

Research article

Open Access

Taking U out, with two nucleases?

I Saira Mian¹, Elizabeth A Worthey and Reza Salavati*^{2,3}

Address: ¹Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720-8265, USA, ²Seattle Biomedical Research Institute, Seattle, Washington, 98109, USA and ³McGill University, Institute of Parasitology, Ste.-Anne-De-Bellevue, Quebec, H9X 3V9, Canada

Email: I Saira Mian - SMian@lbl.gov; Elizabeth A Worthey - Elizabeth_Worthey@rosettabel.com; Reza Salavati* - reza.salavati@mcgill.ca

* Corresponding author

Published: 16 June 2006

Received: 04 April 2006

BMC Bioinformatics 2006, 7:305 doi:10.1186/1471-2105-7-305

Accepted: 16 June 2006

This article is available from: <http://www.biomedcentral.com/1471-2105/7/305>

© 2006 Mian et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: REX1 and REX2 are protein components of the RNA editing complex (the editosome) and function as exouridylylases. The exact roles of REX1 and REX2 in the editosome are unclear and the consequences of the presence of two related proteins are not fully understood. Here, a variety of computational studies were performed to enhance understanding of the structure and function of REX proteins in *Trypanosoma* and *Leishmania* species.

Results: Sequence analysis and homology modeling of the Endonuclease/Exonuclease/Phosphatase (EEP) domain at the C-terminus of REX1 and REX2 highlights a common active site shared by all EEP domains. Phylogenetic analysis indicates that REX proteins contain a distinct subfamily of EEP domains. Inspection of three-dimensional models of the EEP domain in *Trypanosoma brucei* REX1 and REX2, and *Leishmania major* REX1 suggests variations of previously characterized key residues likely to be important in catalysis and determining substrate specificity.

Conclusion: We have identified features of the REX EEP domain that distinguish it from other family members and hence subfamily specific determinants of catalysis and substrate binding. The results provide specific guidance for experimental investigations about the role(s) of REX proteins in RNA editing.

Background

Most mitochondrial mRNAs in trypanosomatid parasites such as *Trypanosoma*, and *Leishmania species* undergo RNA editing [1-3]. This post-transcriptional process produces mature and functional mRNAs through a series of coordinated steps catalysed by a multi-protein complex that inserts and deletes uridylylates (Us) specified by guide RNAs (gRNAs). One hypothesis posits a structural and functional subdivision of the editosome into insertion and deletion subcomplexes [4-8]. Editosome proteins with endonuclease (REN1, REN2) [9,10], terminal uridylyl transferase (TUTase; RET1, RET2) [6,11,12], 3' exouridylylase (exoUase; REX1, REX2) [5,13], Ernst *et al.*,

unpublished), ligase (REL1, REL2) [5,8,14,15], and helicase (REH1) [16] activities have been identified and functionally characterized. Sets of proteins related by sequence similarity exhibit both unique and common functions. For instance, REN1 is an endoribonuclease that is specific for RNA editing deletion sites whereas REN2 is specific for RNA editing insertion sites. RET1 is implicated in the addition of the non-encoded 3'-oligo U tails to gRNAs but RET2 adds Us to pre-edited mRNAs. REL1 may be involved in U-deletion editing and REL2 in U-insertion editing. Six additional editosome proteins, KREPA1-A6, have varying degrees of sequence relatedness with each protein containing a C-terminal motif associated with an

oligonucleotide-binding (OB) fold [5,17-20]. Recent results point to both REX1 and REX2 as candidates for the RNA editing exoUase responsible for deletion of the 3' overhanging U residues from the mRNA 5' cleavage fragment. A U-specific exonuclease, REX1, has been partially purified from *L. tarentolae* [13]. The reconstitution of pre-cleaved U-deletion *in vitro* with recombinant *L. tarentolae* REX1 and REL1 proteins and the *in vivo* RNAi down-regulation of REX1 expression in *T. brucei* suggest that REX1 is the exoUase. However, the closely related REX2 protein (28% overall identity and 46% similarity in *T. brucei*) may be the putative exoUase since tagged *T. brucei* REL1 sub-complex consisting of REX2, REL1 and KREPA2 catalyze accurate U removal and ligation (i.e. pre-cleaved deletion editing) [5]. Thus, the exact roles of REX1 and REX2 in the editosome complex are unclear and the consequences of the presence of two related proteins are not fully understood.

Comparative sequence analysis indicates that both REX1 and REX2 contain a putative C-terminal Endonuclease/Exonuclease/Phosphatase (EEP) domain as well as a region exhibiting subtle, but significant similarity to a known 5'->3' exonuclease domain (*L. major* REX2 lacks an EEP domain because of a truncation at the C-terminus) [17]. Whether REX1 and REX2 have 5'->3' exonuclease activity in the editing complex is unknown.

In this study, we extend our previous analysis of the REX1 and REX2 EEP domains [17]. We use sequence analysis, homology modeling and phylogenetic analysis to enhance understanding of the structure and functions of REX proteins, as well as the relationships amongst EEP family members. Our results suggest that while these enzymes have diverged at the sequence level, the EEP domains share a common catalytic site. Our three-dimensional modeling studies suggest that the REX EEP domains fold in much the same way as other EEP domains whose structures have been determined by X-ray crystallography. We identify features of the REX EEP domain that distinguish it from other family members and hence subfamily specific determinants of catalysis and substrate binding.

Results and discussion

Trypanosomal REX proteins

The REX1 and REX2 proteins from three trypanosomatids show considerable sequence similarity suggesting they are encoded by paralogous genes (Fig. 1). Since the genes are present on non syntenic chromosomal regions in *L. major* and *T. brucei* (chromosome assignments not having been made for the *T. cruzi* genes) it is likely that the ancestral genes diverged prior to the fission/fusion events which resulted in the modern day trypanosomatid genomes.

Sequence and phylogenetic analysis of EEP domains

The EEP domains in proteins from a variety of Eucarya and Bacteria were modelled and analyzed using an HMM-based approach. Members of the EEP domain family include magnesium dependent endonucleases (L1-EN, DNaseI, APE1, APE2) [21-26], exonucleases (ExoIII, REX1, REX2) [5,13,17,27], and phosphatases of lipid second messengers (I5PP) [28] (Fig. 2). Although these proteins have diverse substrate specificities, REX EEP domains possess the conserved sequence motifs that have been used to characterize other EEP domains (I to VI, Fig. 2). A phylogenetic tree of EEP domains indicates that REX and APE proteins form distinct subfamilies (Fig. 3). EEP domains in I5PP proteins are more REX- than APE-like.

Homology modelling of *T. brucei* REX1 and REX2 and *L. major* REX1

The X-ray crystal structures of the EEP domains in two DNA repair enzymes (Fig. 2, APE1_Hs_1HD7, ExoIII_Ec_1AKO) [25,27] were used as the templates to build homology models of the EEP domains in three REX proteins (REX1_Tbrucei, REX2_Tbrucei, REX1_Lmajor). *H. sapiens* APE1 and *E. coli* ExoIII are functional homologs that possess apurinic/apyrimidinic (AP) endonuclease activity and which hydrolyze the phosphodiester bond of DNA at the AP sites by cleaving the DNA in intact strands [29]. Although ExoIII also has 3'->5' exonuclease activity, its biological role remains unclear. Following cleavage by ExoIII or APE1 the bacterium *E. coli* uses DNA polymerase I (pol I) to fill in the single-nucleotide gap whereas the eucaryote *H. sapiens* uses DNA polymerase β (pol β). Pol I has also a 3'->5' proof-reading activity which allows the removal of misincorporated nucleotides [30]. Although pol β is prone to high error (one mistake per 4000 bases inserted) [31], it lacks the proof reading mechanism found in pol I [32]. Instead, APE1 acts as an exonuclease that trims off nucleotides from DNA ends that do not terminate in correct basepairs [33].

