J Mol Model (2010) 16:61–76
DOI 10.1007/s00894-009-0506-1

ORIGINAL PAPER

Kinetoplastid RNA editing ligases 1 and 2 exhibit different electrostatic properties

Alireza Shaneh · Reza Salavati

Received: 18 December 2008 / Accepted: 23 March 2009 / Published online: 27 May 2009

Abstract Kinetoplastid RNA editing ligases 1 and 2 (KREL1 and KREL2) share a significant degree of sequence homology. However, biochemical experiments have reported that KREL1 and KREL2 differ in their functional roles during the RNA editing process. In this study, we hypothesize that dissimilar roles for KREL1 and KREL2 proteins arise from their different physicochemical characteristics. To test our hypothesis at sequence level, we plotted theoretical titration curves for KREL1, KREL2 and their binding partner proteins. The plots showed a lower isoelectric point for KREL1 compared to that for KREL2 as well as more relative alkalinity and acidity for binding partner proteins of KREL1 and KREL2 at net charge zero, respectively. At structure level, based on the available high resolution structure of KREL1 N-terminal domain and strong sequence similarity between KRELs and other ligases, we built the homology model of KREL2 N-terminal domain. Using Poisson-Boltzmann continuum approach, we calculated the electrostatic potential isosurfaces of KREL1 structure and KREL2 model. KREL1 and KREL2 coordinates differed in their electrostatic isopotential patterns. A wider negative patch on the surface of KREL1 suggests differential affinity for another protein compared to KREL2. In contrast, a larger positive patch on the KREL2 surface predicts its differential affinity and/or specificity for its RNA substrate. Subsequently, we employed *in silico* mutational scanning and identified the surface-exposed residues contributing to the long-range electrostatic energy of KRELs. We predict that two structurally conserved loops of KRELs, not previously reported in the literature, also recognize their RNA substrates. Our results provide important information about the physicochemical properties of RNA editing ligases that could contribute to the ligation step of RNA editing.

Keywords Editosome · Electrostatic potential energy · Homology modeling · Kinetoplastid RNA editing ligases · Poisson-Boltzmann equation · RNA editing

Introduction

Trypanosomatidae, a family of kinetoplastids, consist of pathogenic agents of human African trypanosomiasis, Chagas disease, and leishmaniasis in Africa, Latin America, and much of tropical and subtropical countries. *Trypanosoma brucei*, *T. cruzi* and *Leishmania major* are the three major trypanosomatids that undergo remarkable changes in their morphology during their insect and mammalian bloodstream life stages that is coupled to their metabolism [1, 2]. While the major source of energy in the insect stage is oxidative phosphorylation, glycolysis is the main source of energy in the mammalian bloodstream stage. In trypanosomatids, the remodeling of the mitochondrion structure within two life stages requires different levels of gene expression which is regulated by RNA editing [1]. In trypanosomatids, RNA editing is a post-trascriptional modification which alters the mitochondrial mRNA transcripts by insertion or deletion of uridylates (U-insertion and -deletion) [3], creating mature functional mRNAs for multiple components of mitochondrial oxidative phosphorylation system [4, 5]. This process is mediated by small

Electronic supplementary material The online version of this article (doi:10.1007/s00894-009-0506-1) contains supplementary material, which is available to authorized users.

A. Shaneh · R. Salavati (✉)
Institute of Parasitology, McGill Centre for Bioinformatics,
21–111 Lakeshore Road,
Ste. Anne de Bellevue, Quebec H9X 3V9, Canada
e-mail: reza.salavati@mcgill.ca

RNAs known as guide RNAs (gRNAs) [6]. During the RNA editing process, gRNAs form a duplex with their 'cognate' mRNA precursors and specify U-insertion or deletion mediated by a series of coordinated catalytic steps of multiprotein complex, known as editosome (for recent reviews, see [7–10]). Twenty proteins have been found to be associated with the editosome complex. Endonuclease, exonuclease, terminal uridylyl transferase (TUTase), ligase, and helicase activites are among discovered enzymatic activities of the editosome (for review, see [9]). In addition, mitochondrial RNA binding protein (MRP) complex consisting of mitochondrial RNA binding proteins 1 and 2 enhances the formation of RNA editing components in trypanosomatids [7, 11–13].

In this paper, we follow the nomenclature after Stuart and colleagues [9]. As a prefix to acronyms, "K" refers to "kinetoplastid" and includes *T. brucei*, *T. cruzi* and *L. major*. Hence, we use "K" prefix to discuss the respective proteins independent of species. To address specific trypanosomatid species, we use the abbreviated prefix of the trypanosomatid followed by the acronym name of the protein. Kinetoplastid RNA editing ligases 1 and 2 (KREL1 and KREL2) have been characterized in the editosome complex [14–17]. The N-terminal domain of KREL1 and KREL2 share signature motifs with covalent nucleotidyl transferase superfamily which includes DNA ligases and mRNA capping enzymes [18]. In the presence of ATP and Mg^{2+}, the N-terminal catalytic domains of KREL1 and KREL2 interact with gRNA-mRNA duplex to join the pre-cleaved mRNA fragments [16, 19]. Gao and Colleagues [20] have proposed that the C-terminal domains of KREL1 and KREL2 may integrate the RNA editing ligases to the editosome complex. Knockdown expression as well as mutated allele overexpression studies has shown that KREL1 is essential for completion of RNA editing process [21, 22]. On the other hand, knockdown expression of KREL2 has had no effect on RNA editing [23, 24]. The multiple sequence alignment of N-terminal domains of RNA editing ligases [25] has shown that *Tb*REL1 and *Tb*REL2 have 41% amino acid sequence identity and 61% sequence similarity suggesting that *Tb*REL1 may compensate for the loss of *Tb*REL2. On the other hand, tandem affinity purification, co-immunoprecipitation, yeast two-hybrid analysis, and biochemical approaches have suggested deletion and insertion roles for KREL1 and KREL2, respectively [17, 26]. The experimental and *in silico* studies have proposed two separate subcomplexes for RNA editing in trypanosomatids (Fig. 1) [26]. Deletion RNA editing subcomplex consists of KREL1, Kinetoplastid RNA editing protein A2 (KREPA2), and kinetoplastid RNA editing exonuclease 2 (KREX2). Insertion RNA editing subcomplex includes KREL2, Kinetoplastid RNA editing protein A1 (KREPA1), and kinetoplastid RNA editing TUTase 2 (KRET2). Oligonucleotide or oligosaccharide binding (OB-) folds are structurally conserved motifs necessary for enzymatic activities of DNA and RNA ligases [27, 28]. However, the secondary structure analysis of RNA editing ligases shows that KRELs lack the OB-fold in their C-terminal domains and that KREPA1 and KREPA2 have OB-fold in their C-terminal regions. Accordingly, it is proposed that KREPA2 and KREPA1 may provide the necessary OB-fold for KREL1 and KREL2 in *trans*, respectively, to perform the ligation steps of RNA editing [26]. Figure 1 shows that kinetoplastid RNA editing protein A6 (KREPA6), another protein associated with editosome, may interact with KREPA1 and KREPA2 proteins. However, Schnaufer and colleagues [26] have found the interaction of KREPA6 variable between experiments. Accordingly, we have not considered KREPA6 in the present study.

