6966–6977 Nucleic Acids Research, 2012, Vol. 40, No. 14 Published online 4 May 2012
doi:10.1093/nar/gks369

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

Young-Jun Park[1], Tanya Budiarto[1], Meiting Wu[1], Els Pardon[2,3], Jan Steyaert[2,3] and Wim G. J. Hol[1,*]

[1]Biomolecular Structure Center, Department of Biochemistry, School of Medicine, University of Washington, Seattle, WA 98195, USA, [2]Structural Biology Brussels, Vrije Universiteit Brussel and [3]Department of Structural Biology, VIB, Pleinlaan 2, B-1050 Brussels, Belgium

Received February 15, 2012; Revised April 10, 2012; Accepted April 11, 2012

ABSTRACT

Trypanosomatids, such as the sleeping sickness parasite *Trypanosoma brucei*, contain a ~20S RNA-editing complex, also called the editosome, which is required for U-insertion/deletion editing of mitochondrial mRNAs. The editosome contains a core of 12 proteins including the large interaction protein A1, the small interaction protein A6, and the editing RNA ligase L2. Using biochemical and structural data, we identified distinct domains of *T. brucei* A1 which specifically recognize A6 and L2. We provide evidence that an N-terminal domain of A1 interacts with the C-terminal domain of L2. The C-terminal domain of A1 appears to be required for the interaction with A6 and also plays a key role in RNA binding by the RNA-editing ligase L2 *in trans*. Three crystal structures of the C-terminal domain of A1 have been elucidated, each in complex with a nanobody as a crystallization chaperone. These structures permitted the identification of putative dsRNA recognition sites. Mutational analysis of conserved residues of the C-terminal domain identified Arg703, Arg731 and Arg734 as key requirements for RNA binding. The data show that the editing RNA ligase activity is modulated by a novel mechanism, i.e. by the *trans*-acting RNA binding C-terminal domain of A1.

INTRODUCTION

Trypanosomatids are the causative agent of human trypanosomiasis, Chagas disease and leishmaniasis which are important diseases in tropical and subtropical regions of the world. Trypanosomatids undergo remarkable changes in their morphology and energy generation during their insect and mammalian bloodstream life cycles that is coupled to mitochondrial gene expression (1–3). The maturation of most mitochondrial genes in trypanosomes requires major RNA-editing steps (4–6). In trypanosomatids, RNA editing is a post-transcriptional modification which alters the mitochondrial mRNA transcripts by insertion or deletion of uridylates (U insertion/deletion RNA editing). This process is mediated by small RNAs known as guide RNAs (gRNAs) (7–9). RNA-editing steps are catalyzed by several multi-enzyme complexes, a major one of which is called the '~20S editosome and also editosome'. In *Trypanosoma brucei*, 20 proteins have been found to be associated with the editosome. U-insertion/deletion RNA editing is a very complicated process which involves a cascade of enzymatic steps (10–12).

There are three distinct types of ~20S editosomes (13–15), which share a common core of 12 proteins including four enzymes (i.e. the RNA-editing ligase 1 (L1), RNA-editing ligase 2 (L2), the U-specific 3′-exouridylylase X2 and the 3′-terminal uridylyl-transferase T2) and six 'interaction proteins': A1, A2, A3, A4, A5 and A6. (For editosome protein nomenclature, see e.g. (11,16,17).) Many of these enzymes and interaction proteins have been shown to be essential for the function of the editosome (18–26). The six interaction proteins vary greatly in length, but each contains a predicted oligonucleotide-binding (OB)-fold near the C-terminus (27). Comprehensive interaction mapping and mass spectrometry studies revealed that the enzymes are networked by specific associations with interaction proteins (14,17). Four editosome interaction proteins (A1–A4) have been shown to interact with A6 suggesting the formation of 'a five OB-fold center' in the core of the editosome that functions as a scaffold for the entire

*To whom correspondence should be addressed. Tel: +1 206 685 7044; Fax: +1 206 685 7002; Email: wghol@u.washington.edu
Present address:
Meiting Wu, Department of Bioengineering, University of Washington, Seattle, WA 98195, USA.

editosome (15,17,28,29). The largest interaction protein A1 contains two zinc-finger motifs followed by a C-terminal OB-fold (27). The interactions between full length A1 and other editosome proteins, including A6 and L2, have been reported before (17). However, which domains of A1 are responsible for interacting with these two other editosome proteins was unknown so far.

The two editing RNA ligases L1 and L2 share a high degree of amino acid sequence identity (30,31) and belong to the nucleotidyltransferase (NTase) superfamily, which also includes ATP and NAD-dependent DNA ligases as well as eukaryotic mRNA capping enzymes (30). The ligases catalyze the sealing of the 3′-hydroxyl group and the 5′-phosphate group of DNA or RNA termini via a series of nucleotidyl transfer reactions (32–34). All DNA ligases and mRNA capping enzymes contain a common catalytic core structure composed of a NTase domain and an OB-fold domain (32,35–37).

DNA (or RNA) substrate recognition by ligases is a key step of the ligation reaction. It has been shown that the ligase remains bound to DNA (or RNA) in most of the steps of the reaction (35,38–41). Based on several DNA ligase crystal structures, the OB-fold domain of the ligase is required for the initial auto-adenylation step (33,37,42,43). This domain plays a major role in DNA binding and is implicated in specific substrate recognition (35,38,39,44,45). Interestingly, unlike most DNA ligases and mRNA capping enzymes, the RNA ligase family, including T4 RNA ligase 1 (Rnl1), T4 RNA ligase 2 (Rnl2) and RNA-editing ligases L1 and L2, lacks an OB-fold domain (30,40,46).

The RNA-editing ligases L1 and L2 are distinguished from other DNA or RNA ligases in that they are embedded in the editosome, where interaction proteins improve RNA ligation efficiency by an unknown mechanism (15,20). It has been reported that full length A1 and A2 stimulate the auto-adenylation and ligation activity of L2 and L1, respectively (15,47). However, which domains of interaction proteins interact with RNA-editing ligases, and how interaction proteins A1 and A2 stimulate the activity of RNA-editing ligases remained unknown so far.

Through a combination of biochemical and structural studies, we have identified the functions of two domains of the interaction protein A1. An N-terminal domain of A1 appears to interact with the C-terminal region of L2. The C-terminal OB domain of A1 ($A1^{OB}$) links A1 to A6 and therefore to the OB-fold center in the core of the editosome (29). Importantly, the C-terminal OB domain of A1 stimulates nicked dsRNA substrate binding by L2 *in vitro*. It appears that the A1•L2 complex has two RNA-binding modules participating in dsRNA recognition: the NTase domain of L2 and the OB-fold domain of A1, each of which alone fails to bind tightly to nicked dsRNA. We discovered that three basic residues of the $A1^{OB}$ are essential for RNA binding by the A1•L2 complex. It is therefore most likely that, also in the editosome, the C-terminal domain of A1 assists RNA-binding by the editing RNA ligase L2 *in trans*.

MATERIALS AND METHODS

Cloning of *T. brucei* $A1^{OB}$•A6 and *T. brucei* A1•L2

The binary complex of *T. brucei* editosome proteins $A1^{OB}$ and A6 was produced using a co-expression strategy as described previously (17). The gene for *T. brucei* A6 (residues 20–164) and the gene encoding residues 626–762 of *T. brucei* A1 ($A1^{OB}$) were cloned into the bi-cistronic expression vector pRSF. A construct for expressing a variant of $A1^{OB}$ was made with the 38-residue loop L23, comprising residues 658–695, replaced by a linker GASG. The protein expressed by this construct is hereafter called $A1^{OB\Delta}$.

For the expression and purification of wild-type *T. brucei* A1•L2 binary complexes, the gene for *T. brucei* L2 (residues 21–416) was cloned into pACYC without any tag (hereafter called full length L2). The gene for *T. brucei* A1 (residues 196–762) was cloned into pSKB3 for expression with an N-terminal His_6-tag followed by a tobacco etch virus (TEV) protease cleavage site. Full length A1 Δ658–695 (hereafter called $A1^{\Delta}$) construct with 38 residues in loop L23 replaced by GASG was subcloned into a pACYC vector. Full length A1 variants with glutamate substitution in R703, K715, K719, R731, R734, K741 and R742 were introduced by the quick change site-directed mutagenesis method. Wild-type, truncations, L23 deletion and point mutants of A1 and L2 were prepared as summarized in the Supplementary Methods.