As would be expected, the four-layer α/β fold observed in the crystal structures of the templates (APE1/1HD7; ExoIII/1AKO) is reflected in the three-dimensional models of the target proteins (Fig. 4). The roles of amino acids in the conserved sequence motifs found in EEP domains (I to VI, Fig. 2) were examined in APE1/1HD7 and ExoIII/1AKO (Table 1). Based on high resolution X-ray crystal structures, the similar overall catalytic mechanism of APE1 and ExoIII involves the abstraction of a proton from a water molecule by a residue acting as a general base [23,27]. The resultant nucleophilic hydroxide ion attacks a scissile phosphate. The major difference between APE1 and ExoIII is that the catalytic residue which deprotonates the water molecule in APE1 is Asp in motif IV whereas in ExoIII it is the His in motif VI (Table 1 and Fig. 5, 6, 7). In addition, ExoIII appears to be a relatively more powerful

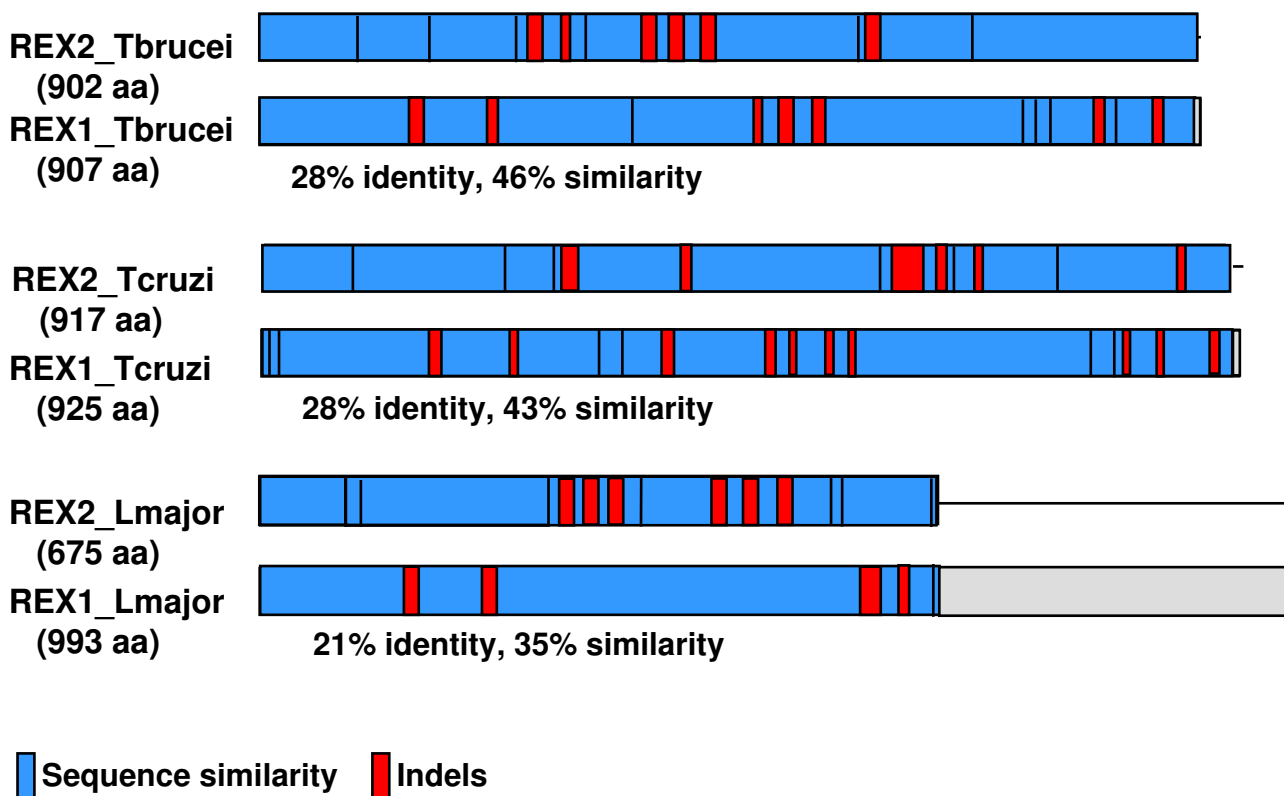


Figure 1

Schematic diagram of the REX1 and REX2 editosome proteins from *Trypanosoma brucei* (REX1_Tbrucei, REX2_Tbrucei), *T. cruzi* (REX1_Tcruzi, REX2_Tcruzi), and *Leishmania major* (REX1_Lmajor, REX2_Lmajor). Each protein is represented as a horizontal bar and the number of amino acids (aa) is given. The percent identity and similarity for each pair of putative paralogs is indicated. The regions of similarity are in blue whereas segments that are added or deleted in one paralog (Indels) are in red. In *L. major*, the C-terminal portion of REX1 which contains an EEP domain (gray) has no counterpart in REX2 because of a truncation.

3'-exonuclease than APE1 [34,35]. This enhanced activity has been attributed to the fewer hydrophobic residues in the active site of ExoIII [23,24] (and see below).

Putative activity and mechanisms of action of REX proteins

To gain insights into the enzymatic activity of the REX proteins, we measured the rmsd values of the superposed models of each of REX EEP domain model to APE1/1HD7 and ExoIII/1AKO. The REX2_Tbrucei model is closer to APE1/1HD7 than to ExoIII/1AKO (2.54 Å versus 4.19 Å). In contrast, both the REX1_Lmajor and REX1_Tbrucei models are closer to ExoIII (2.17 Å, 1.97 Å, respectively) than to APE1/1HD7 (3.41 Å, 3.91 Å, respectively). These results suggest that REX2_Tbrucei may have more in common with APE1/1HD7 whereas REX1_Lmajor and REX1_Tbrucei may be more related to ExoIII/1AKO. The latter data suggest that REX1 exoribonuclease activity in

both *Leishmania* and *Trypanosoma* species may share a similar catalytic mechanism with ExoIII [13]. These results also raise the possibility that if REX2_Tbrucei is similar to APE1, the potential proof-reading function of *Trypanosoma* REX2 may remove (i) the extraneous U residues added during TUTase activity [36-38] and (ii) the Us that result from TUTase function within a U-deletion site [39]. Current data indicate that *Leishmania* REX2 lack a C-terminal EEP domain (Fig. 1, Fig. 2) and hence potential proof-reading activity. Although this function could be compensated for by the related REX1 protein (see below), the absence of an EEP domain could be explained by less extensive editing in *Leishmania* species compared to *Trypanosoma* family members. While both APE1 and ExoIII are known to act as endonucleases, such activity has not yet been demonstrated for the REX family of proteins.