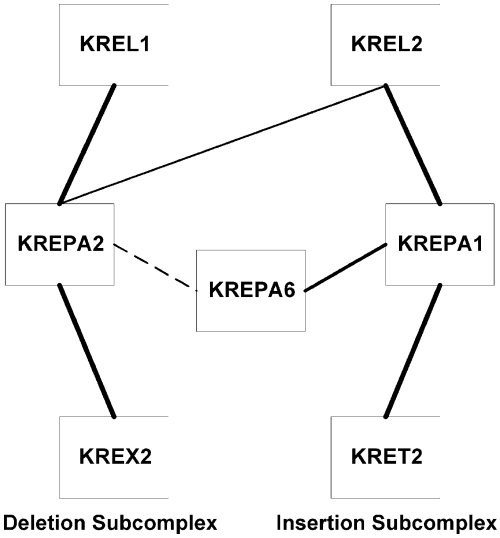

Fig. 1 The deletion and insertion subcomplexes of KRELs. The proposed subcomplexes of KREL1 and KREL2 along with their binding protein partners, KREPA2 and KREPA1 [26]. KREX2 and KRET2 are kinetoplastid RNA editing exonuclease and TUTase, respectively. "Thick solid lines indicate interactions shown by yeast two-hybrid, coimmunoprecipitation, and genetic studies. Thin solid lines indicate reproducible interactions in the yeast two-hybrid assay" [26]. Dotted line between KREPA6 (another editosome protein) and KREPA2 indicates that the interaction between two proteins with yeast two-hybrid experiment had some variability in yeast two-hybrid assay

The crystal structure of N-terminal catalytic domain of *Tb*REL1 in complex with ATP and Mg^{2+} has recently been solved at the atomic resolution of 1.2 Å, the highest resolution of this family to date [25]. The crystal structure of *Tb*REL1 N-terminal domain has revealed the hydrogen bonds between three trapped water molecules and the adenosine moiety of ATP in the deep pocket of the catalytic site. In addition, three other water molecules form a perfect octahedral through their hydrogen bonds to trap Magnesium ion, which in turn coordinates β- and γ- phosphorus

oxygen atoms of ATP. This water-mediated hydrogen bond network has not been observed in any other member of the covalent nucleotidyl transferases including DNA ligases and mRNA capping enzymes [25]. Using molecular dynamics simulation, Amaro and colleagues [29] have calculated the ensemble electrostatic potentials for *Tb*REL1 N-terminal domain structure, and they have confirmed the prediction of Deng and colleagues [25] that two conserved loops, formed by Y167–K177 and V190–Y200, may be involved in RNA recognition and protein-protein interaction.

While KREL1 and KREL2 protein sequences share a high degree of identity and similarity, they differ in the way they join mRNA fragments [17, 19, 30]. Consequently, an examination of previous literature raises a new hypothesis that KREL1 and KREL2 may differ in their physical-chemical properties and necessitates a new study encompassing the physicochemical characteristics of KREL1 and KREL2 at protein sequence and structure levels.

Electrostatic property exemplifies a physical-chemical characteristic of proteins. Accordingly, to examine the differences between physical-chemical characteristics of KREL1 and KREL2, we investigated their electrostatic properties. For this purpose, we built the comparative homology model of N-terminal catalytic domain of *Tb*REL2 and compared the electrostatic potential energy patterns of *Tb*REL2 model with *Tb*REL1 N-terminal structure. We find that such a comparison gives a specific insight into the structural motifs on the surface of the RNA editing ligases responsible for protein or nucleic acid recognition.

Materials and methods

Dataset preparation

We collected the amino acid sequences of KREL1 and KREL2 and their binding partner proteins from *T. brucei*, *T. cruzi*, and *L. major* genomes deposited in GeneDB database Version 2.1 [http://www.genedb.org] (Table 1). KREL1, KREL2, KREPA1, and KREPA2 proteins with a mitochondrial import signal are imported into the mitochondria and are cleaved to produce a functional protein [7]. Therefore, the predicted mitochondrial import signals from these proteins were removed for this study. We obtained the crystal structure of the N-terminal catalytic domain of *Tb*REL1 [25] from Protein Data Bank (PDB) [PDB:1XDN] [31]. We exclude the 12 missing residues reported in the PDB file from our analysis because they belonged to either N- or C-terminal ends of the catalytic domain. We processed the PDB file by visually inspecting incorrect bonds. Furthermore, we treated the selenomethionine residues by mutating them to methionine. All ionizable sidechains were protonated at standard physiological pH. Using Discovery Studio suite (www.accelrys.com), we visually inspected the protonated side-chains to avoid flipped states of charged or polar side-chains.

Table 1 The sequences of proposed deletion and insertion editing subcomplexes

Family	Protein	GeneDB ID
KREL1	*Tb*REL1	Tb09.160.2970
	*Tc*REL1	Tc00.1047053510155.20
	*Lm*REL1	LmjF01.0590
KREL2	*Tb*REL2	Tb927.1.3030
	*Tc*REL2	Tc00.1047053506363.110
	*Lm*REL2	LmjF20.1730
KREPA1	*Tb*REPA1	Tb927.2.2470
	*Tc*REPA1	Tc00.1047053503515.20
	*Lm*REPA1	LmjF02.0410
KREPA2	*Tb*REPA2	Tb10.6 k15.2310
	*Tc*REPA2	Tc00.1047053507611.398
	*Lm*REPA2	LmjF36.6930
KREX2	*Tb*REX2	Tb10.70.3850
	*Tc*REX2	Tc00.1047053508153.1100
	*Lm*REX2	LmjF03.0620
KRET2	*Tb*RET2	Tb07.27 M11.900
	*Tc*RET2	Tc00.1047053503579.150
	*Lm*RET2	LmjF26.0390

KREL refers to kinetoplastid RNA editing ligase, *Tb* to *Trypanosoma brucei*, *Tc* to *Trypanosoma cruzi*, *Lm* to *Leishmania major*, KREPA to kinetoplastid RNA editing protein A, KREX2 to Kinetoplastid RNA editing exo-ribonuclease 2, and KRET2 to Kinetoplastid RNA editing 3′-terminal uridylyl transferase 2

Calculation of theoretical protein charge distribution

Using Protein Workbench [http://www.clcbio.com], we calculated the charge distribution for KREL1, KREL2 and their binding partner proteins using Henderson-Hasselbalch equations for different values of acid dissociation constants (pK_α) of negatively and positively charged amino acids [32, 33]. The net charge of the proteins were plotted versus pH values ($0 \leq pH \leq 14$), and the theoretical isoelectric point was observed at the intersection of the charge distribution curve and the abscissa.

Homology modeling of N-terminal domain of *Tb*REL2

The N-terminal catalytic domain of *Tb*REL2 (residues G20-G285), hereafter termed "TbREL2-NTERM" in this study, was used as the query sequence to search the National Center for Biotechnology Information (NCBI) Non-Redunadant (NR) database (NCBI-NR). We used PSI-BLAST algorithm [36] to search for remote homologous

sequences. The E-value cut-off for each iteration of PSI-BLAST was set to 5×10^{-4}, and the scoring matrix was set to BLOSUM62 [37].

The high scored sequences found from PSI-BLAST search (Table 2) were then aligned using MUSCLE algorithm [38]. MUSCLE algorithm uses profile alignment of distance pairs of sequences along with an Unweighted Pair Group Method with Arithmetic Mean (UPGMA) tree [39, 40] for each pairwise alignment. The optimum alignment reaches when the log-expectation score satisfies. MUSCLE alignment has outperformed the other methods such as T-COFFEE and MAFFT in terms of accuracy and speed [38, 41, 42]. The multiple sequence alignment obtained from MUSCLE algorithm constituted the profile alignment for building a working model for TbREL2-NTERM. The closest homolog of KRELs, T4Rnl2 had low sequence identity with TbREL2-NTERM.