Crystallization of the *T. brucei* $A1^{OB\Delta}$•^{A1}Nb10 complex

Anti-A1 nanobodies (^{A1}Nbs) were generated and used as crystallization chaperones. Crystals were grown via the sitting-drop vapor diffusion method by mixing a $A1^{OB\Delta}$•A6•^{A1}Nb10 protein solution with an equal volume of reservoir solution. Crystals of $A1^{OB\Delta}$ with ^{A1}Nb10 protein were grown in three different conditions. Form I was obtained from 0.05 M ammonium sulfate, 0.05 M sodium acetate trihydrate pH 5.0, 30% w/v polyethylene glycol (PEG) 2000 MME, 0.2 M sodium thiocyanate, belonged to space group $P2_12_12$ and diffracted to 2.0 Å; Form II from 0.2 M magnesium chloride, 0.1 M bis–tris pH 5.5, 25% w/v PEG 3350, belonged to space group C2 and diffracted to 2.65 Å; and, Form III from 0.05 M ammonium sulfate, 0.05 M sodium acetate trihydrate pH 5.0, 30% w/v PEG 2000 MME, 5% v/v jeffamine M600, belonged to space group C222 and diffracted to 2.7 Å. Selenomethionyl derivative protein crystals in Form I were also obtained. Detailed information on mutagenesis, crystallization, structure determination and refinement, production of anti-$A1^{OB}$ nanobodies and RNA-binding assays is included in the Supplementary Methods. Crystallographic data collection and refinement statistics are shown in Supplementary Table S1.

RESULTS

Identification of the region of A1 involved in binding A6

A series of A1 truncation mutants was constructed and used to identify the region of A1 involved in binding to

A6. A1 truncation mutants were designed to avoid disruptions in predicted domains (27). Using pull-down assays, interactions of 10 different truncated A1 variants with His_6-tagged wild-type A6 were tested. A1 with all residues prior to amino acid 626 removed, i.e. $A1^{OB}$, is still capable of binding to A6 (Figure 1A and Supplementary Figure S1). In contrast, any A1 truncation without $A1^{OB}$ failed to interact with A6 suggesting that the OB-fold of A1 is the only A6-interacting domain of A1 (Supplementary Figure S2). In order to obtain direct evidence of interaction between the $A1^{OB}$ and wild-type A6, we co-expressed both protein domains in prey (without any tag) and bait (His_6-tag) position, respectively, resulting in a stable interaction in both cases (Figure 1B). The presence of His_6-tagged A6 and $A1^{OB}$ was also verified by TEV protease cleavage and anti-A1 immunopurification, respectively (Figure 1B, lane 5; Supplementary Figure S3). These findings indicate that $A1^{OB}$ is the domain from A1 responsible for the interaction with A6.

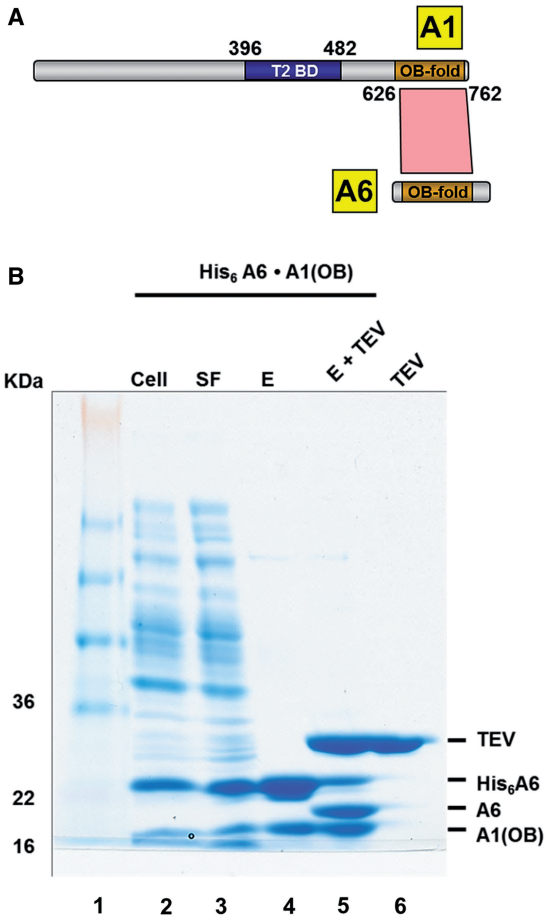

Figure 1. *T. brucei* $A1^{OB}$ and A6 interact with each other. (**A**) The T2-binding domain (T2BD) and OB-fold domain arrangement in A1. A schematic representation of full length A1 and full length A6 with their interaction domains is shown with the T2BD of A1 in blue and the OB folds of A1 and A6 in gold. Direct interactions, as identified by bacterial co-expression and purification, are shown as a red polygon. The T2-binding domain (T2BD) of A1 had been identified before (17). (**B**) Co-transformation and expression of $A1^{OB}$ and A6 constructs. *T. brucei* A6 containing a His_6-tag at the N-terminus co-transformed into *Escherichia coli* cells with $A1^{OB}$ (residues 626–762) was co-expressed. Cells were lysed and the His_6-tagged A6•$A1^{OB}$ complex was captured by incubating the lysate with Ni-NTA beads (lane 2–4). The His_6-tag peptide was cleaved off by TEV. This loss of the His_6-tag peptide is evident in the reduced size of the now untagged A6 (lane 5). Abbreviations used in this and in the following figures: Cell, total lysate from induced cells; SF, soluble fraction; E, Ni-NTA elution fractions; E+TEV, eluate from first Ni-NTA after treatment with TEV protease; TEV, tobacco etch virus (TEV) protease. The positions and sizes (kDa) of marker polypeptides are indicated on the left.

Three crystal structures of $A1^{OB\Delta}$ in complex with ^{A1}Nb10

In order to obtain the three-dimensional structure of $A1^{OB}$, we used several approaches in parallel such as replacing the non-conserved loop L23 by a 4-residue linker (Supplementary Figure S1). However, neither $A1^{OB}$ nor $A1^{OB\Delta}$ crystals could be obtained. Also complexes of $A1^{OB}$ or $A1^{OB\Delta}$ with A6 failed to give suitable crystals. Given these difficulties, anti-A1 nanobodies were generated and used as crystallization chaperones, a strategy which had previously been successfully used in crystallizing challenging proteins by our collaborating laboratories (16,29,48–50). An $A1^{OB\Delta}$•A6•^{A1}Nb10 ternary complex could be obtained (Supplementary Figure S3) indicating that ^{A1}Nb10 does not block A6 binding by $A1^{OB\Delta}$. Crystallization of the $A1^{OB\Delta}$•A6 in complex with ^{A1}Nb10 yielded three crystal forms (I, II and III). Surprisingly, in all three crystal forms, the asymmetric unit contained two copies of $A1^{OB\Delta}$ and two of ^{A1}Nb10 indicating that A6 dissociated during crystallization. The three structures were refined at 2.0, 2.7 and 2.65 Å resolution, with R_{free} values of 22.9, 26.4 and 26.7%, respectively (Supplementary Table S1). The structures of the six $A1^{OB\Delta}$ subunits in the three crystal forms superpose mutually with root-mean-square deviations 0.6 ~ 0.8 Å (Supplementary Figure S4). In the following, we therefore describe the structure of crystal Form I, since this is the structure with the highest resolution (Figure 2).

Although $A1^{OB}$ has a low level of sequence similarity with the previously determined crystal structures of editosome OB domains of A6 and $A3^{OB}$ (16,29), the overall structure of $A1^{OB\Delta}$ shares the conserved OB-fold architecture (Figure 2 and Supplementary Figure S5). Superimposing the $A1^{OB\Delta}$ monomer onto A6 (PDB-ID: 3K7U) results in a root-mean-square deviation of 1.5 Å for 80 equivalent Cα atoms with 29% amino acid sequence identity. The superposition shows the conserved core β-barrel composed of six β strands (Supplementary Figure S5). A significant difference is the absence of helix α1 in $A1^{OB\Delta}$ (Figure 2A and Supplementary Figure S5). No electron density was observed for the putative helix α1 linking β3 and β4 in any of the six $A1^{OB\Delta}$ subunits in our three crystal structures, suggesting that residues which form usually this helix in other OB folds, are flexible in $A1^{OB\Delta}$.

$A1^{OB\Delta}$ dimerization involves the side-by-side arrangement of β strands from the two subunits resulting in a contiguous six-stranded β-sheet with in the center an anti-parallel pair of N-terminal β1 strands

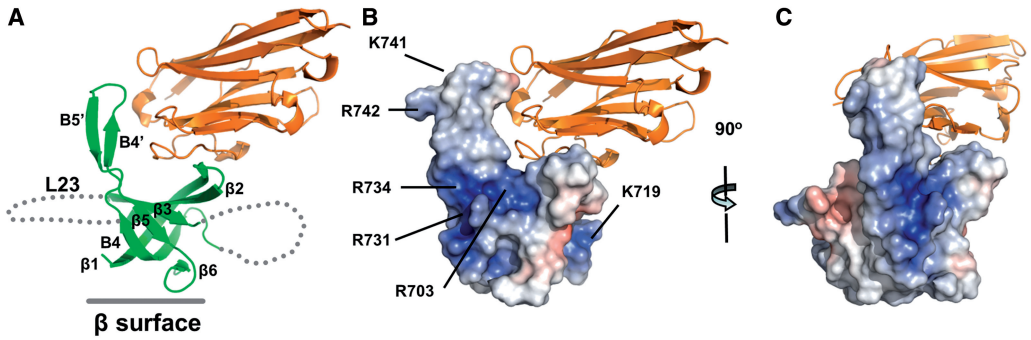

Figure 2. The crystal structure of *T. brucei* $A1^{OB\Delta}$ in complex with ^{A1}Nb10. (**A**) Overall structure of the $A1^{OB\Delta}$ •^{A1}Nb10 complex. The structure is shown as a ribbon diagram with the $A1^{OB\Delta}$ domain and ^{A1}Nb10 depicted in green and gold, respectively. Selected secondary structure elements are labeled. The approximate location of the β-surface is indicated with a solid line. Two disordered regions are shown with dashed lines. (**B**) Electrostatic surface representations for $A1^{OB\Delta}$. The $A1^{OB\Delta}$ domain is shown with its electrostatic surface charge and the nanobody ^{A1}Nb10 as a ribbon diagram. The electrostatic potential surface of $A1^{OB\Delta}$ was calculated using APBS (51). Regions with potentials above $+5\ k_bTe_c^{-1}$ and below $-5\ k_bTe_c^{-1}$ are shown in blue and red, respectively. The locations of several conserved basic residues are indicated. (**C**) A side view of the complex shows that the basic residues on the surface of the $A1^{OB\Delta}$ are facing away from nanobody binding site.

