design of compounds that will recognize trypanosomatid REL1s with high specificity and affinity. Therefore, the structure presented provides a new platform for the battle against several of the most notorious tropical diseases.

Methods and Materials

Protein expression and purification

The N-terminal 50 amino acid residues of TbREL1, predicted to be a mitochondrial importing signal, were omitted from our constructs. A number of constructs from both *Trypanosoma brucei* and *Leishmania* species have been made. The crystallized *T. brucei* REL1 construct consists of residues 51–324 with a TEV-cleavable N-terminal 6xHis tag. Protein was expressed in *E. coli* cell strain BL21goldDE3. The cells were grown at 37 °C until A_{600} reached 0.8, then cooled to 18 °C. The cells were induced with 1 mM IPTG and incubated further for 12 hours at 18 °C. SeMet-substituted protein was expressed in M9 minimal medium supplemented with amino acids as described by Van Duyne *et al*.[41] The cells were harvested by centrifugation and resuspended in buffer A (20 mM Tris, 300 mM NaCl, 10% (v/v) glycerol, 20 mM imidazole, pH 8.0). Cells were lysed by three cycles in a French press and insoluble material was removed by centrifugation. Subsequent purification steps were carried out at 4 °C. The supernatant was passed through a Ni-NTA column, washed with buffer A, and eluted with buffer A containing 250 mM imidazole. TEV was added immediately into the collected fractions after Ni-NTA purification and the mixture was incubated at 4 °C overnight. The TEV-treated REL1 protein mixture was then buffer-exchanged into 20 mM Tris, 150 mM NaCl, 1 mM EGTA, pH 7.8, using a PD-10 desalting column. The protein solution was subsequently passed through a second Ni-NTA column to remove the cleaved His-tag, His-tagged rTEV protease and any uncleaved TbREL1. The flow-through was concentrated to 15 mg/ml using a Centricon spin-column and applied onto a Superdex-75 column for a size-exclusion chromatography step. The protein appeared to be monomeric with an apparent mass of 31 kDa. The peak fraction was concentrated to 10 mg/ml and 5 mM Na-ATP was added. The purified protein was flash-frozen in liquid nitrogen in 50 μl aliquots and preserved at −80 °C.[42]

The full-length TbREL1 alone failed to be expressed in soluble form in *E. coli*. A bi-cistronic co-expression in *E. coli* of full-length TbREL1 together with its protein partner TbMP63, yielded enough soluble protein for activity assays. In the bi-cistronic construct, there is an N-terminal cleavable 6xHis tag on TbREL1 and no tags on TbMP63. The protein purification process started with Ni-NTA column purification as described above, followed by TEV cleavage of the 6xHis tag, and finished by flowing through the second Ni-NTA column after buffer exchange into buffer A as above.

Crystallization and data collection

The protein aliquot was thawed in hand immediately before crystallization experiments. Sitting-drop, vapor-diffusion with 100 mM HEPES, 100 mM Mg^{2+} and 35% (w/v) PEG3350 at pH 7.0 reservoir solution yielded promising but small crystals. Micro-seeding resulted in single crystals with excellent diffraction quality. The above-mentioned crystallization condition in 20% PEG3350 plus 15% glycerol was used as cryo-protectant. The data collection was done at beam line 19-ID at APS. Three-wavelength MAD data (inflection, peak and remote) around the Se absorption K-edge were collected to 1.8 Å resolution at 100 K. Two passes were used for high-resolution data collection. All the data were processed and reduced by HKL2000.[43] The crystal belongs to space group $P2_1$, with cell dimensions of $a=44.9$ Å, $b=58.6$ Å, $c=53.0$ Å and $\beta=100.2°$, and one molecule in the asymmetric unit. The Matthew's coefficient is 2.28 Å^3/Da, corresponding to 46.0% (v/v) solvent content. For further data collection statistics, see Table 1.