To examine whether the solved structure of T4Rnl2 [PDB:1S68] along with *Tb*REL1 N-terminal structure [PDB:1XDN] can guide the multiple sequence alignment used for modeling *Tb*REL2-NTERM (Fig. 2), we built UPGMA phylogram tree after MUSCLE alignment of KRELs, [PDB:1XDN] and [PDB:1S68]. The phylogram tree was cross-validated 100 times in a bootstrap analysis, and [PDB:1S68] was found to be a remote structural neighbor of *Tb*REL2. Consequently, we considered *Tb*REL1 N-terminal structure as the only structural template to guide the multiple sequence alignment profile toward building the comparative model of TbREL2-NTERM. We employed MODELLER [43] to build the homology model of TbREL2-NTERM. MODELLER builds a three-dimensional model of a protein based on comparative multiple sequence alignment and known structural homologs as templates to guide sequence alignment. We built 42 models of *Tb*REL2 and refined the loops in the models based on discrete optimized protein energy (DOPE) that is a statistical potential for protein model evaluation [44] and the deviation of the models from the crystal structure. We evaluated the models using Profiles-3D [45]. As Profiles-3D parameters, the smoothing-per-residue window size was set to 10. To calculate the scores, Kabsch-Sander algorithm [46] was selected as secondary structure prediction definition. The Profiles-3D score for the best model, hereafter called *Tb*REL2 N-terminal model, was 124.800. The score was higher than the expected maximum score, estimated by Profiles-3D (expected minimum score = 52. 354 and expected maximum score = 116.343). This assessment confirmed the reliability of the model. The model, therefore, was considered as a working model for structural analysis in this study (see Supplementary Figure 1). Figure 3 displays the secondary structure of the *Tb*REL2 N-terminal domain model. 3DMA program, distributed by Accelrys, aligns the backbones of one or more proteins by segment matching of similar C_α pseudotorsion angle. We superimposed the C_α atoms of crystal and model coordinates using 3DMA program. Following the structure-based alignment, we recorded the root mean square deviation (RMSD), as a measure of deviation between the model and crystal coordinates. Aligned by C_α backbone alignment over 244 residues, the working model deviated 0.50 Å from the crystal template.

Electrostatic potential calculation

To calculate electrostatic potential surfaces, we numerically solved the Poisson-Boltzmann equation using DelPhi package [47–50]. Employing a continuum approach, DelPhi calculates the electrostatic surface potentials by assigning different dielectric constants to the solvent and protein with the protein surface as the dielectric boundary. We included ATP and Mg^{2+} in the electrostatic calculations. Electrostatic potential surface with focusing was the strategy employed in this study. Using 50% fill, we run the potential calculations on a coarse grid with Debye-Hückel conditions. The accuracy of the generated grids were increased by a second finite difference lattice, and 90% of the molecules occupied the grids. The dielectric constants of the solvent and protein were assigned to 80

Table 2 The high scored sequences obtained from PSI-BLAST search of TbREL2-NTERM

Hit	Accession #	E-value	Length	Identity (%)
T. brucei REL1	EMBL:AAG27062	2.26693×10^{-99}	272	47
T. cruzi REL1	EMBL:AAR10841	3.06529×10^{-98}	272	45
T. cruzi REL2	EMBL:AAR10840	3.58043×10^{-109}	263	82
L. major REL1	EMBL:AAR10824	5.31657×10^{-98}	271	44
L. major REL2	EMBL:AAR10823	3.76424×10^{-109}	264	70
T. brucei REL1 (N-domain structure)	PDB:1XDN	2.02205×10^{-94}	272	46
T4 Rnl 2 (Without RNA)	PDB:1S68	2.87753×10^{-57}	273	24
T4 Rnl2 (partially length with RNA)	PDB:2HVR	1.91977×10^{-30}	274	23
T4 Rnl2 (full length with RNA)	PDB:2HVQ	2.50657×10^{-31}	274	23

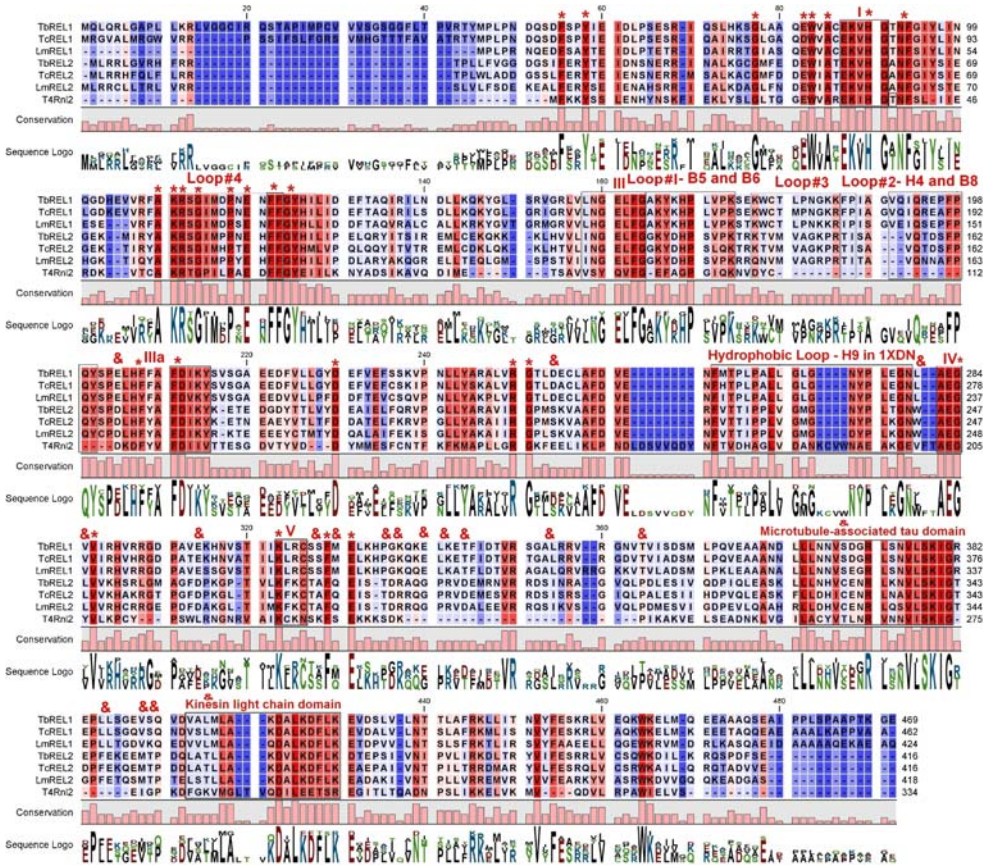

Fig. 2 Multiple sequence alignment profile of KREL proteins. KREL1 and KREL2 proteins are aligned with bacteriophage T4 RNA ligase 2 (T4 Rnl2). The five ligase signature motifs are labeled as I, III, IIIa, IV and V. The color intensity indicate the degree of conservation. The highly conserved amino acids are shown in red and the poorly conserved residues in blue. The residues conserved among KRELs and T4 Rnl2 are labeled with star (*). The columns whose corresponding amino acids are conserved within each KREL family are labeled by ampersand (&). For example, the column position 205 in *Tb*REL1 shows that glutamic acid (**E**) residues are conserved in KREL1 proteins, whereas in KREL2 family, aspartic acid residues (**D**) are conserved. The loops reported by Deng and coworkers [25] are blocked and labeled as Loop#1 (Y167-K177) and Loop#2 (V190-Y200 in *Tb*REL1 N-terminal structure). Loop#1 is located between β-stands 5 and 6 on *Tb*REL1 N-terminal structure. Loop#2 is located between α-helix 4 and β-strand 8. The hydrophobic loop (H9) unique to trypanosomatids is also shown by a block in the alignment. The Loop#3 (W178-P188 in *Tb*REL1 N-terminal structure) and Loop#4 (R111-Y124 in *Tb*REL1 N-terminal structure) identified in this study are also labeled. Microtubule-associated tau [34] and Kinesin light chain [35] domains are indicated in the alignment. The numbers on top of the residues indicate the ruler, and the numbers at the end of each sequence indicate the position of the last amino acid in that sequence. For example, the last amino acid in the first row for *Tb*REL1 sequence refers to N99