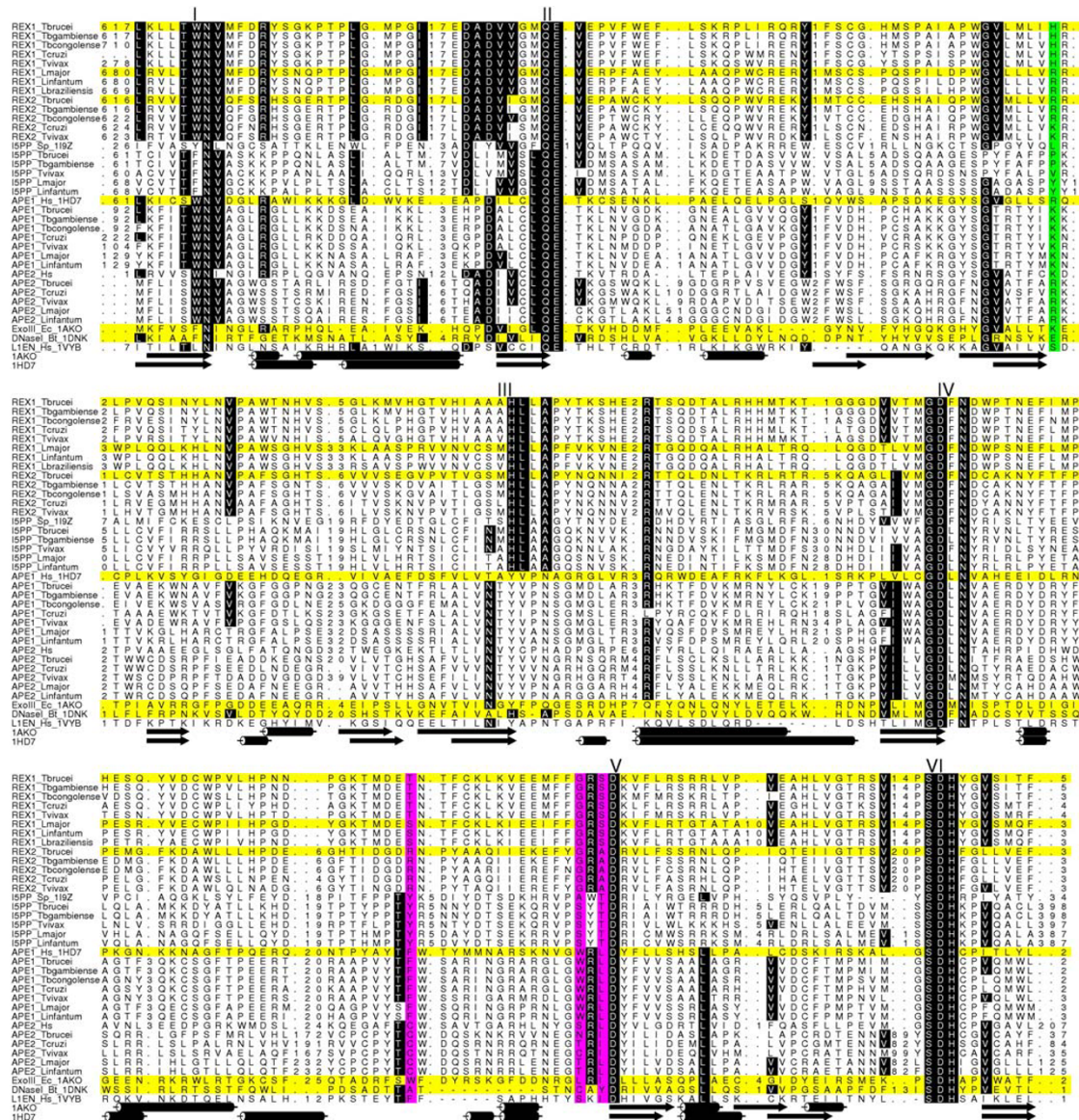


Figure 2
HMM-generated multiple sequence alignment of the EEP domain from a variety of Bacteria and Eucarya. The EEP domain at the C-terminus of REX2 and REX1 is the putative exonuclease component of the editosome. Residues that are conserved in >50% of the sequences are in black and numbers indicate the number of amino acids not shown explicitly. Six EEP domains of particular interest are in yellow. The Roman numerals above the alignment mark the six conserved sequence motifs that have been used to characterize EEP domain (see also Table 1). Columns in pink indicate the putative substrate specificity active-site hydrophobic pocket. The column in green marks the location of the essential Glu in DNaseI (DNaseI_Bt_IDNK, the last sequence highlighted in yellow). The proteins shown are REX1 (RNA editing exonuclease 1), REX2 (RNA editing exonuclease 2), ISPP (inositol polyphosphate 5'-phosphatase), APE1 (apurinic/aprimidinic endonuclease 1), and APE2 (apurinic/aprimidinic endonuclease 2); these sequences are from *T. brucei* (Tbrucei), *T. gambiense* (Tgambiense), *T. congolense* (Tbcongolense), *T. cruzi* (Tcruzi), *T. vivax* (Tvivax), *L. major* (Lmajor), *L. infantum* (Linfantum), *L. braziliensis* (Lbraziliensis), *Schizosaccharomyces pombe* (Sp), *Homo sapiens* (Hs), *Escherichia coli* (Ec), and *Bos taurus* (Bt). EEP domains whose structures have been determined experimentally are ISPP_Sp_I19Z (*S. pombe* phosphatidylinositol phosphate phosphatase, RCSB code I19Z), APE1_Hs_IHD7 (*H. sapiens* AP endonuclease 1, IHD7), ExoIII_Ec_IAKO, (*E. coli* exonuclease III, IAKO); DNaseI_Bt_IDNK (*B. taurus* deoxyribonuclease I, IDNK) and LIEN_Hs_IVYB: (*H. sapiens* LI endonuclease, IVYB). Cylinders and arrows denote the α -helices and β -strands given in the RCSB entries IAKO and IHD7. The protein sequences, EEP domain HMM and alignment are available as Supplementary material.

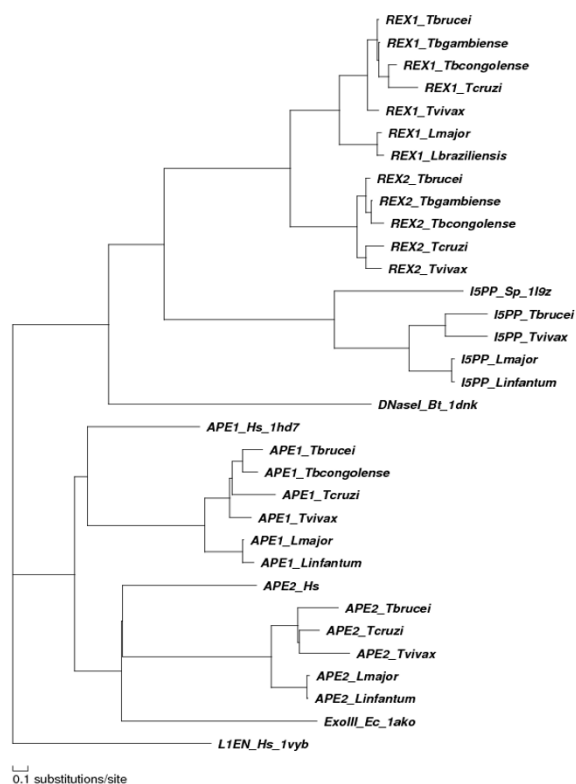


Figure 3
A maximum likelihood tree estimated from an HMM-generated multiple sequence alignment of EEP protein domains (Fig. 2 provides information on the proteins shown).