(Supplementary Figure S6). The three $A1^{OB\Delta}$ homodimers in the three new crystal structures form different dimers (Supplementary Figure S6), with the dimers of crystal Forms I and II being quite similar to each other but distinctly different from the dimer in crystal Form III. The latter is closer in structure to that of canonical OB-fold dimers, while there is a gap between the neighboring β1 strands from the two monomers in crystal Forms I and II. When superimposing one subunit of crystal Form I or II onto a subunit of Form III, it appears that the second subunit in Form I or II is rotated by 8.6° and shifted by 2 Å compared to the second subunits in Form III (Supplementary Figure S6B). Size exclusion chromatography studies reveal an equilibrium of $A1^{OB}$ monomers and dimers in solution (Supplementary Figure S7) which, combined with the different dimers in our crystal structures, suggests that there might be a mixture of monomers and multiple types of dimers in solution.

The charge distribution on the surface of $A1^{OB\Delta}$ reveals a distinctive feature on the surface of $A1^{OB\Delta}$: a large positively charged region with contributing charged side chains mainly from strands β3, β4, β4′ and β5′ (Figure 2 and Supplementary Figure S1). Interestingly, seven conserved basic residues are responsible for this highly positively charged surface. An alignment of the primary structures of $A1^{OB}$ domains from various species reveals that the R703, K715, R731, R734 and R742 side chains are identical in all of the A1 proteins from these species (Supplementary Figure S1). Two additional residues, K719 and K741 of *T. brucei* are either a Lys or and Arg in all trypanosomes species. This conserved positively charged region could play a critical role in the recognition of negatively charged dsRNA substrates or other editosome proteins, as will be further investigated below.

Identification of interacting domains of A1 and L2

To identify which domains of A1 and L2 interact with each other, truncated forms of A1 and L2 were constructed and analyzed in pull-down assays after co-expression as described above. For this purpose, we generated expression systems for only the NTase domain of L2, comprising residues 21–285, and various N-terminal truncations of L2 with the NTase domain removed. We also constructed various non-tagged domain variants of A1. Co-expression and Ni-NTA pull-down experiments of untagged A1 variants with His_6-tagged L2 variants showed interesting results (Figure 3B). The C-terminal region of L2, spanning residues 308–416, was capable of interacting with the N-terminal region (residues 196–331) of A1 (Figure 3B). No interactions between A1 (residues 335–762) and full length L2 (see Supplementary Figure S8), and neither between full length A1 and the $L2^{NTase}$ domain were observed (data not shown). These results suggest that amino acids 196–331 of A1 and amino acids 308–416 of L2 are functionally important for the interaction of A1 and L2. Hence, residues 308–416 of L2 are hereafter called the A1-binding domain or $L2^{A1BD}$, and residues 196–331 of A1 the L2-binding domain or $A1^{L2BD}$ (Figure 3A).

A1 promotes dsRNA binding by the RNA-editing ligase L2

Based on A1's association with L2, we hypothesized that domains from both A1 and L2 might be required to bind a nicked gRNA:mRNA duplex in the RNA ligation process. To test this hypothesis, a native gel mobility shift assay was used to examine the binding the nicked dsRNA substrates to length variants of L2 and A1. Binding reactions were performed in the absence of ATP so as to preclude conversion of substrate to product during the incubation.

To determine the role of different parts of L2 in nicked dsRNA recognition, two length variants of L2 were made and co-expressed with full length A1. The first L2 variant retains residues 21–416 (i.e. comprises the NTase and A1BD domains) of L2, while the second contained only the A1BD of L2. Full length L2 in complex with full length A1 readily binds dsRNA, but the $L2^{A1BD}$ alone

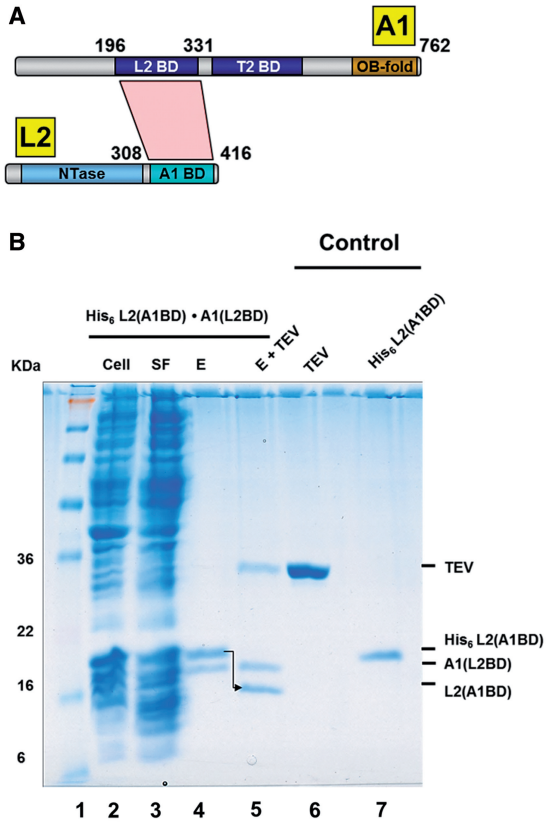

Figure 3. Identification of interaction domains of *T. brucei* A1 and L2. (**A**) $A1^{L2BD}$ and $L2^{A1BD}$ interact with each other. A schematic representation of the domains of full length A1 and full length L2, i.e. the L2-binding domain (L2BD), the T2-binding domain (T2BD), and the OB-fold domain (OB-Fold) of A1, and the nucleotidyl-transferase domain (NTase) and the A1-binding domain (A1BD) of L2. The direct interaction between the L2BD and the A1BD, as identified by bacterial co-expression and purification (see Figure 3B), is shown as a red polygon. (**B**) Co-transformation and co-expression of $A1^{L2BD}$ and $L2^{A1BD}$ constructs. *T. brucei* $A1^{L2BD}$ (residues 196–331) and $L2^{A1BD}$ (residues 308–416) were co-expressed in *E. coli* and co-purified by Ni-NTA chromatography via the His_6-tag of L2 (lanes 2–4). The His_6-tag of L2 was removed with TEV protease. This loss of the His_6-tag of L2 is indicated by an arrow (lane 5). The positions and sizes (kDa) of marker polypeptides are indicated on the left (lane 1). See Figure 1B for abbreviations.

in complex with full length A1 does not interact with dsRNA (Figure 4A). Since the L2 NTase domain alone does also not interact with dsRNA (Figure 4B), this indicates that both the NTase domain and parts of A1 are essential for RNA binding by the L2•A1 complex.

To determine the roles of different parts of A1 in interacting with L2 and RNA, three different domain variants of A1 were made, $A1^{(L2BD+T2BD+OB)}$, $A1^{(L2BD+T2BD)}$ and $A1^{(L2BD)}$. Subsequently, these domain variants were co-expressed with full length L2, and tested for dsRNA binding (Figure 4C). $A1^{(L2BD+T2BD+OB)}$ in complex with full length L2 did bind to dsRNA, but full length L2 in complex with $A1^{(L2BD+T2BD)}$ without the C-terminal OB-fold domain did not. In addition, $A1^{L2BD}$ in complex with full length L2 lost overall RNA-binding capability. These results suggest that the OB-fold domain of A1 is essential for dsRNA binding by the A1•L2 complex.

Interestingly, the $A1^{OB}$ domain by itself failed to bind tightly to nicked dsRNA (Figure 4D). As expected, the $L2^{A1BD}$, $A1^{L2BD}$ and $A1^{T2BD}$ domains individually did also not display any RNA-binding capability (data not shown). These observations suggest that the $A1^{OB}$ domain and the $L2^{NTase}$ domain bind nicked dsRNA in a cooperative manner.

Discovering critical residues of $A1^{OB}$ in dsRNA binding by the complex of the RNA ligase L2 and A1

We next asked which area of $A1^{OB}$ assists L2 in dsRNA binding. We first tested the possible function in RNA binding of the flexible L23 loop of $A1^{OB}$ which was deleted in our constructs in order to obtain diffraction quality crystals (Figure 5A and Supplementary Figure S1). Interestingly, the removal of the 38 amino acid residues of L23 from A1 did not affect its ability to bind dsRNA (Figure 5B).