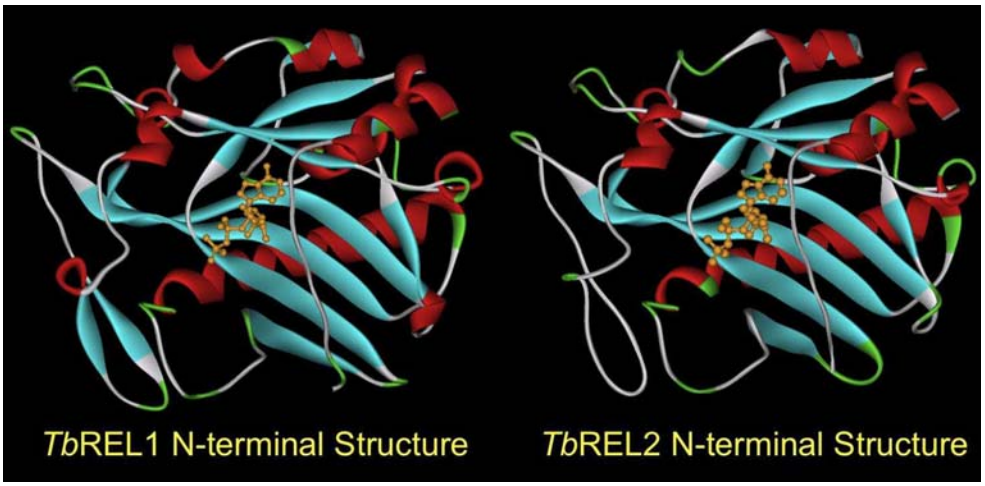

Fig. 3 *Tb*REL1 N-terminal structure and *Tb*REL2 N-terminal model. *Tb*REL1 and *Tb*REL2 coordinates are shown as solid ribbon diagrams. The molecules are in the same orientation. In the structures, the ATP ligand is shown in ball-and-stick display and color-coded as golden

and 4 respectively. The solvent ionic strength was set to 0.145 M, and the ion exclusion radius was set to 2 Å. The electrostatic potential values were computed by mapping onto a 101×101×101 cubic grids with a final grid resolution of 1 Å by grid unit. Net charges were taken from Merck Molecular Force Field (MMFF94) [51–57]. We generated the isopotential surfaces at $-1\frac{kT}{e}$ and $+1\frac{kT}{e}$, where k is the Boltzmann constant, T is the temperature (K), and e is the charge of one electron (C).

Furthermore, we employed protein interaction similarity analysis (PIPSA) [58–61] to cluster the crystal structure of *Tb*REL1 and *Tb*REL2 based on their electrostatic similarity. We used Hodgkin similarity index [62] as a metric for electrostatic similarity between proteins.

Results and discussion

In this section, we will discuss the results of calculating theoretical protein charge distribution of KREL1, KREL2 and their binding partner proteins. In addition, we will analyze the structural difference between N-terminal domains of *Tb*REL1 and *Tb*REL2 according to their electrostatic potential surfaces.

Theoretical protein charge distribution comparison

To investigate the protein net charge as a function of pH in each proposed subcomplex, we calculated titration curves for KREL1 and KREL2 proteins (Fig. 4(a)) and for their binding partner proteins (Fig. 4(b), (c)). Consequently, it was possible to obtain the theoretical pI for each protein of the subcomplexes. The different pI values observed for KREL1 and KREL2 proteins (Fig. 4(a)) show relative acidity for KREL1 proteins compared to relative alkalinity for KREL2 proteins. The results of theoretical pI values for KREPA1 show relative acidity compared to relative alkalinity for their proposed binding partners, KREL2 proteins (Fig. 4(a), (b)). On the other hand, we observed that KREPA2 proteins show relative alkalinity compared to the relative acidity of their proposed interacting partners, KREL1 proteins. In addition, the pI values for KREX2 proteins indicate relative acidity compared to the relative alkalinity of their suggested binding partners, KREPA2

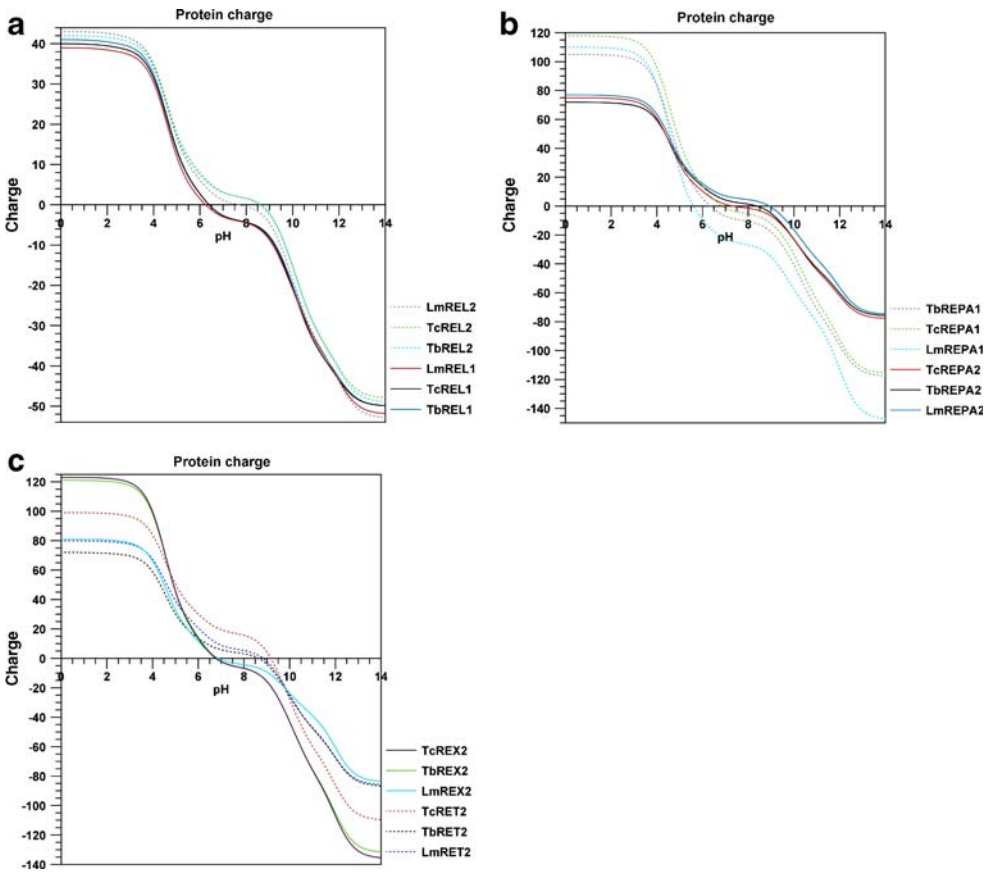

Fig. 4 Theoretical charge distribution of (**a**) KREL1, KREL2, (**b**) KREPA1, KREPA2, and (**c**) KREX2 and KRET2 sequences. The ordinate shows the calculated net charge on the proteins. The absscisa indicates the pH values. To name KREL proteins, we followed Stuart and colleagues [9]. For example, *Tb*REL1 represents *T. brucei* RNA editing ligase 1. "K" refers to the kinetoplastids discussed in this paper and encompasses *T. brucei* (Tb), *T. cruzi* (Tc) and *L. major* (Lm)

proteins (Fig. 4(b), (c)). The results of theoretical pI values also show relative alkalinity for KRET2 proteins compared to the relative acidity for their proposed interacting partners, KREPA1 proteins.

These observations suggest a potential acid-base complementarity that could play a role in protein-protein interactions within the proposed RNA editing insertion and deletion subcomplexes (Fig. 5) [26].

Electrostatic isopotential surface analysis of *Tb*REL1 and *Tb*REL2 coordinates

The difference in calculated pI between the KREL1 and KREL2 N-terminal domains suggested that their electrostatic potential profiles would be different. To test this hypothesis, using a strong degree of homology between the KREL1 and KREL2 sequences and available crystal structure of TbREL1 N-terminal domain, we built various models of of TbREL2 N-terminal domain based on different structural templates, especially, PDB:1XDN and PDB:1S68 (Table 2). After verifying the models with Profiles-3D scores, we selected the working model whose root mean square deviation (RMSD) of backbone C_α deviated 0.50 Å from the available crystal of PDB:1XDN.