In addition to the primary catalytic residues, the active site of APE1/1HD7 and ExoIII/1AKO contains a bulky hydrophobic pocket that has been proposed to act as a sequence specific "gate-keeper" able to accommodate only abasic sites [23,24]. In APE1/1HD7, the hydrophobic pocket is composed of Phe266, Trp280, and Leu282 (Fig. 2, Fig. 5). The equivalent pocket in ExoIII/1AKO is larger and consists of Trp212, Leu226, and Ile228 (Fig. 2, Fig. 6, 7).

Mutations of the APE1 hydrophobic pocket that results in smaller residues (e.g. Phe266 to Ala/Cys, Trp280 to Ile/Leu/Ser), can enhance its 3'-exonuclease activity [24]. ExoIII possesses a hydrophobic pocket containing only one aromatic residue and the enzyme is a better 3' exonuclease than APE1. These two findings support the idea that the hydrophobic pocket in EEP domains plays a significant role in nucleotide binding and specificity. Our sequence and structural analysis suggests that REX1 and REX2 do not have a bulky hydrophobic pocket but instead share a pocket composed of smaller residues. The pocket is formed by Arg834, Gly848, and Ala850 in REX2_T.brucei, Thr825, Gly839, and Ser841 in REX1_T.brucei, and by Ser916, Gly930, and Ser932 in REX1_Lmajor (Fig. 2, Fig. 5, 6, 7). The equivalent of Trp280 in APE1/1HD7 or Leu226 in ExoIII/1AKO is Gly839 in REX1_T.brucei, Gly848 in REX2_T.brucei, and Gly930 in REX1_Lmajor. Thus, we predict that REX1 and REX2 have the potential to accommodate an extrahelical residue (i.e. uridine) downstream (3') of the scissile bond. The conserved polar Ser/Thr in REX1, and positively charged Arg in REX2 may form hydrogen bonds with the extrahelical base (column in pink between conserved motifs IV and V in Fig. 2). In REX1, a Ser in *Leishmania* is a Thr in *Trypanosoma* (Fig. 6, 7). This suggests altered substrate specificity for *Leishmania* species, which may partially compensate for the absence of an EEP domain in REX2.

Comparison of the REX subfamily of EEP domains with that of the *B. taurus* DNase I (DnaseI_Bt_1DNK, Fig. 2) and inositol polyphosphate 5'-phosphatase (I5PP, Fig. 2) family members reveals a conserved His in motif III (Fig. 2 and Table 1) [22,28,40]. In DNase I, this residue is part of the essential His-Glu catalytic pair located within the active site and is proposed to act as a general acid acting to stabilize the leaving group [22]. Mutagenesis of Glu in DNase I has also shown the importance of this residue for catalytic activity. However, the sequence analysis of the REX and other family members of EEP domains does not reveal a conserved Glu or other negatively charged residue

Table 1: Conserved sequence motifs found in EEP domains (labeled I through VI in Fig. 2) and the roles of specific amino acids in two EEP domains of known three-dimensional structure (APE1_Hs_1HD7, ExoIII_Ec_1AKO in Fig. 5-7).

REX1/REX2	APE1	EXO III
I. TWN	Asn forms a hydrogen bond to catalytic Asp in motif IV	Asn interacts with 5' phosphate group
II. QE	Glu coordinates Mg ²⁺ ion	Glu coordinates Mg ²⁺ ion
III. HL	His is substituted by Tyr in APE1 but is the catalytic His in DNase I family	His is substituted by Tyr
IV. GDFN	Catalytic Asp deprotonates the water molecule; Asn forms a hydrogen bond to scissile phosphate	Asn forms a hydrogen bond to scissile phosphate
V. GRXD (X is S in REX1 and A in REX2)	Asp forms a hydrogen bond to His in motif VI	Asp forms a hydrogen bond to His in motif VI
VI. SDH	His forms a hydrogen bond to Asp in motif V and to scissile phosphate	Catalytic His deprotonates the water molecule and forms a hydrogen bond to Asp in motif V

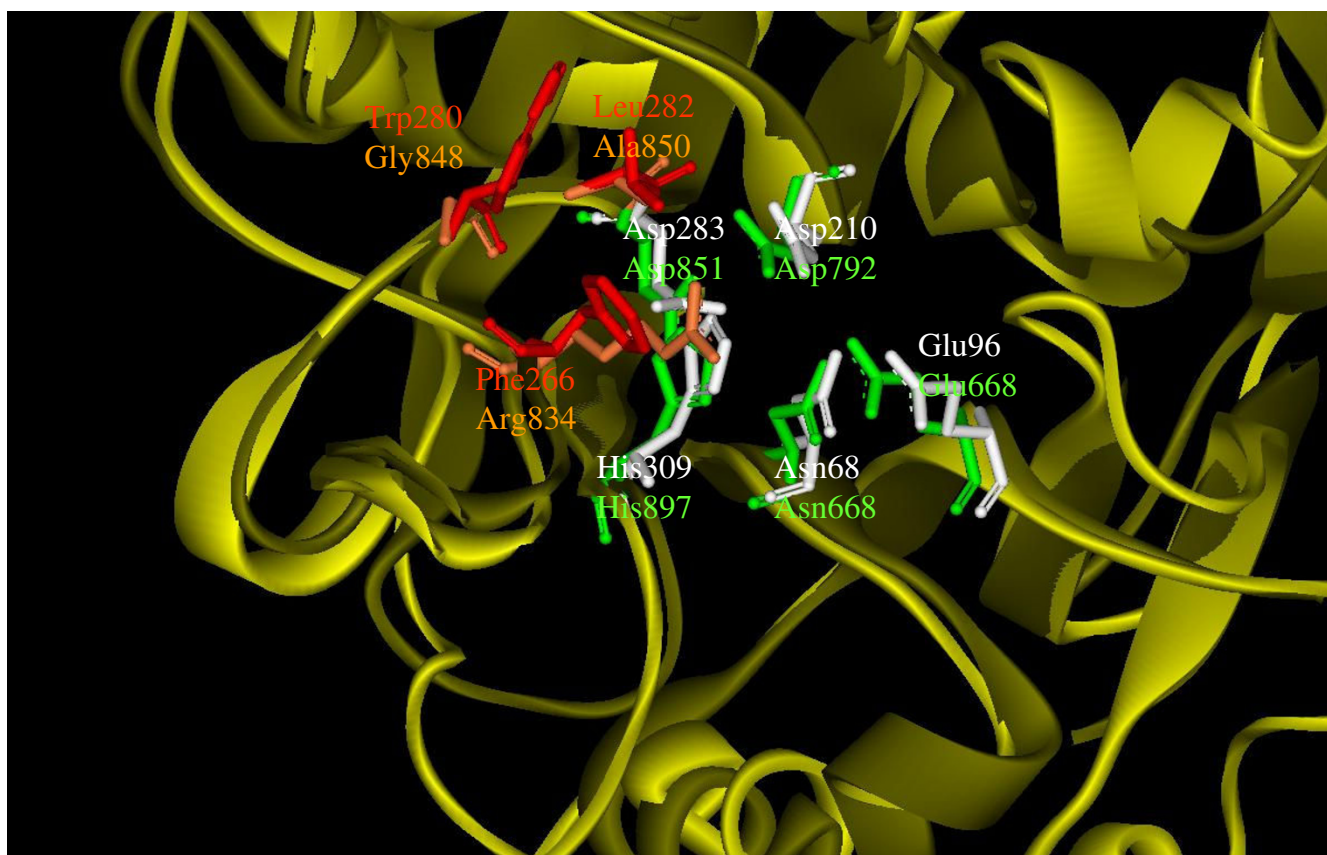


Figure 5

The active sites of APEI_Hs_IHD7 (green) superposed onto REX2_Tbrucei (yellow) with the side chains of critical amino acids shown explicitly. The APEI (REX2) catalytic residues are in white (green) and the hydrophobic residues are in red (orange).

that could pair with the His residue (Fig. 2) [41]. Therefore, we predict that in the REX EEP domains, the His residue forms a hydrogen bond to and further polarizes the scissile phosphate, as previously proposed for the I5PP family of proteins [41].