In the structure of $A1^{OB\Delta}$ in complex with ^{A1}Nb10, ~800 $Å^2$ solvent accessible area of $A1^{OB\Delta}$ (13% of its total accessible area) is buried by the ^{A1}Nb10 nanobody (Figure 5A). It was of interest to determine whether $A1^{OB}$ residues involved in nanobody binding had any RNA recognition function *in vitro*. Figure 5C shows that the A1•L2•^{A1}Nb10 complex was able to bind dsRNA, i.e. the nanobody did not prevent A1•L2's ability to bind dsRNA, and hence the surface of $A1^{OB}$ occupied by ^{A1}Nb10 is not playing a significant role in RNA binding by the A1•L2 heterodimer.

We next examined the role of the strictly conserved, solvent exposed basic residues found in a closely spaced cluster of positively charged residues, referred to as 'the arginine cluster' (Figures 5A and 6). To probe whether any of the conserved side chains in this region are functionally relevant, each of the seven basic residues R703, K715, K719, R731, R734, K741 and R742 of $A1^{OB}$ were replaced by glutamate, thereby creating the potential for electrostatic repulsion between the OB-fold domain and the RNA phosphodiester backbone. Figure 5D shows that three A1 mutants, R703E, R731E and R734E completely lost their RNA-binding property, while no effect on RNA binding by other single or double mutants was observed. These results indicate that the cluster of three arginines formed by R703, R731 and R734 of $A1^{OB}$ plays a crucial role in the RNA-binding process by the A1•L2 complex.

To investigate whether the A1•L2 complex had any preference for dsRNA over dsDNA substrates, we compared the binding of the A1•L2 enzyme complex with both nicked RNA duplex substrates and nicked DNA duplex substrates. The A1•L2 complex has a high affinity for various dsRNA substrates. However, A1•L2 complex does not have a similar high affinity for dsDNA (Figure 5E).

DISCUSSION

The studies described in this paper present major progress in understanding properties of domains from two key editosome proteins: the largest interaction protein

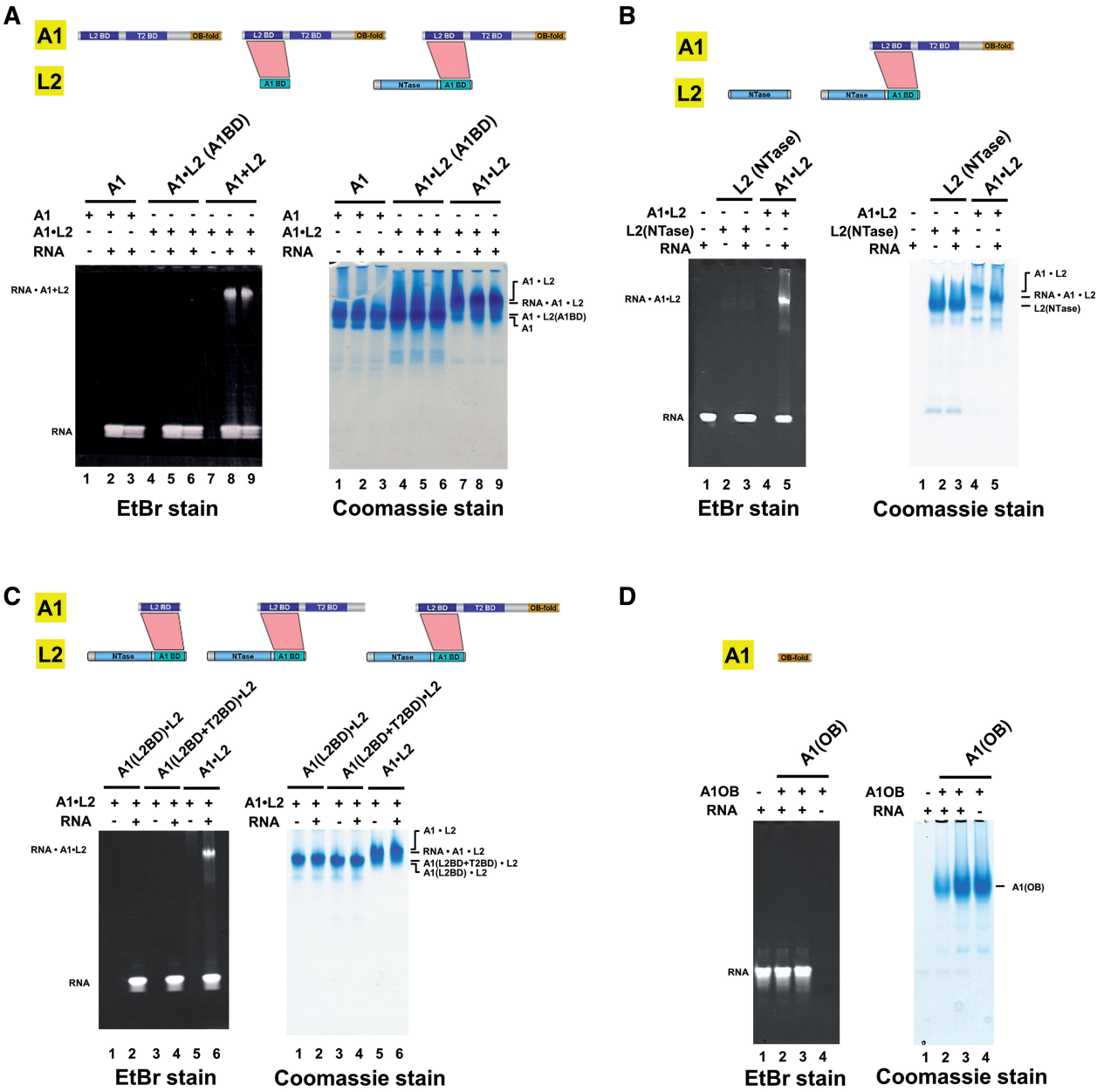

Figure 4. The C-terminal domain of *T. brucei* A1 promotes dsRNA binding by the RNA-editing ligase L2. Electrophoretic mobility shift assays were used to examine the binding of A1•L2 to nicked dsRNA substrate. Wild type and variants of A1•L2 binary complexes were co-expressed and co-purified as explained in Supplementary Methods. A total of 30 μM double-stranded 26-mer nicked RNA (see the sequence in Supplementary Table S2) was incubated with 15 μM proteins or protein complexes for 30 min at 4°C in 20 mM Tris (pH 7.5), 2 mM DTT and 250 mM NaCl. The RNA and A1•L2 complex were resolved on a 4–15% polyacrylamide gel and stained by ethidium bromide and Coomassie Blue, shown, respectively, at the left and the right of A–D. Each protein and the A1•L2 constructs used for the RNA-binding assays are illustrated in a schematic diagram above the gels. The positions of the RNA•A1•L2 complex and of RNA alone are indicated on the left. (**A**) The role of different domains of L2 in dsRNA recognition. The RNA binding abilities of (i) full length A1 (lanes 1–3) by itself, (ii) $L2^{A1BD}$ in complex with full length A1 (lanes 4–6) and (iii) full length $L2^{(NTase+A1BD)}$ with full length A1 (lanes 7–9), show that the NTase domain of L2 is critical for nicked dsRNA recognition. (**B**) The role of the $L2^{NTase}$ domain in dsRNA recognition. The RNA binding abilities of $L2^{NTase}$ domain by itself (lanes 2–3) and of wild-type full length A1 in complex with full length L2 (lanes 4–5), show that $L2^{NTase}$ alone fails to bind tightly to nicked dsRNA. Lane 1 contains 26-mer nicked dsRNA alone as a control. (**C**) The role of different domains of A1 in dsRNA recognition. The RNA binding abilities of (i) $A1^{L2BD}$ with full length L2 (lanes 1–2), (ii) $A1^{(L2BD+T2BD)}$ with full length L2 (lanes 3–4) and (iii) $A1^{(L2BD+T2BD+OB)}$ with full length L2 (lanes 5–6), show that the $A1^{OB}$ domain promotes RNA binding by L2. (**D**) The $A1^{OB}$ domain alone has little affinity for dsRNA. The RNA binding abilities of 15 and 45 μM $A1^{OB}$ alone (lanes 2–3) were tested. Lane 1 contains 26-mer nicked dsRNA alone as a control.