Structure analysis and refinement

The structure was solved by the multiple-wavelength anomalous dispersive (MAD) method using the program SOLVE.[44] Three out of four Se sites were clearly located by SOLVE. After density modification by RESOLVE,[44] 256 out of a total of 277 residues were automatically built by

Table 1. Data and statistics

	Inflection	Peak	Remote	High resolution
A. *Crystal data*				
Wavelength (Å)	0.97899	0.97880	0.96112	0.97903
Cell constants				
a, *b*, *c* (Å)	44.9, 58.6, 53.0			
β (deg.)	100.23°			
Resolution (Å)	1.77	1.77	1.77	1.1
Total reflections	94,872	94,641	95,873	667,327
Unique reflections	26,366	26,332	26,429	95,711
Completeness[a] (%)	100 (99.7)	100 (99.7)	100 (99.8)	86.8 (34.6)
I/o[a]	21.8 (11.6)	30.5(15.7)	22.4(5.4)	33.8 (3.2)
R_{sym}[a] (%)	4.6 (9.3)	5.0 (9.6)	4.1 (9.1)	6.4 (27.2)
B. *Refinement statistics*				
Resolution range (Å)		20–1.2		
No. reflections		78,835		
R_{work}/R_{free} (%)		12.8/14.8		
rmsd bond lengths (Å)		0.010		
rmsd bond angles (deg.)		1.420		
rmsd chiral (deg.)		0.084		

[a] Overall data statistics and (in parentheses) outer-shell data statistics.

Arp/Warp.[45] Seven additional residues were manually built using the program O.[46] The final protein model contains residues 5–269. The N-terminal four residues as well as the eight residues at the C terminus were not observed in the electron density map. The experimental electron density map showed one ATP molecule together with one Mg^{2+} with great clarity. The program Refmac5[47] was used in subsequent refinement. The *B*-factors were first isotropically refined to 1.3 Å resolution and then anisotropically for data up to 1.2 Å resolution. Refinement statistics are shown in Table 1. The final crystallographic R_{work} and R_{free} are 12.8% and 14.8%, respectively. The final protein model has 89.5% of all residues residing in the most-favored region of the Ramachandran plot and 10.5% in additionally allowed region, as calculated by PROCHECK.[48]

Adenylylation assays

Adenylylation reactions were performed as described[15] with 100 ng of protein in a total volume of 20 µl and in a buffer containing 12.5 mM Hepes (pH 7.9), 25 mM KCl, 5 mM magnesium acetate, 0.25 mM DTT, 10% (v/v) DMSO, and 5 µCi of [α-^{32}P]ATP. Reactions were incubated at 28 °C for 15 minutes, mixed with SDS sample buffer, fractionated by SDS-PAGE, and analyzed by phosphorimaging (Storm, Molecular Dynamics).

In vitro ligation assay

Ligation assays using a nicked, double-stranded RNA substrate were performed as described.[40] Briefly, 0.125 pmol of RNA substrates ^{32}P-5′Cleav18 (acceptor RNA), 0.5 pmol of 3′Clepp (donor RNA), and 0.25 pmol of ArtgA6 (gRNA) were pre-annealed and incubated with the indicated amounts of protein and 10 µM ATP in HHE buffer[37] in a total reaction volume of 15 µl for 60 minutes at 28 °C. Reactions were quenched with SDS, phenol-extracted, precipitated, resuspended in urea buffer, fractionated on 7 M urea, 9% (w/v) polyacrylamide gels, and analyzed by phosphorimaging (Storm, Molecular Dynamics).

Protein Data Bank accession code

The atomic coordinates and diffraction amplitudes have been deposited in the Protein Data Bank with accession code 1XDN.

Acknowledgements

We gratefully acknowledge the staff of beam-line 19ID at the Advanced Photon Source for their support. W.G.J.H. acknowledges the Murdock Charitable Trust for a major equipment grant to the Biomolecular Structure Center at the University of Washington. This work received support from NIH grant GM42188 and built on work supported by NIH grant AI14102 to K.D.S.

References

1. World Health Organization (2003). *WHO Report on Global Surveillance of Epidemic-prone Infectious Diseases–African trypanosomiasis* WHO, Rome.
2. Stuart, K., Allen, T. E., Heidmann, S. & Seiwert, S. D. (1997). RNA editing in kinetoplastid protozoa. *Microbiol. Mol. Biol. Rev.* **61**, 105–120.
3. Stuart, K., Panigrahi, A. K., Schnaufer, A., Drozdz, M., Clayton, C. & Salavati, R. (2002). Composition of the editing complex of *Trypanosoma brucei*. *Phil. Trans. Roy. Soc. Ser. B*, **357**, 71–79.
4. Simpson, L., Sbicego, S. & Aphasizhev, R. (2003). Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business. *RNA*, **9**, 265–276.
5. Madison-Antenucci, S., Grams, J. & Hajduk, S. L. (2002). Editing machines: the complexities of trypanosome RNA editing. *Cell*, **108**, 435–438.
6. Ernst, N. L., Panicucci, B., Igo, R. P., Panigrahi, A. K., Salavati, R. & Stuart, K. (2003). TbMP57 is a 3′ terminal uridylyl transferase (TUTase) of the *Trypanosoma brucei* editosome. *Mol. Cell*, **11**, 1525–1536.