We calculated electrostatic isopotential surfaces at $+1\frac{kT}{e}$ and $-1\frac{kT}{e}$ for *Tb*REL1 and *Tb*REL2 N-terminal domain coordinates. To better compare the isopotential distributions, we positioned the domains at the same orientation (Fig. 6a, b). Figure 6c also displays the structural backbone alignment of the *Tb*REL1 and *Tb*REL2 N-terminal catalytic domains with the same orientation as that of electrostatic isopotential surfaces. Our analysis indicated that the charge distribution patterns on the surface of the proteins were different (Fig. 6a, b). On *Tb*REL1 N-terminal structure, a large negative patch predominated. The secondary structure corresponding to this patch consists of an antiparallel β-sheet (labeled "Negative sheet") and a loop (F262-L282; labeled "H9") which is unique in trypanosomatids (Fig. 2). Deng et al. [25] argue that this loop may promote protein-protein interaction. Our observation of the electropositive and electronegative regions in vicinity of the H9 loop (Fig. 7) shows that the loop is more surrounded by negatively charged patches in *Tb*REL1. This observation can be supported by extended mutational analysis on H9 loops to determine their possible interaction with KRELs binding partners.

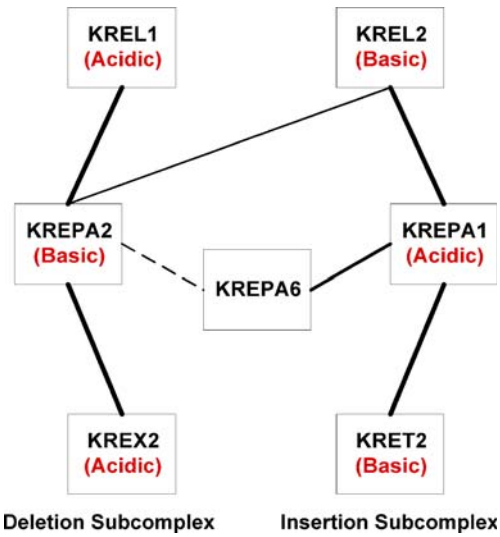

Fig. 5 Relative Acid-base complementarity observed in the proposed deletion and insertion subcomplexes. The results of theoretical pI values for KREL1, KREL2, KREPA1, KREPA2, KRET2, and KREX2 with the proposed acid-base complementarity of the interacting proteins are shown. In the proposed RNA editing deletion and insertion subcomplexes, "Acidic" and "Basic" labels indicate the relative acidity or alkalinity in each binary interaction

Furthermore, we observed a negatively charged surface in a helical motif, labeled as Helical loop (V217-F224 in Figs. 2 and 6c). Although the structural motif corresponding to this particular negative region was highly conserved, we did not detect a similar patch on the TbREL2 N-terminal surface (Fig. 6a, b). Similarly, the negative sheet in the vicinity of highly conserved H9 loop contributed less to the negative patch of *Tb*REL2 N-terminal model (Fig. 6b). These results predict that KREL1 and KREL2 proteins may have differential affinities for their binding partners.

Comparison of the electrostatic potentials associated with corresponding structural motifs of *Tb*REL1 and *Tb*REL2 N-terminal coordinates also revealed a large positive patch on the surface of both proteins (Fig. 6a, b). The large positively charged region encompassed two unique and structurally conserved loops in trypanosomatids. The first loop, hereafter called Loop#1, is formed by Y167-K177 residues of *Tb*REL1 N-terminal structure (Figs. 2 and 6c). The second loop, Loop#2, is formed by V190-Y200 residues of *Tb*REL1 N-terminal structure. Deng and colleagues [25] have predicted that Loop#1 and Loop#2 may interact with RNA substrate [25]. The overlapped positive and negative contours associated with Loop#2 supports Deng et al.'s prediction. Electrostatic isosurface analysis also predicted two structurally conserved loops in the N-terminal coordinates of *Tb*REL1 and *Tb*REL2. The loops are labeled Loop#3 and Loop#4 (Fig. 6c) and are formed by W178-P188 and R111-Y124 amino acid chains (Fig. 2) in *Tb*REL1 N-terminal structure, respectively. Comparing isopotential surfaces of *Tb*REL1 and *Tb*REL2, one can conclude that these loops are associated with larger positive patches in TbREL2 model. These patches are associated with Loop #3 and Loop#4 whose function has not previously been reported. We then anticipate an RNA recognition role for Loop#3 and Loop#4 and that N-terminal domain of *Tb*REL2 may have different

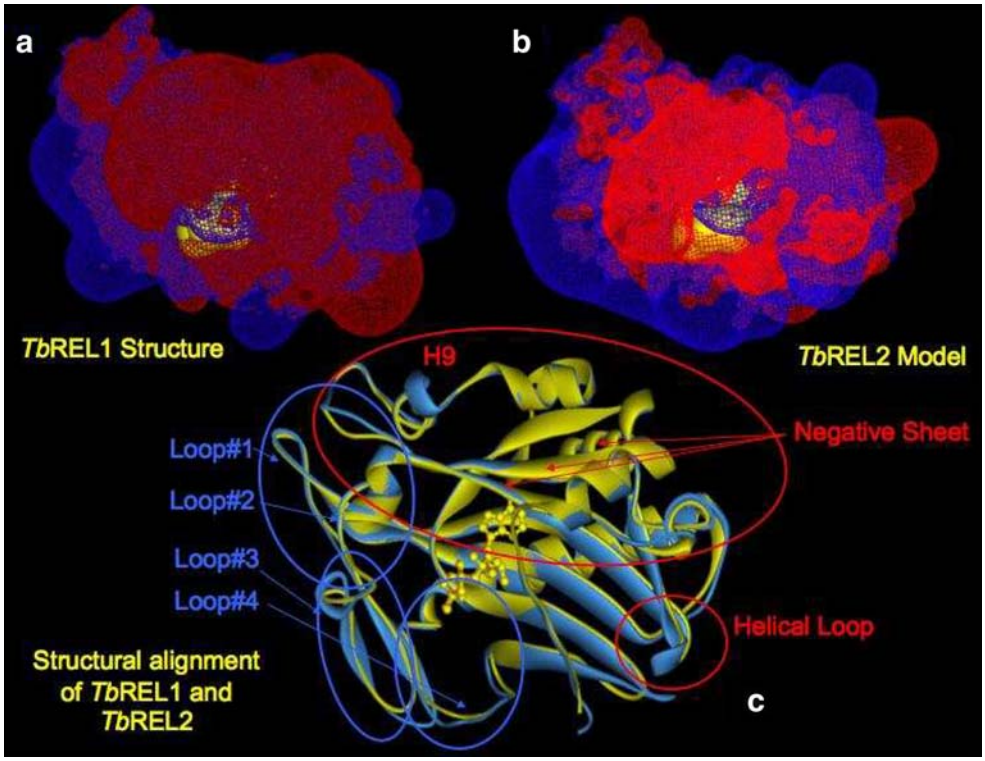

Fig. 6 Electrostatic potential isosurfaces at $+1\frac{kT}{e}$ and $-1\frac{kT}{e}$ for *Tb*REL1 (**a**) and *Tb*REL2 (**b**). The surfaces and structures are shown in the same orientation with positive charges in blue and negative charges in red. The superimposed structures of *Tb*REL1 (blue) and *Tb*REL2 model (yellow) are shown (**c**). Regions contributing to negative charges anticipated to interact with proteins are marked by red ovals. Regions contributing to positive charges anticipated to interact with nucleic acids are marked by blue ovals

affinity and/or specificity for RNA substrate compared to *Tb*REL1.