A1 and the RNA-editing ligase L2. In particular, the multiple functions of the OB-fold of A1 are most intriguing including the role it plays in the action of the RNA-editing ligase L2. Our new findings can be well-incorporated into the recently proposed model of an OB-fold center in the core of the editosome (29). It is remarkable that nanobodies were again required to obtain well-diffracting crystals illustrating once again the power

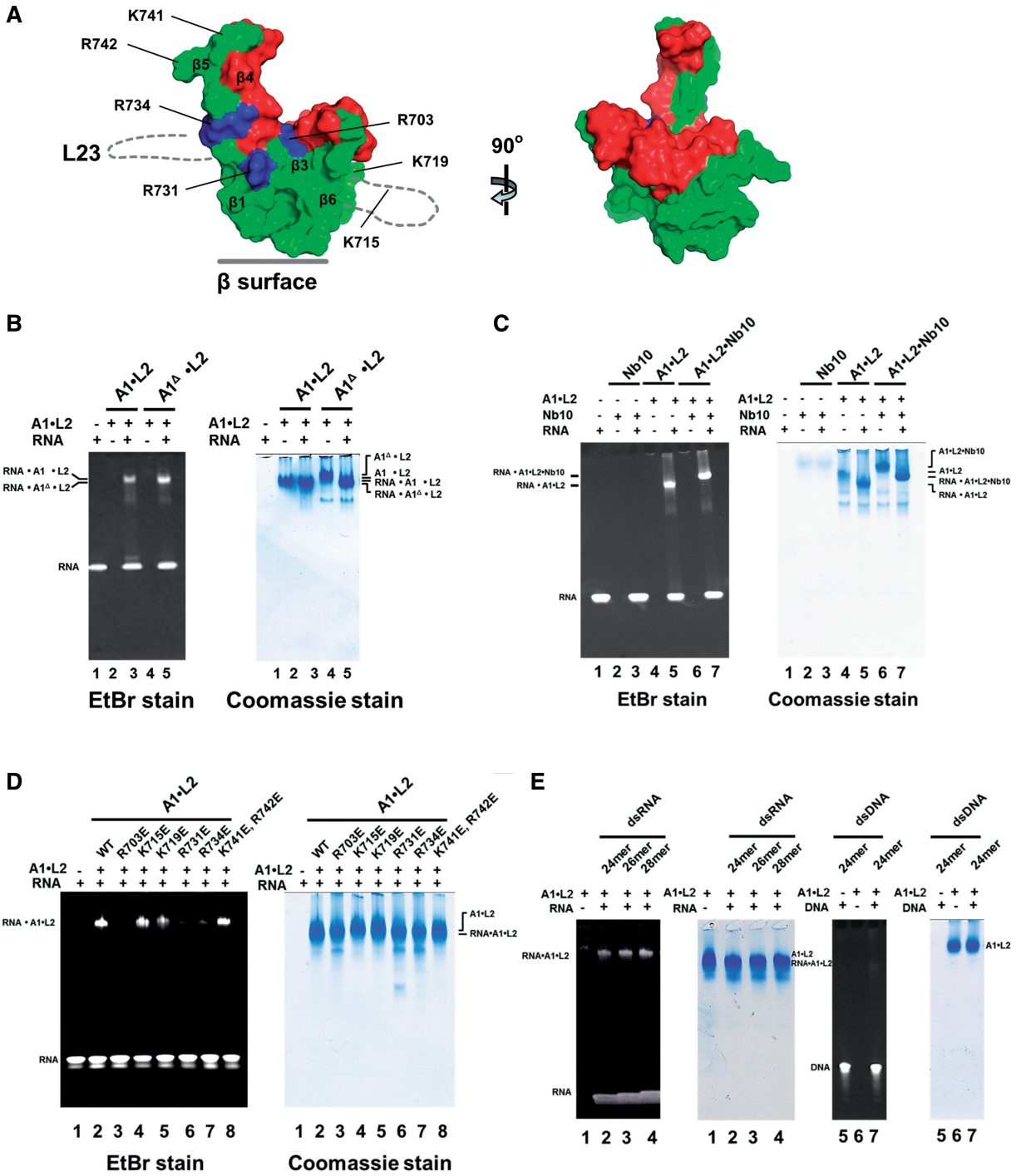

Figure 5. Identification of specific arginine residues involved in dsRNA binding by the *T. brucei* $A1^{OB}$ domain in complex with the RNA-editing ligase L2. (**A**) Surface presentation of the $A1^{OB\Delta}$ structure. Left: the structure of $A1^{OB\Delta}$, with ^{A1}Nb10 omitted for clarity, is shown with (i) the residues involved in binding of ^{A1}Nb10 in red, (ii) the completely conserved and solvent exposed basic residues R703, R731 and R734 in blue and (iii) the remainder of the $A1^{OB\Delta}$ surface in green. The approximate locations of the two disordered regions are sketched with dotted lines. Right: the $A1^{OB\Delta}$ structure rotated ~90° with respect to the left panel illustrating the ^{A1}Nb10-binding surface of $A1^{OB}$ which is not engaged in RNA-binding (Figure 5C). (**B–E**) Wild-type and variants of A1•L2 binary complexes were co-expressed and co-purified. Electrophoretic mobility shift assays—as explained in the Legend to Figure 4—were used to evaluate the RNA-binding capability of the set of A1•L2 variants obtained. (B) The effect of the deletion of L23 on RNA affinity of $A1^{OB}$. The RNA binding abilities of wild-type A1•L2 (lanes 2 and 3) and $A1^{\Delta}$•L2 (lanes 4 and 5) are compared. Lane 1 contains 26-mer nicked dsRNA alone as a control. (C) The effect of ^{A1}Nb10 binding on RNA-binding activity. The RNA binding abilities of ^{A1}Nb10 alone (lanes 2 and 3), A1•L2 (lanes 4 and 5), and A1•L2• ^{A1}Nb10 (lanes 6 and 7) were compared. Lanes 1 contains 26-mer nicked dsRNA alone as a control. (D) Mutational effects on RNA-binding activity. The RNA-binding assay was carried out with single or double mutants of A1 with wild-type L2. Each of seven basic residues (R703, K715, K719, R731, R734, K741 and R742) in full length A1 was replaced by glutamate as indicated (lanes 3–8). The RNA binding abilities of wild-type and mutant A1•L2 were tested. 26-mer nicked dsRNA alone (lane 1) and wild-type A1•L2 (lane 2) were used as a control. (E) The preference for dsRNA over dsDNA substrates on RNA-binding activity. The binding of both nicked RNA duplex substrates and nicked DNA duplex substrates by recombinant A1•L2 was tested. To probe the possibility of sequence specificity on dsDNA recognition, two different random sequence dsDNA were also tested as a control and appeared not to bind to A1•L2 (data not shown).

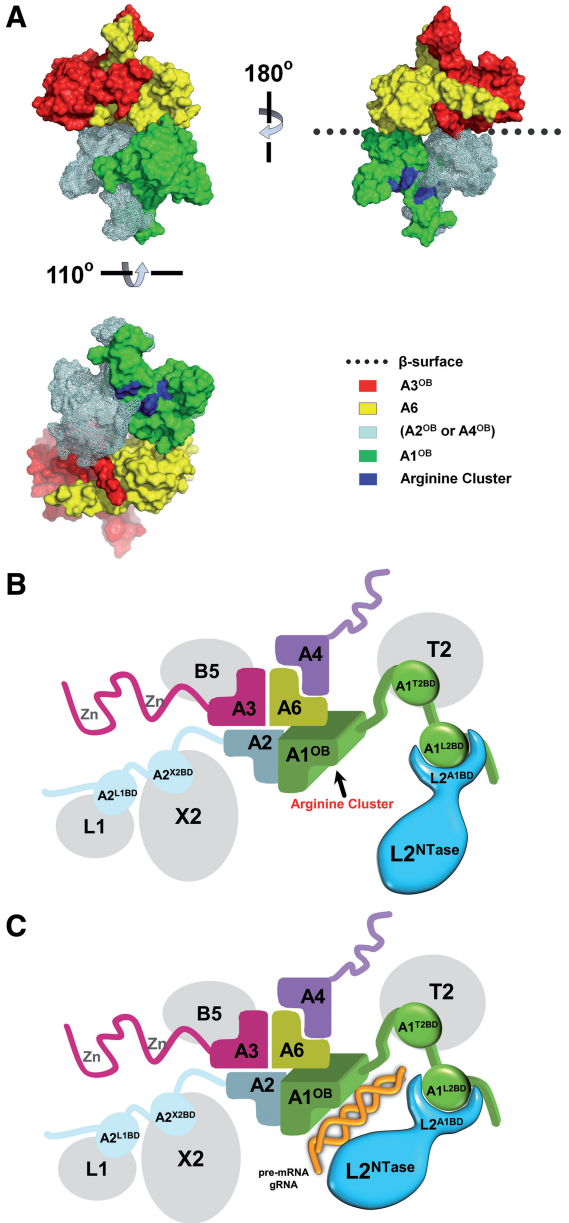

Figure 6. Model of the OB-fold center and associated proteins in the core of the editosome. (**A**) A 'shifted heterotetramer' model of the OB folds of A1, A2 or A4, A3 and A6. The shifted heterotetramer described in (29) was modified by replacing the model structure for $A1^{OB}$ by the actual structure of $A1^{OB}$ described in this article. The heterotetramer depicted contains the actual structure of the $A3^{OB}$•A6 heterodimer (29), the new structure of $A1^{OB}$ (Figure 5A and Supplementary Figure S4) and a model of $A2^{OB}$ based on the structure of $A1^{OB}$. The $A2^{OB}$ domain might be an $A4^{OB}$ domain instead, as discussed in (29). (**B**) The 'shifted heterotetramer' model with four associated enzymes depicted schematically. The OB domains are colored as in (A); the editing RNA ligase L2 is shown in blue; four additional editosome proteins in gray. The three domains of A1 are schematically shown interacting with partner proteins in the core: the $A1^{L2BD}$ with the $L2^{A1BD}$, $A1^{T2BD}$ domain with T2 and $A1^{OB}$ with A6. The editosome proteins schematically depicted are: TUTase 2 (T2), Exonuclease 2 (X2), Ligase 1 (L1), Ligase 2 (L2), interaction proteins A1 to A6 (A1–A6) and KREPB5 (B5). Interaction data is from references (15,17) and from Figures 1, 3 and 4. (**C**). A1 and L2 recognizing dsRNA. Color code as in (B) above with dsRNA schematically added in gold, and with the $L2^{NTase}$ domain closer to $A1^{OB}$. The $A1^{L2BD}$ and $L2^{A1BD}$ domains interact with each other, thereby enabling the $A1^{OB}$ and $L2^{NTase}$ domains to cooperatively bind dsRNA substrate. For further explanation see 'Discussion' section.

of these crystallization chaperones which had been successful in many other studies in our laboratories (16,29,48–50).