Conclusion

Using a variety of computational approaches, we have identified conserved motifs and a critical substrate binding pocket in the REX subfamily of EEP domains. Our results suggest experiments that could be performed to examine the distinct catalytic roles of REX proteins in the editosome.

Methods

Trypanosomal proteins

The genomic locations of the trypanosomal proteins discussed in this work were determined using the GeneDB Artemis interface to data from the TriTryp genomes sequencing consortium [42]. Putative orthologs were initially identified using BLAST searches of the unfinished

genome sequences. These findings were later confirmed through mutual best BlastP analysis amongst the unique portions of the three essentially complete trypanosomatid genomes. The sequences of these homologous genes were also confirmed against the high coverage sequences available from the genome projects. The available, unfinished genome sequences for the remaining trypanosomatids discussed in this manuscript have been made available through GeneDB, and were searched using the blast algorithm to identify putative orthologues in these species. In a number of instances, the matches were not full length, and the gene was present in more than a single contig, thereby requiring assembly of the sequence to obtain full-length genes.

Sequence and phylogenetic analysis of EEP domains

The sequence and phylogenetic analysis of EEP domains was performed using a hidden Markov model (HMM)-based approach that has been employed successfully elsewhere (see, for example, the following refs [43-47]).

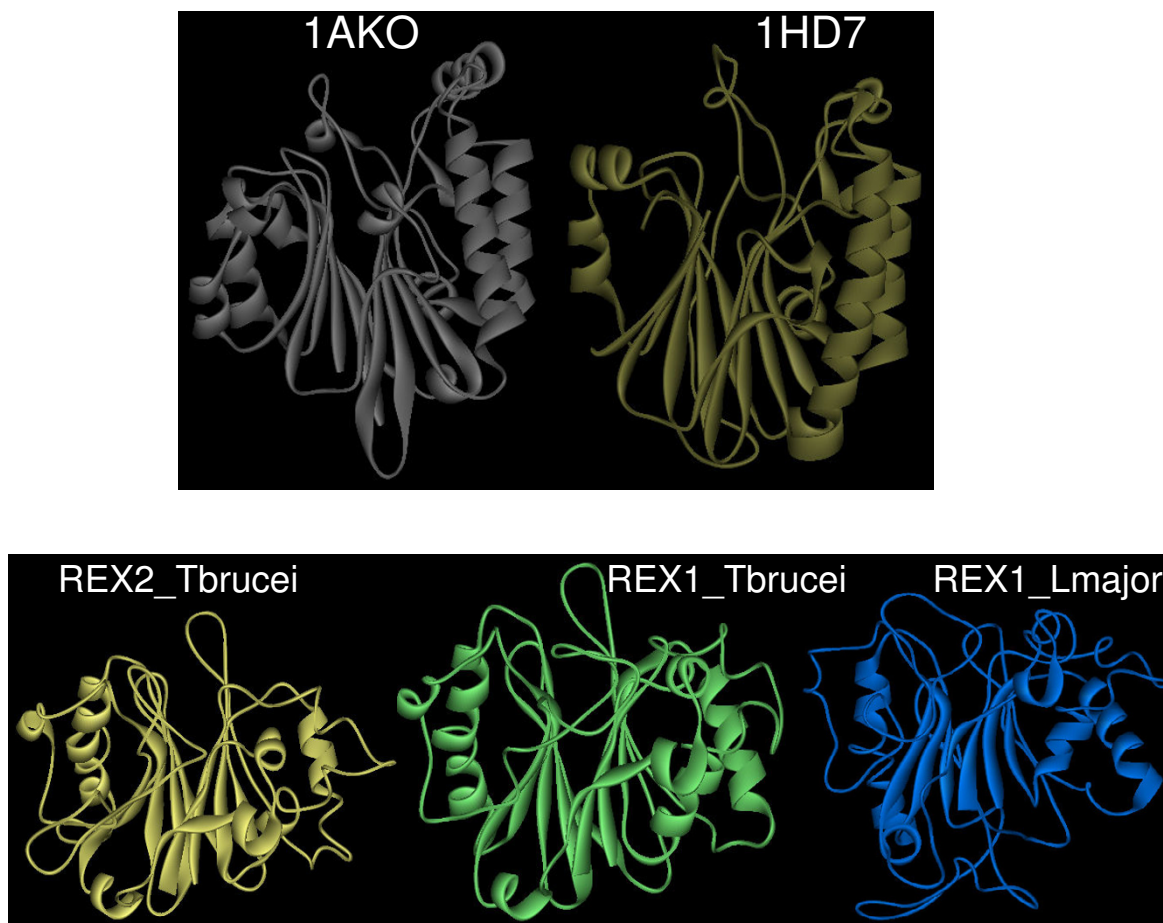


Figure 4

Three-dimensional models of trypanosomal REX EEP domains built using EEP domains whose structures have been determined by X-ray crystallography. The upper panel shows ribbon diagrams for two experimentally determined structures: 1AKO is *E. coli* exonuclease III (Fig. 2, ExoIII_Ec_1AKO) and 1HD7 is *H. sapiens* APE1 (APE1_Hs_1HD7). The lower panel shows ribbon diagrams of the homology-built structures: *T. brucei* REX2 (REX2_Tbrucei), *T. brucei* REX1 (REX1_Tbrucei) and *L. major* REX1 (REX1_Lmajor). The four-layered α , β sandwich fold in each EEP domain is shown in the same orientation with the substrate binding surface at the top.

Previously, we estimated an HMM of the EEP domain using the SAM software suite version 3.3.1 [48] and a limited number of protein sequences [17]. For this work, the parameters of this initial HMM were updated using an expanded training set that included additional eukaryotic (including trypanosomatid) and bacterial sequences. The ensuing EEP domain HMM was used to generate a multiple sequence alignment of all the EEP domains in the training set and the alignment was annotated with known structural information for some members of the EEP domain family (Fig. 2).

A phylogenetic tree for EEP domains was estimated using an HMM-generated multiple sequence alignment of the

training set and ProtML in the MOLPHY software suite version 2.3b3s. Since insert states in the HMM are uninformative, the alignment consisted only of residues aligned to match states of the EEP HMM. ProtML infers an evolutionary tree from amino acid sequences using the Maximum Likelihood (ML) method. The tree with the maximum likelihood was used to understand the relationships between EEP domains.