7. Aphasizhev, R., Sbicego, S., Peris, M., Jang, S.-H., Aphasizheva, I., Simpson, A. M. *et al*. (2002). Trypanosome mitochondrial 3′ terminal uridylyl transferase (TUTase): the key enzyme in U-insertion/deletion RNA editing. *Cell*, **108**, 649–660.
8. Pollard, V. W., Harris, M. E. & Hajduk, S. L. (1992). Native mRNA editing complexes from *Trypanosoma brucei* mitochondria. *EMBO J.* **11**, 4429–4438.
9. Rusche, L. N., Cruz-Reyes, J., Piller, K. J. & Sollner-Webb, B. (1997). Purification of a functional enzymatic editing complex from *Trypanosoma brucei* mitochondria. *EMBO J.* **16**, 4069–4081.
10. Madison-Antenucci, S. & Hajduk, S. L. (2001). RNA editing-associated protein 1 is an RNA binding protein with specificity for preedited mRNA. *Mol. Cell*, **7**, 879–886.
11. Panigrahi, A. K., Gygi, S. P., Ernst, N. L., Igo, R. P., Jr, Palazzo, S. S., Schnaufer, A. *et al*. (2001). Association of two novel proteins, TbMP52 and TbMP48, with the *Trypanosoma brucei* RNA editing complex. *Mol. Cell. Biol.* **21**, 380–389.
12. Panigrahi, A. K., Schnaufer, A., Carmean, N., Igo, R. P., Jr, Gygi, S. P. & Ernst, N. L. (2001). Four related proteins of the *Trypanosoma brucei* RNA editing complex. *Mol. Cell. Biol.* **21**, 6833–6840.
13. Panigrahi, A. K., Schnaufer, A., Ernst, N. L., Wang, B., Carmean, N., Salavati, R. & Stuart, K. (2003). Identification of novel components of *Trypanosoma brucei* editosomes. *RNA*, **9**, 484–492.
14. Aphasizhev, R., Aphasizheva, I., Nelson, R. E., Gao, G., Simpson, A. M., Kang, X. *et al*. (2003). Isolation of a U-insertion/deletion editing complex from *Leishmania tarentolae* mitochondria. *EMBO J.* **22**, 913–924.
15. Schnaufer, A., Panigrahi, A. K., Panicucci, B., Igo, R. P., Jr, Wirtz, E., Salavati, R. & Stuart, K. (2001). An RNA ligase essential for RNA editing and survival of the bloodstream form of *Trypanosoma brucei*. *Science*, **291**, 2159–2162.
16. Drozdz, M., Palazzo, S. S., Salavati, R., O'Rear, J., Clayton, C. & Stuart, K. (2002). TbMP81 is required for RNA editing in *Trypanosoma brucei*. *EMBO J.* **21**, 1791–1799.
17. Gao, G. & Simpson, A. M. (2003). Is the *Trypanosoma brucei* REL1 RNA ligase specific for U-deletion RNA editing, and is the REL2 RNA ligase specific for U-insertion editing? *J. Biol. Chem.* **278**, 27570–27574.
18. Huang, C. E., Cruz-Reyes, J., Zhelonkina, A. G., O'Hean, S., Wirtz, E. & 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**, 4694–4703.
19. McManus, M. T., Shimamura, M., Grams, J. & Hajduk, S. L. (2001). Identification of candidate mitochondrial RNA editing ligases from *Trypanosoma brucei*. *RNA*, **7**, 167–175.
20. Rusche, L. N., Huang, C. E., Piller, K. J., Hemann, M., Wirtz, E. & Sollner-Webb, B. (2001). The two RNA ligases of the *Trypanosoma brucei* RNA editing complex: cloning the essential band IV gene and identifying the band V gene. *Mol. Cell. Biol.* **21**, 979–989.
21. Worthey, E. A., Schnaufer, A., Mian, I. S., Stuart, K. & Salavati, R. (2003). Comparative analysis of editosome proteins in trypanosomatids. *Nucl. Acids Res.* **31**, 6392–6408.
22. Ho, C. K. & Shuman, S. (2002). Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proc. Natl Acad Sci. USA*, **99**, 12709–12714.
23. Yin, S., Ho, C. K. & Shuman, S. (2003). Structure–function analysis of T4 RNA ligase 2. *J. Biol. Chem.* **278**, 17601–17608.
24. Ho, C. K., Wang, L. K., Lima, C. D. & Shuman, S. (2004). Structure and mechanism of RNA ligase. *Structure*, **12**, 327–339.
25. Schnaufer, A., Ernst, N. L., Palazzo, S. S., O'Rear, J., Salavati, R. & Stuart, K. (2003). Separate insertion and deletion subcomplexes of the *Trypanosoma brucei* RNA editing complex. *Mol. Cell*, **12**, 307–319.