Functions of editosome protein domains unraveled

By expressing, purifying and analyzing a large number of variants of the interaction protein A1 and the RNA-editing ligase L2, and complexes of the two, it has become clear that:

(i) The OB fold of A1 is engaged in an interaction with A6 (Figure 1). This result is consistent with previous results where it was shown that A6 forms a complex with both $A1^{(L2BD+T2BD+OB)}$ and $A1^{(T2BD+OB)}$ (17). The A6•$A1^{OB}$ interaction is probably not very strong since, in three different crystal forms obtained under mild conditions, the A6 chain is absent and only $A1^{OB\Delta}$ in complex with nanobody ^{A1}Nb10 is observed, while the crystallization experiments started with a solution containing a ternary A6•$A1^{OB\Delta}$•^{A1}Nb10 complex (Supplementary Figure S3).

(ii) The region surrounding the putative first zinc-binding site of A1, spanning residues 196–331, is responsible for interacting with the RNA Ligase L2, and therefore called $A1^{L2BD}$ (Figure 3).

(iii) the C-terminal domain of L2, and comprising residues 308–416, is involved in associating L2 with A1, and is therefore called $L2^{A1BD}$ (Figure 3). The interactions of the $A1^{L2BD}$ and the $L2^{A1BD}$ are key to making it possible for $A1^{OB}$ to assist the catalytic domain of L2 ($L2^{NTase}$) in performing its function discussed as follows.

Since it has been shown (17) that the region surrounding the second putative zinc-binding site of A1, comprising residues 396–482, is involved in binding to the enzyme T2 and is called the $A1^{T2BD}$, there are now three well-defined regions of the large A1 protein responsible for contacts with three different editosome proteins (Figure 1A). The name 'interaction protein' for A1 appears therefore to be very well chosen. It is also of interest to consider the global similarities in the domain structures of A1 and A2 (27), and the fact that full length A2 has been shown to bind to the editing RNA Ligase L1 and the editosomal exonuclease X2 (17,20). Given the similarities in domain structure of the two large interaction proteins A1 and A2, combined with the fact that the C-terminal domain of A2 ($A2^{OB}$) interacts with A6 [Figures 2A, 4 A and B in (17)], it is tempting to suggest that the region in A2 corresponding to the L2BD of A1 is engaged in interactions with L1, and that the region in A1 corresponding to the T2BD of A2 is engaged in binding to X2.

In other words, the order of domains with known functions in A1 is L2BD–T2BD–OB and in A2 the order is presumably L1BD–X2BD–OB, where the X2-binding function of the putative X2BD of A2 still has to be demonstrated but the role of the L1BD of A2 and the interactions of the OB fold of A2 with A6 have been shown before (17). Given the similarities of A1 and

A2, and of L1 and L2, it is likely that the C-terminal domain of L1 is responsible for the interactions of L1 with A2, specifically through interactions with the $A2^{L1BD}$ surrounding the first putative zinc-binding region of A2 (17,20).

Three crystal structures of the OB fold of the interaction protein A1

The structure of the $A1^{OB\Delta}$ monomer appears to have, in all three crystal forms we studied, a number of special characteristics. The residues expected to form the canonical helix α1 between strands β3 and β4 is not represented by well-defined electron density in any of the six independent copies of $A1^{OB\Delta}$ in the three crystal structures determined. This helix has been observed to be well defined in over 10 other OB-fold structures reported in the literature (Supplementary Figure S5), and its flexibility in $A1^{OB\Delta}$ is hence highly unusual, but might be partly due to a conserved proline residue in this region (Supplementary Figure S1). The significance of this flexibility is unknown at this moment but it might be that this region only becomes better defined in structure when $A1^{OB}$ interacts with other, yet to be discovered, proteins in the editosome.

While the structures of the six $A1^{OB\Delta}$ monomers are very similar, the three dimers in our structures are different which is highly unusual among OB-fold dimers. The dimers in crystal Form I and II are deviating substantially from the dimer in crystal Form III, with a 2 Å shift perpendicular to the interacting β1 and β1′ strands, and a 9° difference in orientation of the two subunits in the dimers (Supplementary Figure S6), and a concomitant loss of several β1–β1′ main chain hydrogen bonds. However, the interface between OB folds in a dimer is extensive (29), hence the $A1^{OB\Delta}$ dimers are stabilized by the interactions in other regions, e.g. by interfaces involving strands β3, β4 and β5. In all three dimers, $A1^{OB\Delta}$ interacts with ^{A1}Nb10 in essentially the same manner, hence nanobody binding is not responsible for the differences in $A1^{OB\Delta}$ dimer organization in the three crystal Forms.

Solution studies provide evidence for a monomer–dimer equilibrium of the $A1^{OB}$ domain (Supplementary Figure S7), which, together with the variability of the dimer structures observed in the three crystal structures, suggests that the $A1^{OB}$ dimer can easily be disrupted. This would be in agreement with its participation in the OB-fold center of the editosome where only one $A1^{OB}$ domain is expected to be present since, as a very large interaction protein, A1 is most likely only represented by a single copy in the editosome (29,52,53).

The OB fold of A1 assists RNA binding by RNA-editing ligase L2 *in trans*

With the increased knowledge of the functions of domains in the A1•L2 complex, and the new structures of $A1^{OB\Delta}$, we carried out several further studies of this complex to unravel features essential for RNA binding. It appeared that neither the $A1^{OB}$ nor the $L2^{NTase}$ domains by themselves bound nicked dsRNA, and nor did the complex of full length L2 with a variant of A1 with the $A1^{OB}$ removed (Figure 4). Hence, $A1^{OB}$ is essential for dsRNA binding by the A1•L2 complex. In DNA ligases and RNA capping enzymes known so far, an OB fold from these enzymes is involved in nucleic acid binding *in cis* (35,37,39,42). It has also been suggested that the C-terminal non-OB-fold domain of bacteria phage RNA ligases is required for nick recognition and ligation (54,55). This C-terminal domain of the phage RNA ligase then also promotes RNA binding *in cis*. In contrast, the OB fold of A1 is able to assist RNA binding by L2, due to interactions between the $L2^{A1BD}$ and the $A1^{L2BD}$ which bring A1 and L2 together. Flexible regions between domains of both proteins may bring the catalytic domain of L2 and the OB fold of A1 in close proximity to enable RNA binding by the A1•L2 complex. To the best of our knowledge, this is the first time that an OB fold has observed to increase the affinity for RNA by a member of the nucleotidyl-transferase family via an *in trans* mechanism.

Additional experiments provided important information which part of $A1^{OB}$ is crucial for its function in RNA binding by the A1•L2 complex. For example, the ternary A1•L2•^{A1}Nb10 complex did interact with nicked dsRNA with a similar affinity as the binary A1•L2 complex and, hence, the surface area of $A1^{OB}$ used for ^{A1}Nb10 binding is not critical for the interaction of A1•L2 with RNA (Figure 5C). It also appeared that A1•L2 and $A1^{\Delta}$•L2 exhibited a similar affinity for nicked dsRNA binding and therefore loop L23 is not a key element of A1 for engaging RNA (Figure 5B). Next, the possible role of a highly charged region on the $A1^{OB}$ surface was probed by mutating each of seven Arg residues into glutamates. Substituting a glutamate for three of these seven arginines caused loss of nicked dsRNA binding by the A1•L2 complex (Figure 5D). This established that a highly positively charged surface region is essential for RNA binding by $A1^{OB}$. It is very likely that also in the assembled editosome the $A1^{L2BD}$ plays a critical role to by bringing together the editing RNA ligase L2 and the interaction protein A1, and that the same positively charged residues of $A1^{OB}$ enhance RNA binding by the ligase L2 *in trans*.