To obtain a better insight into the possible functional loops, using 3DMA module of Discovery Studio suite, we aligned the protein backbone chains of *Tb*REL1 and *Tb*REL2 N-terminal domains capturing isocontours on the protein chain of T4 RNA ligase 2 co-crystalized with DNA strands (Fig. 8a, b). Amaro and colleagues [29] have reported that Loops #1 and #2 are possibly important in RNA substrate recognition. In addition to Loop#1 and Loop#2, Loops#3 and #4 also lie on the nucleic acid, making our anticipation stronger that the two latter loops may also play roles in RNA substrate recognition.

Structure-based mutagenesis *in silico*

We used single- and multiple-site charged-to-alanine scanning *in silico* to identify the charged residues which may promote distinct positive and negative regions and contribute to protein-protein and protein-RNA interactions on

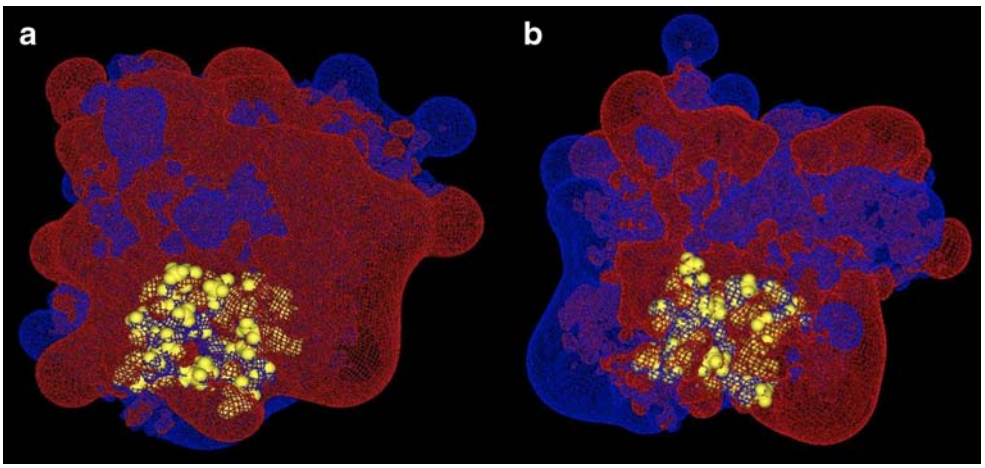

Fig. 7 H9 loop in *Tb*REL1 (**a**) and the corresponding loop in *Tb*REL2 (**b**) N-terminal domains. The residues on the loops are shown in CPK model and color coded as yellow

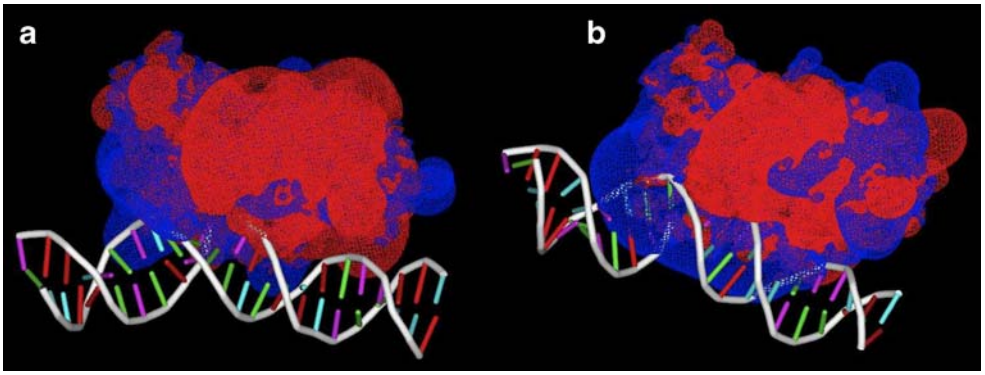

Fig. 8 Superimposition of isopotential surfaces of TbREL1 (**a**) and TbREL2 (**b**) N-terminal domains on the double stranded DNA chains of T4 RNA ligase 2

*Tb*REL1 and *Tb*REL2 protein surfaces. Alanine has a hydrophobic, uncharged methyl side-chain. Consequently, mutating a charged backbone residue to alanine disrupts or perturbs the associated charge without affecting the backbone conformation [63]. For the purpose of alanine scanning, we visually inspected the structural motifs associated with large electrostatic patches for the charged residues whose side-chains were pointing outward to the protein surfaces. We then carried out systematic replacement of the predicted charged side chains with alanine on *Tb*REL1 and *Tb*REL2 N-terminal surfaces and examined their electrostatic isosurfaces (Figs. 9, 10, 11, 12, 13 and 14). Figure 9 shows *Tb*REL1 D258A-E260A double mutation that disrupted the associated negative patch on the surface of the *Tb*REL1 N-terminal structure. Figure 9 also displays that E60A-D62A-E66A triple mutation disrupted the related negative electrostatic pattern on the surface of *Tb*REL1 N-terminal structure. In addition, triple mutation of E219A-E220A-D221A disrupted the negative patch associated with these residues and increased the neighboring positive patch (Fig. 9). E219, E220, and D221 lie in a structurally conserved and surface exposed helical loop in *Tb*REL1 N-terminal structure ("Helical Loop" in Fig. 6c). Consequently, we propose that this helical loop can potentially contribute to protein-protein interactions in *Tb*REL1. In contrast, such negatively charged residues on the conserved loop is absent in *Tb*REL2 N-terminal model. In *Tb*REL1 N-terminal structure, E278A disrupted the negative charge associated with H9 loop (Fig. 10). E278A promoted a similar electrostatic pattern to that of T219, the aligned

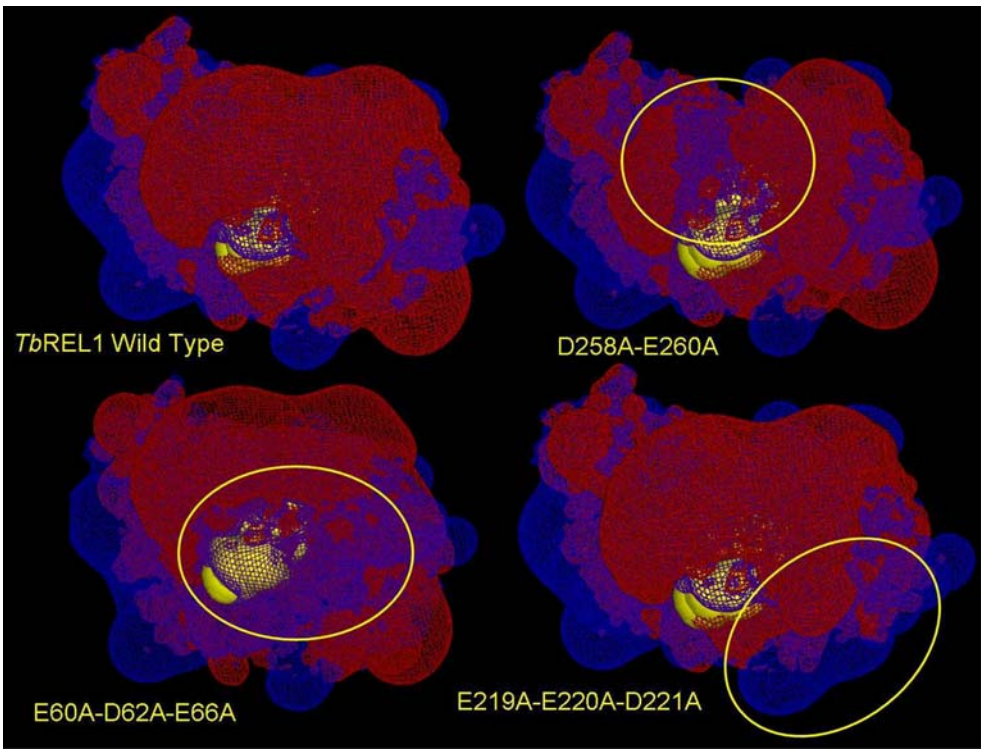

Fig. 9 The effect of double and triple mutations on *Tb*REL1 N-terminal structure. The isoelectric potential surface comparison of N-terminal catalytic domains of *Tb*REL1 wild type protein with *Tb*REL1 D258A-E260A double mutant as well as E60A-D62A-E66A and E219A-E220A-E221A triple mutants is shown. The mutant regions are marked by yellow ovals