Homology modeling

Three-dimensional models of selected REX EEP domains were built as described previously [49] using the MODELLER program [50] using software programs from Accelrys Inc., DS Modeling 1.1 and an alignment of a domain of

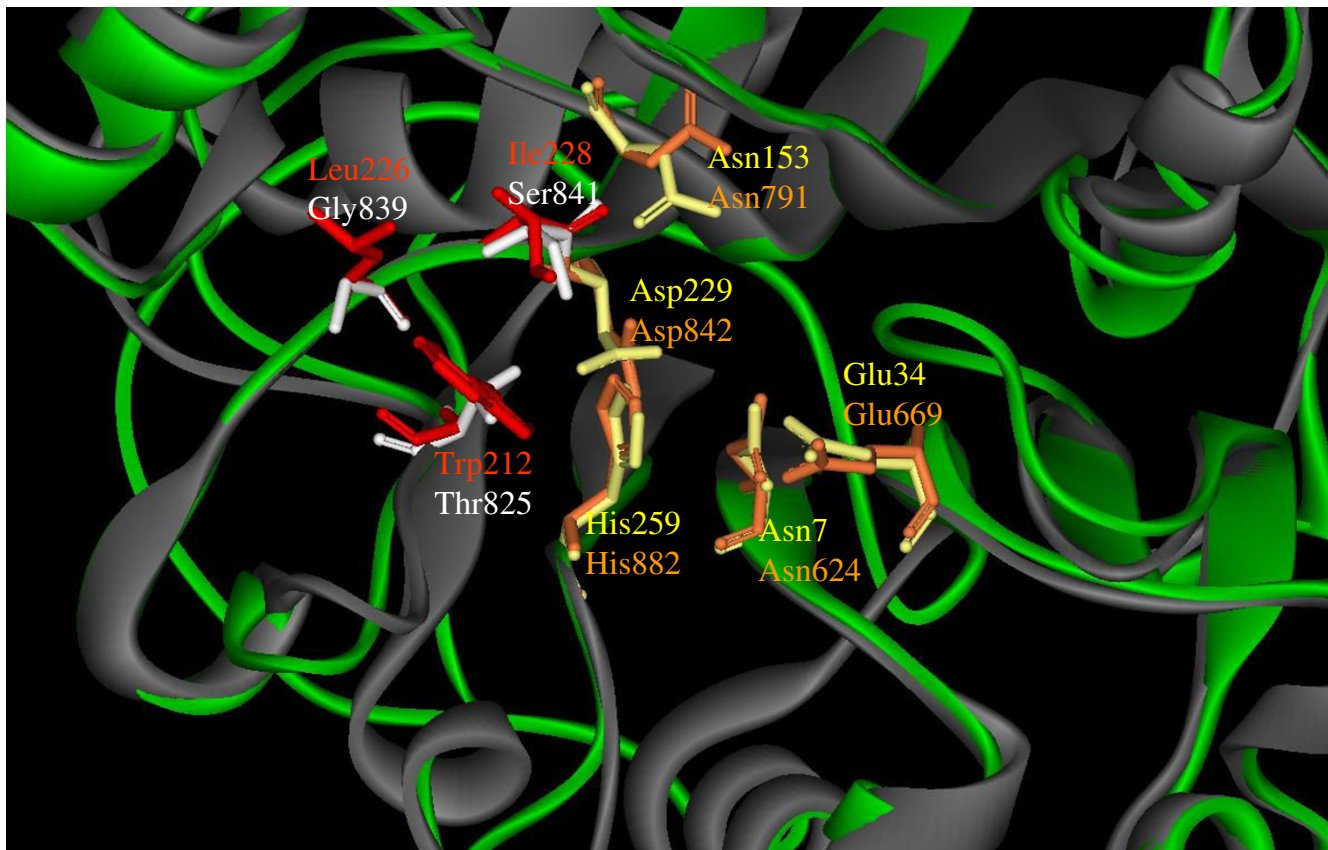


Figure 6

The active sites of ExoIII_Ec_1AKO (grey) superposed on REX1_Tbrucei (green) with the side chains of critical amino acids shown explicitly. The ExoIII (REX1) catalytic residues are in yellow (orange) and the hydrophobic residues are in red (white).

unknown structure against a domain of known structure (Fig. 2). The sequences/structures of APE1_Hs_1HD7 and ExoIII_Ec_1AKO were used as the templates for constructing models of three targets, REX1_Tbrucei, REX2_Tbrucei and REX1_Lmajor. This particular choice was based on (i) the functional homology and multiple sequence alignment (Fig. 2), (ii) a statistically significant PSI-BLAST score between the target and an EEP family protein (E-value = 5e-12 REX1_Tbrucei, 6e-08 REX1_Lmajor, and 2e-08 REX2_Tbrucei), and (iii) a statistical significant score produced by 3D-Jury (120–156, well above the cutoff value of 50). The 3D-Jury metaserver [51] selects the most abundant models from the set of 3D models generated by various independent prediction providers. To measure the r.m.s. deviation of the superposed template and the target, the complete sequences of the predicted EEP domains aligned in Figure 2 were used to measure the r.m.s. devi-

ation values. The quality of predicted modeled structures were checked with the Profiles_3D program [52] in DS Modeling 1.1.

Abbreviations

REX, RNA editing exouridylylase; EEP, Endonuclease/Exonuclease/Phosphatase; REN, RNA editing endonuclease; RET, RNA editing terminal uridylyl transferase; TUTase, terminal uridylyl transferase; exoUase, exouridylylase; REL, RNA editing ligase; REH, RNA editing helicase, APE, apurinic/aprimidinic endonuclease; OB, oligonucleotide binding; L1-EN, L1 endonuclease; ExoIII, exonuclease III; I5PP, inositol polyphosphate 5'-phosphatase; pol I, DNA polymerase I; pol β , DNA polymerase β .