The OB center in the core of the editosome

It is of interest to combine A6 and RNA-binding functions of $A1^{OB}$ with the proposed 'shifted heterotetramer' of the OB-fold center in the core of the editosome (29). When taking the model of the shifted heterotetramer proposed by these authors, and replacing the modeled $A1^{OB}$ fold in that tetramer by the $A1^{OB}$ structure determined in this study, we obtain a heterotetramer based on the experimentally observed $A3^{OB}$•A6 heterodimer and the new structure of $A1^{OB}$ (Figure 6A). In this heterotetramer, where the fourth OB fold can either be that of A2 or of A4 [for discussion, see (29)], the following features of $A1^{OB}$ with respect to the other three OB folds are of interest as follows.

(i) The flexible region between β3 and β4, which is usually helical in OB folds as mentioned above, is not in contact with any of the three other OB folds

in the heterotetramer. This highly conserved (Supplementary Figure S1) region could be of importance in the assembled editosome to interact with other editosome proteins.

(ii) The region most likely occupied by loop L23, which was deleted in our studies in order to obtain diffraction quality crystals, is also accessible to solvent in the shifted heterotetramer (Figure 6A). Loop L23 is not conserved in trypanosomatids (Supplementary Figure S1) and hence its role in editosome function, if any, will require further study.

(iii) The highly positively charged region including the three arginines R703, R731 and R734, demonstrated to be involved in RNA binding (Figure 5D), is fully accessible to solvent.

The highly positively charged region on $A1^{OB}$ is very likely interacting with nicked dsRNA in the A1•L2 binary complex (Figure 5E), and most probably the same will happen, at least transiently, in the core of the editosome during the RNA-editing process. Consistent with this model with its exposed positively charged surface region of $A1^{OB}$, the A1•A6•L2 ternary complex did bind dsRNA (data not shown).

To illustrate, more clearly, the possible role of the OB fold center, additional domains of A1, A2 and A3 have been added schematically to the shifted heterotetramer of OB folds (Figure 6B). In the absence of RNA, the $A1^{OB}$ and the $L2^{NTase}$ domains might be at a distance from each other, but the presence of nicked dsRNA may induce at a certain moment a conformational change in L2, and possibly also in the linkers between the different domains of A1, and bring the $A1^{OB}$ and the $L2^{NTase}$ closer together to enhance RNA binding and to enable L2 to carry out the ligation reaction (Figure 6B and C). This results in a sealed mRNA strand which is subject to subsequent editing steps, except of course when this is the final editing reaction for this mRNA-editing site.

Obviously, major questions still remain to be answered regarding the mechanism of action and the motions in the RNA-loaded editosome. Yet, our studies have provided new insights in the mechanism of RNA editing by the editosome, in particular by identifying (i) the domains of A1 and L2 which interact with each other and (ii) a critically important region for RNA binding on the surface of the C-terminal OB fold of A1, which acts *in trans* to enhance RNA binding by the editing ligase L2.

ACCESSION NUMBERS

Coordinates and structure factors for the crystals of $A1^{OB\Delta}$ •^{A1}Nb10 have been deposited with the Protein Data Back under accession codes 4DK3, 4DK6 and 4DKA.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2, Supplementary Figures 1–9, Supplementary Methods, and Supplementary References [16,17,29,48,50,51,56–68].

ACKNOWLEDGEMENTS

We thank the staff of BL8.2.1, BL5.0.1, and BL8.2.2 beam line at ALS for invaluable assistance with data collection. We also thank Stewart Turley for scientific discussions and technical support; Jonathan Kay for maintaining the computing environment of the Biomolecular Structure Center; and Jungpeng Deng for contributions during the early stages of the project.

FUNDING

National Institute of Health [RO1 GM077418 and RO1 GM077418-04S1 to W.G.J.H.]. Vrije Universiteit Brussel, JS's laboratory [VUB, GOA65], the Vlaams Instituut Biotechnologie (VIB), the Fund for Scientific Research of Flanders (FWOAL551), the Hercules Foundation (HERC2); Institute for the encouragement of Scientific Research and Innovation of Brussels (ISRIB). Financial support was provided by the Belgian Government under the framework of the Interuniversity Attraction Poles [I.A.P. P6/19 to E.P.]. Funding for open access charge: National Institute of Health [RO1 GM077418, RO1 GM077418-04S1] and Interuniversity Attraction Poles [I.A.P. P6/19].

Conflicts of interest statement. None declared.

REFERENCES

1. Schnaufer,A., Domingo,G.J. and Stuart,K. (2002) Natural and induced dyskinetoplastic trypanosomatids: how to live without mitochondrial DNA. *Int. J. Parasitol.*, **32**, 1071–1084.
2. Vickerman,K. (1965) Polymorphism and mitochondrial activity in sleeping sickness trypanosomes. *Nature*, **208**, 762–766.
3. Koslowsky,D.J., Riley,G.R., Feagin,J.E. and Stuart,K. (1992) Guide RNAs for transcripts with developmentally regulated RNA editing are present in both life cycle stages of Trypanosoma brucei. *Mol. Cell Biol.*, **12**, 2043–2049.
4. Stuart,K. (1991) RNA editing in trypanosomatid mitochondria. *Annu. Rev. Microbiol.*, **45**, 327–344.
5. Aphasizhev,R. and Aphasizheva,I. (2011) Mitochondrial RNA processing in trypanosomes. *Res. Microbiol.*, **162**, 655–663.
6. Hajduk,S. and Ochsenreiter,T. (2010) RNA editing in kinetoplastids. *RNA Biol.*, **7**, 229–236.
7. Kable,M.L., Seiwert,S.D., Heidmann,S. and Stuart,K. (1996) RNA editing: a mechanism for gRNA-specified uridylate insertion into precursor mRNA. *Science*, **273**, 1189–1195.
8. Koslowsky,D.J., Bhat,G.J., Read,L.K. and Stuart,K. (1991) Cycles of progressive realignment of gRNA with mRNA in RNA editing. *Cell*, **67**, 537–546.
9. Sturm,N.R. and Simpson,L. (1990) Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome oxidase subunit III mRNA. *Cell*, **61**, 879–884.
10. Madison-Antenucci,S., Grams,J. and Hajduk,S.L. (2002) Editing machines: the complexities of trypanosome RNA editing. *Cell*, **108**, 435–438.
11. Stuart,K.D., Schnaufer,A., Ernst,N.L. and Panigrahi,A.K. (2005) Complex management: RNA editing in trypanosomes. *Trends Biochem. Sci.*, **30**, 97–105.
12. Simpson,L., Sbicego,S. and Aphasizhev,R. (2003) Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business. *RNA*, **9**, 265–276.
13. Panigrahi,A.K., Ernst,N.L., Domingo,G.J., Fleck,M., Salavati,R. and Stuart,K.D. (2006) Compositionally and functionally distinct editosomes in Trypanosoma brucei. *RNA*, **12**, 1038–1049.