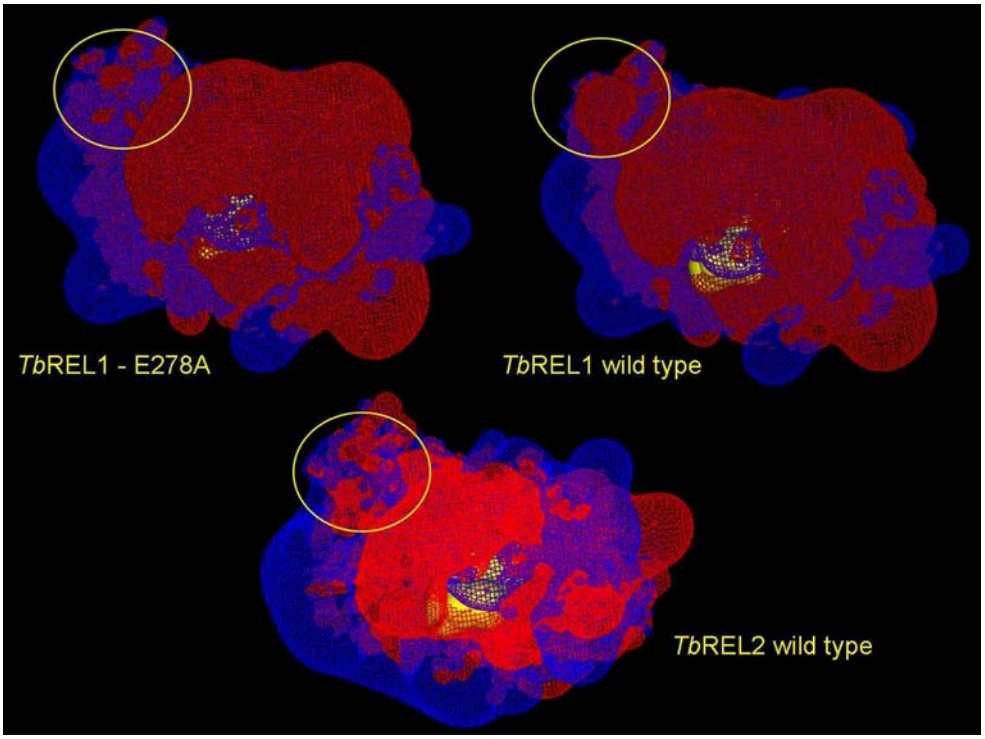

Fig. 10 E278A single mutation disrupts the corresponding negative charges. The positive long-range electrostatic isosurfaces are shown in blue and the negative charges in red for *Tb*REL1 E278A mutant, TbREL1 wild type, and *Tb*REL2 N-terminal coordinates. The region covering E278 is marked by a yellow circle

residue in *Tb*REL2 N-terminal model, that suggests different functional roles for E278 and T219. Spatially distant residues could also affect the electrostatic isopotential surfaces associated with *Tb*REL1 structure. Figure 11 illustrates that double mutation of distant residues, E253A-E315A, disrupted the corresponding negative contours. The mutagenesis effects seen for E253 and E315 of *Tb*REL1 N-terminal domain suggest that these residues may have

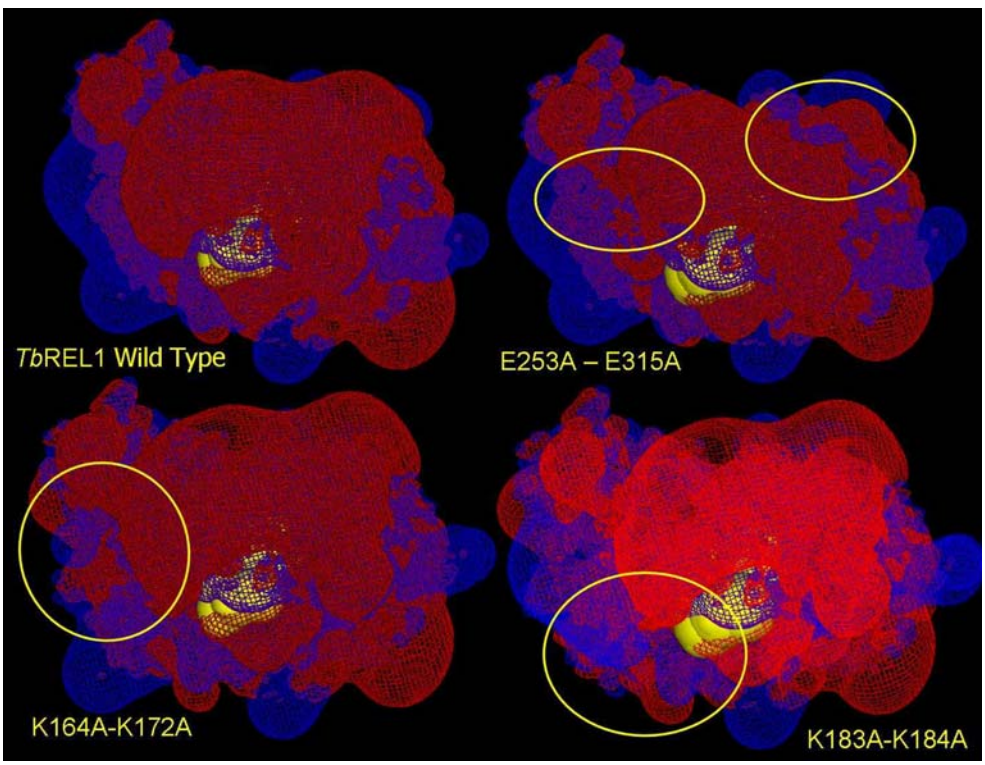

Fig. 11 The effect of distant as well as positive residue mutations on *Tb*REL1 N-terminal structure.The isoelectric potential surface comparison of N-terminal catalytic domains of *Tb*REL1 wild type model with TbREL1 E253A-E315A, K164A-K172A and K183A-K184A double mutants is shown. The mutant regions are marked by yellow ovals