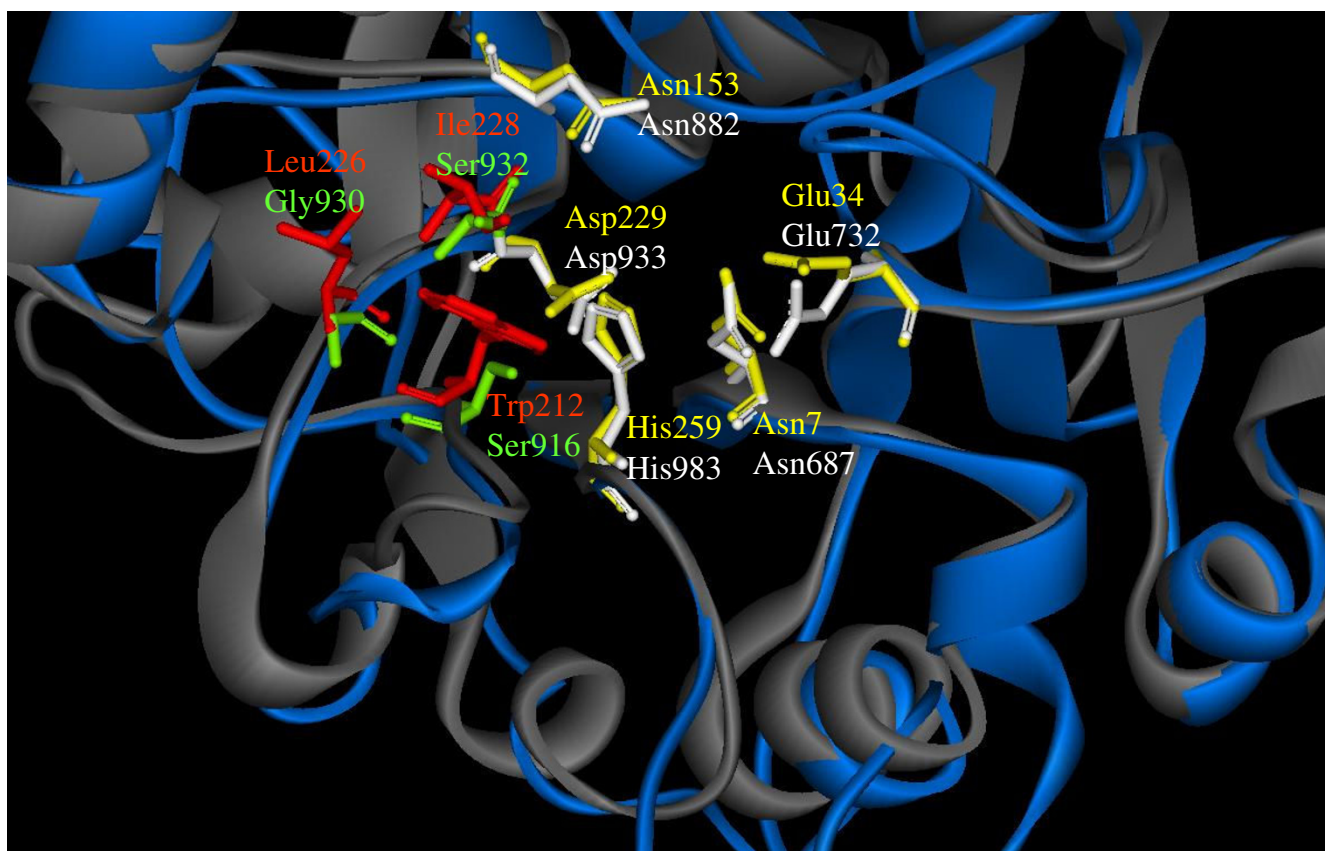


Figure 7

The active sites of ExoIII_Ec_IAKO (grey) superposed on REXI_Lmajor (blue) with the side chains of critical amino acids shown explicitly. The ExoIII (REXI) catalytic residues are in yellow (white) and the hydrophobic residues are in red (green).

Authors' contributions

ISM carried out the HMM analysis and phylogenetic inference of the EEP domains and contributed to writing the manuscript. EAW identified the EEP domain homologs and contributed to the phylogenetic inference analyses. RS carried out the homology modeling analysis, drafted the manuscript, conceived and coordinated the study. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the TriTryp Genome Consortium for their considerable effort that made this work possible. This work was supported by NIH grant IR21AI053784-01 to R.S and funds from the National Institute on Aging, National Institute of Environmental Health Sciences, U.S. Department of Energy (OBER) and California Breast Cancer Research Program to I.S.M.

References

- Madison-Antenucci S, Grams J, Hajduk SL: **Editing machines: the complexities of trypanosome RNA editing.** *Cell* 2002, **108**:435-438.
- Simpson L, Aphasizhev R, Gao G, Kang X: **Mitochondrial proteins and complexes in Leishmania and Trypanosoma involved in U-insertion/deletion RNA editing.** *RNA* 2004, **10**:159-170.
- Stuart KD, Schnauffer A, Ernst NL, Panigrahi AK: **Complex management: RNA editing in trypanosomes.** *Trends Biochem Sci* 2005, **30**:97-105.
- Aphasizhev R, Aphasizheva I, Nelson RE, Gao G, Simpson AM, Kang X, Falick AM, Sbicego S, Simpson L: **Isolation of a U-insertion/deletion editing complex from Leishmania tarentolae mitochondria.** *EMBO J* 2003, **22**:913-924.
- Schnauffer A, Ernst NL, Palazzo SS, O'Rear J, Salavati R, Stuart K: **Separate insertion and deletion subcomplexes of the Trypanosoma brucei RNA editing complex.** *Mol Cell* 2003, **12**:307-319.
- Ernst NL, Panicucci B, Igo RPJ, Panigrahi AK, Salavati R, Stuart K: **TbMP57 is a 3' terminal uridylyl transferase (TUTase) of the Trypanosoma brucei editosome.** *Mol Cell* 2003, **11**:1525-1536.
- Huang CE, O'Hearn SF, Sollner-Webb B: **Assembly and function of the RNA editing complex in Trypanosoma brucei requires band III protein.** *Mol Cell Biol* 2002, **22**:3194-3203.
- Cruz-Reyes J, Zhelonkina AG, Huang CE, Sollner-Webb B: **Distinct functions of two RNA ligases in active Trypanosoma brucei RNA editing complexes.** *Mol Cell Biol* 2002, **22**:4652-4660.
- Trotter JR, Ernst NL, Carnes J, Panicucci B, Stuart K: **A deletion site editing endonuclease in Trypanosoma brucei.** *Mol Cell* 2005, **20**:403-412.
- Carnes J, Trotter JR, Ernst NL, Steinberg A, Stuart K: **An essential RNase III insertion editing endonuclease in Trypanosoma brucei.** *Proc Natl Acad Sci U S A* 2005, **102**:16614-16619.
- Aphasizhev R, Sbicego S, Peris M, Jang SH, Aphasizheva I, Simpson AM, Rivlin A, Simpson L: **Trypanosome mitochondrial 3' terminal uridylyl transferase (TUTase): the key enzyme in U-insertion/deletion RNA editing.** *Cell* 2002, **108**:637-648.