14. Carnes,J., Soares,C.Z., Wickham,C. and Stuart,K. (2011) Endonuclease associations with three distinct editosomes in Trypanosoma brucei. *J. Biol. Chem.*, **286**, 19320–19330.
15. 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.
16. Wu,M., Park,Y.J., Pardon,E., Turley,S., Hayhurst,A., Deng,J., Steyaert,J. and Hol,W.G. (2011) Structures of a key interaction protein from the Trypanosoma brucei editosome in complex with single domain antibodies. *J. Struct. Biol.*, **174**, 124–136.
17. Schnaufer,A., Wu,M., Park,Y.J., Nakai,T., Deng,J., Proff,R., Hol,W.G. and Stuart,K.D. (2010) A protein-protein interaction map of trypanosome ~20S editosomes. *J. Biol. Chem.*, **285**, 5282–5295.
18. Brecht,M., Niemann,M., Schluter,E., Muller,U.F., Stuart,K. and Goringer,H.U. (2005) TbMP42, a protein component of the RNA editing complex in African trypanosomes, has endo-exoribonuclease activity. *Mol. Cell*, **17**, 621–630.
19. Drozdz,M., Palazzo,S.S., Salavati,R., O'Rear,J., Clayton,C. and Stuart,K. (2002) TbMP81 is required for RNA editing in Trypanosoma brucei. *EMBO J.*, **21**, 1791–1799.
20. Kang,X., Falick,A.M., Nelson,R.E., Gao,G., Rogers,K., Aphasizhev,R. and Simpson,L. (2004) Disruption of the zinc finger motifs in the Leishmania tarentolae LC-4 (=TbMP63) L-complex editing protein affects the stability of the L-complex. *J. Biol. Chem.*, **279**, 3893–3899.
21. Law,J.A., O'Hearn,S. and Sollner-Webb,B. (2007) In Trypanosoma brucei RNA editing, TbMP18 (band VII) is critical for editosome integrity and for both insertional and deletional cleavages. *Mol. Cell Biol.*, **27**, 777–787.
22. Salavati,R., Ernst,N.L., O'Rear,J., Gilliam,T., Tarun,S. Jr and Stuart,K. (2006) KREPA4, an RNA binding protein essential for editosome integrity and survival of Trypanosoma brucei. *RNA*, **12**, 819–831.
23. Huang,C.E., O'Hearn,S.F. and Sollner-Webb,B. (2002) Assembly and function of the RNA editing complex in Trypanosoma brucei requires band III protein. *Mol. Cell. Biol.*, **22**, 3194–3203.
24. Aphasizhev,R., Sbicego,S., Peris,M., Jang,S., Aphasizheva,I., Simpson,A., Rivlin,A. and Simpson,L. (2002) Trypanosome mitochondrial 3′ terminal uridylyl transferase (TUTase): the key enzyme in U-insertion/deletion RNA editing. *Cell*, **108**, 637–648.
25. Schnaufer,A., Panigrahi,A.K., Panicucci,B., Igo,R.P. Jr, 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.
26. Tarun,S.Z. Jr, Schnaufer,A., Ernst,N.L., Proff,R., Deng,J., Hol,W. and Stuart,K. (2008) KREPA6 is an RNA-binding protein essential for editosome integrity and survival of Trypanosoma brucei. *RNA*, **14**, 347–358.
27. Worthey,E.A., Schnaufer,A., Mian,I.S., Stuart,K. and Salavati,R. (2003) Comparative analysis of editosome proteins in trypanosomatids. *Nucleic Acids Res.*, **31**, 6392–6408.
28. Panigrahi,A.K., Schnaufer,A., Carmean,N., Igo,R.P. Jr, Gygi,S.P., Ernst,N.L., Palazzo,S.S., Weston,D.S., Aebersold,R., Salavati,R. *et al.* (2001) Four related proteins of the Trypanosoma brucei RNA editing complex. *Mol. Cell. Biol.*, **21**, 6833–6840.
29. Park,Y.J., Pardon,E., Wu,M., Steyaert,J. and Hol,W.G. (2012) Crystal structure of a heterodimer of editosome interaction proteins in complex with two copies of a cross-reacting nanobody. *Nucleic Acids Res.*, **40**, 1828–1840.
30. Panigrahi,A.K., Gygi,S.P., Ernst,N.L., Igo,R.P. Jr, Palazzo,S.S., Schnaufer,A., Weston,D.S., Carmean,N., Salavati,R., Aebersold,R. *et al.* (2001) Association of two novel proteins, TbMP52 and TbMP48, with the Trypanosoma brucei RNA editing complex. *Mol. Cell. Biol.*, **21**, 380–389.
31. Deng,J., Schnaufer,A., Salavati,R., Stuart,K.D. and Hol,W.G. (2004) High resolution crystal structure of a key editosome enzyme from Trypanosoma brucei: RNA editing ligase 1. *J Mol Biol*, **343**, 601–613.
32. Pascal,J.M. (2008) DNA and RNA ligases: structural variations and shared mechanisms. *Curr. Opin. Struct. Biol.*, **18**, 96–105.
33. Shuman,S. and Lima,C.D. (2004) The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases. *Curr. Opin. Struct. Biol.*, **14**, 757–764.
34. Lehman,I.R. (1974) DNA ligase: structure, mechanism, and function. *Science*, **186**, 790–797.
35. 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.
36. Chu,C., Das,K., Tyminski,J.R., Bauman,J.D., Guan,R., Qiu,W., Montelione,G.T., Arnold,E. and Shatkin,A.J. (2011) Structure of the guanylyltransferase domain of human mRNA capping enzyme. *Proc. Natl Acad. Sci USA*, **108**, 10104–10108.
37. Hakansson,K., Doherty,A.J., Shuman,S. and Wigley,D.B. (1997) X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. *Cell*, **89**, 545–553.
38. Nandakumar,J., Nair,P.A. and Shuman,S. (2007) Last stop on the road to repair: structure of E. coli DNA ligase bound to nicked DNA-adenylate. *Mol. Cell*, **26**, 257–271.
39. Nair,P.A., Nandakumar,J., Smith,P., Odell,M., Lima,C.D. and Shuman,S. (2007) Structural basis for nick recognition by a minimal pluripotent DNA ligase. *Nat. Struct. Mol. Biol.*, **14**, 770–778.
40. Nandakumar,J., Shuman,S. and Lima,C.D. (2006) RNA ligase structures reveal the basis for RNA specificity and conformational changes that drive ligation forward. *Cell*, **127**, 71–84.
41. Doherty,A.J. and Dafforn,T.R. (2000) Nick recognition by DNA ligases. *J. Mol. Biol.*, **296**, 43–56.
42. Nishida,H., Kiyonari,S., Ishino,Y. and Morikawa,K. (2006) The closed structure of an archaeal DNA ligase from Pyrococcus furiosus. *J. Mol. Biol.*, **360**, 956–967.
43. Pascal,J.M., Tsodikov,O.V., Hura,G.L., Song,W., Cotner,E.A., Classen,S., Tomkinson,A.E., Tainer,J.A. and Ellenberger,T. (2006) A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA. *Mol. Cell*, **24**, 279–291.
44. Sawaya,R. and Shuman,S. (2003) Mutational analysis of the guanylyltransferase component of Mammalian mRNA capping enzyme. *Biochemistry*, **42**, 8240–8249.
45. Sriskanda,V. and Shuman,S. (1998) Mutational analysis of Chlorella virus DNA ligase: catalytic roles of domain I and motif VI. *Nucleic Acids Res.*, **26**, 4618–4625.
46. El Omari,K., Ren,J., Bird,L.E., Bona,M.K., Klarmann,G., LeGrice,S.F. and Stammers,D.K. (2006) Molecular architecture and ligand recognition determinants for T4 RNA ligase. *J. Biol. Chem.*, **281**, 1573–1579.
47. Gao,G., Rogers,K., Li,F., Guo,Q., Osato,D., Zhou,S.X., Falick,A.M. and Simpson,L. (2010) Uridine insertion/deletion RNA editing in Trypanosomatids: specific stimulation in vitro of Leishmania tarentolae REL1 RNA ligase activity by the MP63 zinc finger protein. *Protist*, **161**, 489–496.
48. Lam,A.Y., Pardon,E., Korotkov,K.V., Hol,W.G. and Steyaert,J. (2009) Nanobody-aided structure determination of the EpsI:EpsJ pseudopilin heterodimer from Vibrio vulnificus. *J. Struct. Biol.*, **166**, 8–15.
49. Korotkov,K.V., Johnson,T.L., Jobling,M.G., Pruneda,J., Pardon,E., Heroux,A., Turley,S., Steyaert,J., Holmes,R.K., Sandkvist,M. *et al.* (2011) Structural and functional studies on the interaction of GspC and GspD in the type II secretion system. *PLoS Pathog.*, **7**, e1002228.
50. Korotkov,K.V., Pardon,E., Steyaert,J. and Hol,W.G.J. (2009) Crystal structure of the N-terminal domain of the secretin GspD from ETEC determined with the assistance of a nanobody. *Structure*, **17**, 255–265.
51. Baker,N.A., Sept,D., Joseph,S., Holst,M.J. and McCammon,J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
52. Li,F., Ge,P., Hui,W.H., Atanasov,I., Rogers,K., Guo,Q., Osato,D., Falick,A.M., Zhou,Z.H. and Simpson,L. (2009) Structure of the core editing complex (L-complex) involved in uridine insertion/deletion RNA editing in trypanosomatid mitochondria. *Proc. Natl Acad. Sci. USA*, **106**, 12306–12310.
53. Golas,M.M., Bohm,C., Sander,B., Effenberger,K., Brecht,M., Stark,H. and Goringer,H.U. (2009) Snapshots of the RNA editing machine in trypanosomes captured at different assembly stages in vivo. *EMBO J.*, **28**, 766–778.

54. Ho,C.K., Wang,L.K., Lima,C.D. and Shuman,S. (2004) Structure and mechanism of RNA ligase. *Structure*, **12**, 327–339.
55. Wang,L.K., Nandakumar,J., Schwer,B. and Shuman,S. (2007) The C-terminal domain of T4 RNA ligase 1 confers specificity for tRNA repair. *RNA*, **13**, 1235–1244.
56. Van Duyne,G.D., Standaert,R.F., Karplus,P.A., Schreiber,S.L. and Clardy,J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.*, **229**, 105–124.
57. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
58. Cowtan,K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta. Crystallogr. D Biol. Crystallogr.*, **62**, 1002–1011.
59. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta. Crystallogr. D. Biol. Crystallogr.*, **60**, 2126–2132.
60. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta. Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.
61. Adams,P.D., Grosse-Kunstleve,R.W., Hung,L.W., Ioerger,T.R., McCoy,A.J., Moriarty,N.W., Read,R.J., Sacchettini,J.C., Sauter,N.K. and Terwilliger,T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta. Crystallogr. D Biol. Crystallogr.*, **58**, 1948–1954.
62. Collaborative Computational Project Number 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta. Cryst.*, **D50**, 760–763.
63. Krissinel,E. and Henrick,K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.
64. Gouet,P., Courcelle,E., Stuart,D.I. and Metoz,F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.
65. Lefranc,M.P. (2005) IMGT, the international ImMunoGeneTics information system: a standardized approach for immunogenetics and immunoinformatics. *Immun. Res.*, **1**, 3.
66. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
67. Chen,V.B., Arendall,W.B. III, Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta. Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.
68. Painter,J. and Merritt,E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta. Crystallogr. D Biol. Crystallogr.*, **62**, 439–450.