26. Shuman, S. & Schwer, B. (1995). RNA capping enzyme and DNA ligase: a superfamily of covalent nucleotidyl transferases. *Mol. Microbiol.* **17**, 405–410.
27. Subramanya, H. S., Doherty, A. J., Ashford, S. R. & Wigley, D. B. (1996). Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. *Cell*, **85**, 607–615.
28. Hakansson, K., Doherty, A. J., Shuman, S. & Wigley, D. B. (1997). X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. *Cell*, **89**, 545–553.
29. Singleton, M. R., Hakansson, K., Timson, D. J. & Wigley, D. B. (1999). Structure of the adenylation domain of an NAD^+-dependent DNA ligase. *Struct. Fold Des.* **7**, 35–42.
30. Odell, M., Sriskanda, V., Shuman, S. & Nikolov, D. B. (2000). Crystal structure of eukaryotic DNA ligase-adenylate illuminates the mechanism of nick sensing and strand joining. *Mol. Cell*, **6**, 1183–1193.
31. Lee, J. Y., Chang, C., Song, H. K., Moon, J., Yang, J. K., Kim, H. K. *et al*. (2000). Crystal structure of NAD(+)-dependent DNA ligase: modular architecture and functional implications. *EMBO J.* **19**, 1119–1129.
32. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.
33. Cherepanov, A. V. & de Vries, S. (2002). Kinetic mechanism of the Mg^{2+}-dependent nucleotidyl transfer catalyzed by T4 DNA and RNA ligases. *J. Biol. Chem.* **277**, 1695–1704.
34. Doherty, A. J. & Dafforn, T. R. (2000). Nick recognition by DNA ligases. *J. Mol. Biol.* **296**, 43–56.
35. Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. & Frappier, L. (1997). Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature*, **385**, 176–181.
36. Bochkareva, E., Korolev, S., Lees-Miller, S. P. & Bochkarev, A. (2002). Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. *EMBO J.* **21**, 1855–1863.
37. Peersen, O. B., Ruggles, J. A. & Schultz, S. C. (2002). Dimeric structure of the Oxytricha nova telomere end-binding protein alpha-subunit bound to ssDNA. *Nature Struct. Biol.* **9**, 182–187.
38. Lei, M., Podell, E. R., Baumann, P. & Cech, T. R. (2003). DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA. *Nature*, **426**, 198–203.
39. Cruz-Reyes, J., Zhelonkina, A. G., Huang, C. E. & Sollner-Webb, B. (2002). Distinct functions of two RNA ligases in active *Trypanosoma brucei* RNA editing complexes. *Mol. Cell. Biol.* **22**, 4652–4660.
40. Palazzo, S. S., Panigrahi, A. K., Igo, R. P., Salavati, R. & Stuart, K. (2003). Kinetoplastid RNA editing ligases: complex association, characterization, and substrate requirements. *Mol. Biochem. Parasitol.* **127**, 161–167.
41. Van Duyne, G. D., Standaert, R. F., Karplus, P. A.,

Schreiber, S. L. & Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124.
42. Deng, J., Davies, D. R., Wisedchaisri, G., Wu, M., Hol, W. G. & Mehlin, C. (2004). An improved protocol for rapid freezing of protein samples for long-term storage. *Acta Crystallog. sect. D*, **60**, 203–204.
43. Otwinowski, Z. & Minor, W. (1997). DENZO. Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* (Carter, C. W. J. & Sweet, R. M., eds), vol. 276, pp. 307–326, Academic Press, New York.
44. Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallog. sect. D*, **4**, 849–861.
45. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463.
46. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
47. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallog. sect. D*, **55**, 247–255.
48. Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallog. sect. D*, **55**, 191–205.
49. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
50. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.

Edited by J. Doudna

(Received 12 May 2004; received in revised form 11 August 2004; accepted 12 August 2004)