12. Aphasizhev R, Aphasizheva I, Simpson L: **A tale of two TUTases.** *Proc Natl Acad Sci U S A* 2003, **100**:10617-10622.
13. Kang X, Rogers K, Gao G, Falick AM, Zhou S, Simpson L: **Reconstitution of uridine-deletion precleaved RNA editing with two recombinant enzymes.** *Proc Natl Acad Sci U S A* 2005, **102**:1017-1022.
14. Huang CE, Cruz-Reyes J, Zhelonkina AG, O'Hearn S, Wirtz E, Sollner-Webb B: **Roles for ligases in the RNA editing complex of Trypanosoma brucei: band IV is needed for U-deletion and RNA repair.** *EMBO J* 2001, **20**:4694-4703.
15. Drodz M, Palazzo SS, Salavati R, O'Rear J, Clayton C, Stuart K: **TbMP81 is required for RNA editing in Trypanosoma brucei.** *EMBO J* 2002, **21**:1791-1799.
16. Missel A, Souza AE, Norskau G, Goring HU: **Disruption of a gene encoding a novel mitochondrial DEAD-box protein in Trypanosoma brucei affects edited mRNAs.** *Mol Cell Biol* 1997, **17**:4895-4903.
17. Worthey EA, Schnauffer A, Mian IS, Stuart K, Salavati R: **Comparative analysis of editosome proteins in trypanosomatids.** *Nucleic Acids Res* 2003, **31**:6392-6408.
18. Law JA, Huang CE, O'Hearn SF, Sollner-Webb B: **In Trypanosoma brucei RNA editing, band II enables recognition specifically at each step of the U insertion cycle.** *Mol Cell Biol* 2005, **25**:2785-2794.
19. Brecht M, Niemann M, Schluter E, Muller UF, Stuart K, Goring HU: **TbMP42, a protein component of the RNA editing complex in African trypanosomes, has endo-exoribonuclease activity.** *Mol Cell* 2005, **17**:621-630.
20. Panigrahi AK, Schnauffer A, Carmean N, Igo RPJ, Gygi SP, Ernst NL, Palazzo SS, Weston DS, Aebersold R, Salavati R, Stuart KD: **Four related proteins of the Trypanosoma brucei RNA editing complex.** *Mol Cell Biol* 2001, **21**:6833-6840.
21. Weichenrieder O, Repanas K, Perrakis A: **Crystal structure of the targeting endonuclease of the human LINE-1 retrotransposon.** *Structure* 2004, **12**:975-986.
22. Suck D, Oefner C: **Structure of DNase I at 2.0 Å resolution suggests a mechanism for binding to and cutting DNA.** *Nature* 1986, **321**:620-625.
23. Mol CD, Izumi T, Mitra S, Tainer JA: **DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination [corrected].** *Nature* 2000, **403**:451-456.
24. Hadi MZ, Ginalski K, Nguyen LH, Wilson DMIII: **Determinants in nuclease specificity of Ape1 and Ape2, human homologues of Escherichia coli exonuclease III.** *J Mol Biol* 2002, **316**:853-866.
25. Beernink PT, Segelke BV, Hadi MZ, Erzberger JP, Wilson DMIII, Rupp B: **Two divalent metal ions in the active site of a new crystal form of human apurinic/apyrimidinic endonuclease, Ape1: implications for the catalytic mechanism.** *J Mol Biol* 2001, **307**:1023-1034.
26. Hadi MZ, Wilson DMIII: **Second human protein with homology to the Escherichia coli abasic endonuclease exonuclease III.** *Environ Mol Mutagen* 2000, **36**:312-324.
27. Mol CD, Kuo CF, Thayer MM, Cunningham RP, Tainer JA: **Structure and function of the multifunctional DNA-repair enzyme exonuclease III.** *Nature* 1995, **374**:381-386.
28. Tsujishita Y, Guo S, Stolz LE, York JD, Hurley JH: **Specificity determinants in phosphoinositide dephosphorylation: crystal structure of an archetypal inositol polyphosphate 5-phosphatase.** *Cell* 2001, **105**:379-389.
29. Mol CD, Hosfield DJ, Tainer JA: **Abasic site recognition by two apurinic/apyrimidinic endonuclease families in DNA base excision repair: the 3' ends justify the means.** *Mutat Res* 2000, **460**:211-229.
30. Tabor S, Huber HE, Richardson CC: **Escherichia coli thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7.** *J Biol Chem* 1987, **262**:16212-16223.
31. Osheroff WP, Jung HK, Beard WA, Wilson SH, Kunkel TA: **The fidelity of DNA polymerase beta during distributive and processive DNA synthesis.** *J Biol Chem* 1999, **274**:3642-3650.
32. Kunkel TA: **The mutational specificity of DNA polymerase-beta during in vitro DNA synthesis. Production of frameshift, base substitution, and deletion mutations.** *J Biol Chem* 1985, **260**:5787-5796.
33. Chou KM, Cheng YC: **An exonucleolytic activity of human apurinic/apyrimidinic endonuclease on 3' mispaired DNA.** *Nature* 2002, **415**:655-659.
34. Demple B, Harrison L: **Repair of oxidative damage to DNA: enzymology and biology.** *Annu Rev Biochem* 1994, **63**:915-948.
35. Seki S, Hatsushika M, Watanabe S, Akiyama K, Nagao K, Tsutsui K: **cDNA cloning, sequencing, expression and possible domain structure of human APEX nuclease homologous to Escherichia coli exonuclease III.** *Biochim Biophys Acta* 1992, **1131**:287-299.
36. Byrne EM, Connell GJ, Simpson L: **Guide RNA-directed uridine insertion RNA editing in vitro.** *EMBO J* 1996, **15**:6758-6765.
37. McManus MT, Adler BK, Pollard VW, Hajduk SL: **Trypanosoma brucei guide RNA poly(U) tail formation is stabilized by cognate mRNA.** *Mol Cell Biol* 2000, **20**:883-891.
38. Igo RPJ, Weston DS, Ernst NL, Panigrahi AK, Salavati R, Stuart K: **Role of uridylate-specific exoribonuclease activity in Trypanosoma brucei RNA editing.** *Eukaryot Cell* 2002, **1**:112-118.
39. Zhelonkina AG, O'Hearn SF, Law JA, Cruz-Reyes J, Huang CE, Alatorsev VS, Sollner-Webb B: **T. brucei RNA editing: action of the U-insertional TUTase within a U-deletion cycle.** *RNA* 2006, **12**:476-487.
40. Matsuo Y, Yamada A, Tsukamoto K, Tamura H, Ikezawa H, Nakamura H, Nishikawa K: **A distant evolutionary relationship between bacterial sphingomyelinase and mammalian DNase I.** *Protein Sci* 1996, **5**:2459-2467.
41. Whisstock JC, Romero S, Gurung R, Nandurkar H, Ooms LM, Bottomley SP, Mitchell CA: **The inositol polyphosphate 5-phosphatases and the apurinic/apyrimidinic base excision repair endonucleases share a common mechanism for catalysis.** *J Biol Chem* 2000, **275**:37055-37061.
42. Hertz-Fowler C, Peacock CS, Wood V, Aslett M, Kerhornou A, Mooney P, Tivey A, Berriman M, Hall N, Rutherford K, Parkhill J, Ivens AC, Rajandream MA, Barrell B: **GeneDB: a resource for prokaryotic and eukaryotic organisms.** *Nucleic Acids Res* 2004, **32**:D339-D343.
43. Dalgaard JZ, Klar AJ, Moser MJ, Holley WR, Chatterjee A, Mian IS: **Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family.** *Nucleic Acids Res* 1997, **25**:4626-4638.
44. Dalgaard JZ, Moser MJ, Hughey R, Mian IS: **Statistical modeling, phylogenetic analysis and structure prediction of a protein splicing domain common to inteins and hedgehog proteins.** *J Comput Biol* 1997, **4**:193-214.
45. Mian IS: **Sequence, structural, functional, and phylogenetic analyses of three glycosidase families.** *Blood Cells Mol Dis* 1998, **24**:83-100.
46. Mian IS, Moser MJ, Holley WR, Chatterjee A: **Statistical modelling and phylogenetic analysis of a deaminase domain.** *J Comput Biol* 1998, **5**:57-72.
47. Moser MJ, Holley WR, Chatterjee A, Mian IS: **The proofreading domain of Escherichia coli DNA polymerase I and other DNA and/or RNA exonuclease domains.** *Nucleic Acids Res* 1997, **25**:5110-5118.
48. System SAM: <http://www.so.eucsc.edu/research/compbio/sam.html>. URL 2005 [<http://www.so.eucsc.edu/research/compbio/sam.html>].
49. Salavati R, Ernst NL, O'Rear J, Gilliam T, Tarun SJR, Stuart K: **KREPA4, an RNA binding protein essential for editosome integrity and survival of Trypanosoma brucei.** *RNA* 2006, **12**:819-31.
50. Sali A, Blundell TL: **Comparative protein modelling by satisfaction of spatial restraints.** *J Mol Biol* 1993, **234**:779-815.
51. Ginalski K, Elofsson A, Fischer D, Rychlewski L: **3D-Jury: a simple approach to improve protein structure predictions.** *Bioinformatics* 2003, **19**:1015-1018.
52. Bowie JU, Luthy R, Eisenberg D: **A method to identify protein sequences that fold into a known three-dimensional structure.** *Science* 1991, **253**:164-170.