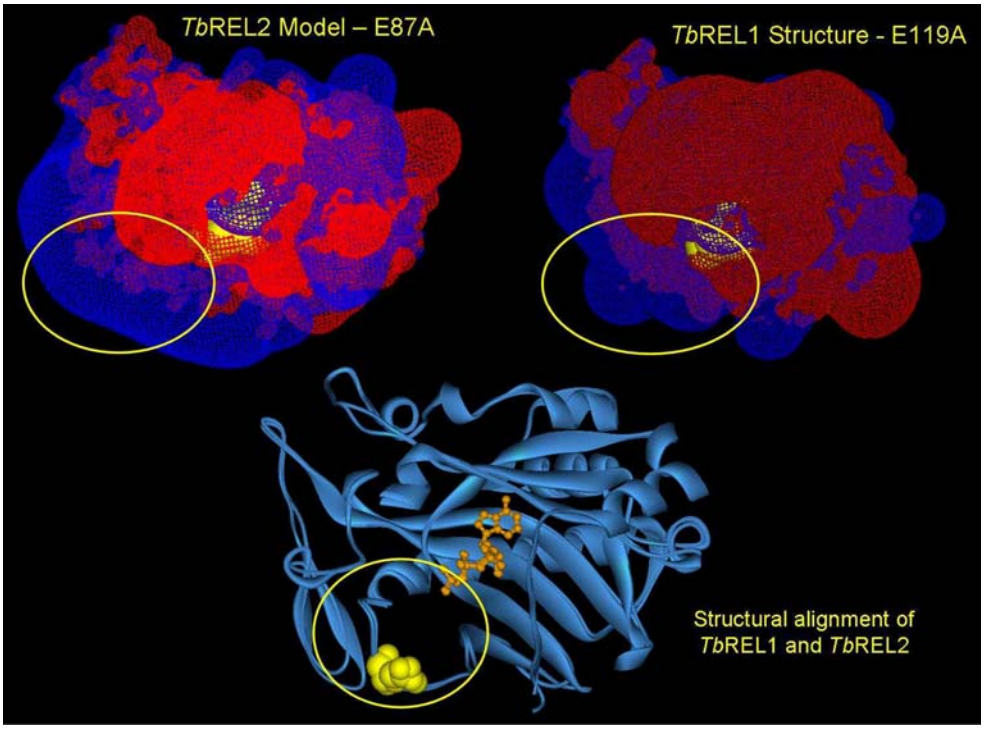

Fig. 12 The effect of single mutation of a conserved glutamic acid in Loop#4 of KRELs. The isoelectric potential surface comparison of N-terminal catalytic domains of *Tb*REL1 E119A mutant and the corresponding E87A mutant of *Tb*REL2 model is shown. The lower panel highlights the side-chain of glutamic acid residue explicitly on superposed N-terminal catalytic domains of *Tb*REL1 and *Tb*REL2 model. The mutant regions are marked by yellow ovals

functional roles. In contrast, the counterpart of the H9 loop in *Tb*REL2 N-terminal model, shows a less prominent neighboring negative patch (Fig. 6a, b).

On the surface of *Tb*REL1 N-terminal structure, K164A-K172A double mutation representing Loop#1 and Loop#2 disrupted the positively charged isosurfaces in Loop#1 more significantly than those in Loop#2 (Fig. 11). Furthermore, K183A-K184A mutation of *Tb*REL1 N-terminal structure disturbed the positive patch associated with Loop#3 (Fig. 11).

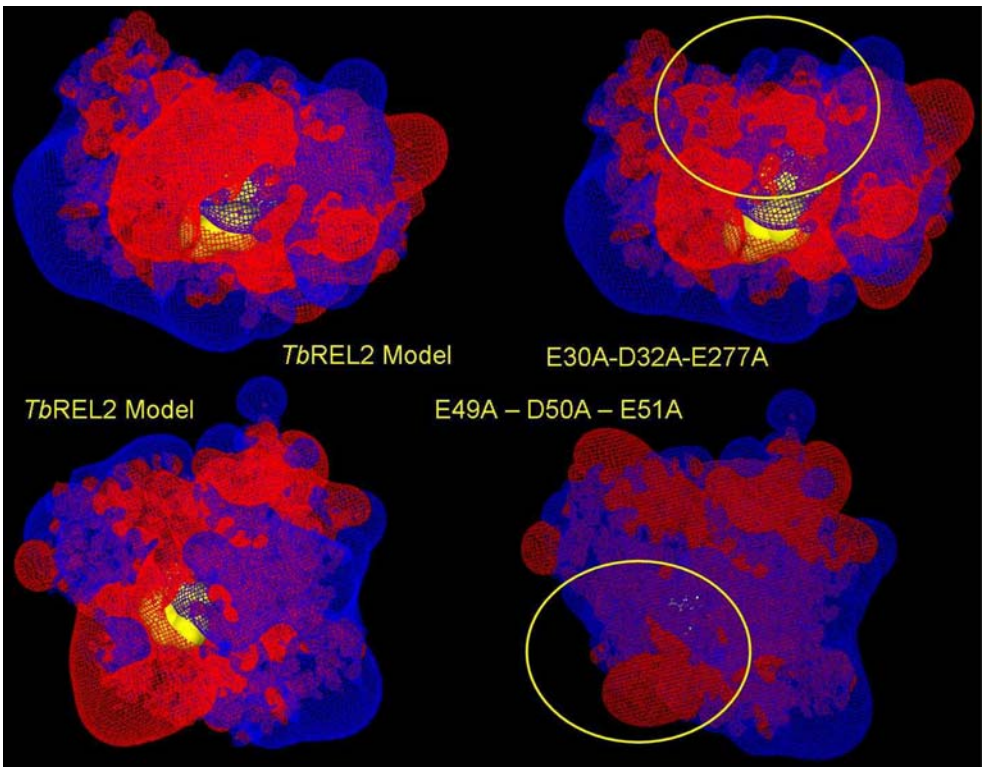

Fig. 13 The effect of negative residue mutation on *Tb*REL2 N-terminal model. The isoelectric potential surface comparison of N-terminal catalytic domains of *Tb*REL2 wild type model with *Tb*REL2 E30A-D32A-E277A and E49A-D50A-E51A triple mutants is shown. In the bottom panel, the *Tb*REL2 model is rotated 180°C around the plane facing the reader. The mutant regions are marked by yellow ovals

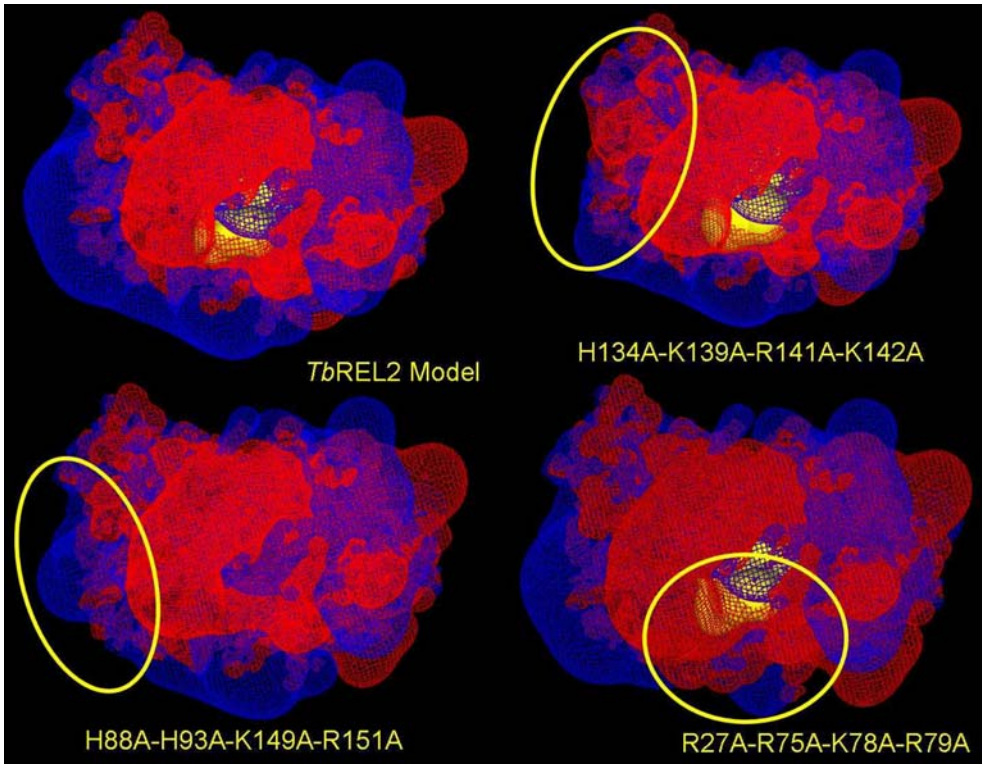

Fig. 14 The effect of positively charged residue mutation in *Tb*REL2 N-terminal model. The isoelectric potential surface comparison of N-terminal catalytic domains of *Tb*REL2 wild type model with a series of quadruple mutations in *Tb*REL2 model (H134A-K139A-R141A-K142A corresponding to loop #1, H88A-H93A-K149A-R151A corresponding to loop#3, and R27A-R75A-K78A-R79A corresponding to loop#4 in Fig. 6) is shown. The mutant regions are marked by yellow ovals

In Loop#4 of *Tb*REL1 N-terminal structure, E119A disrupted the negative charge and increased the positive contours associated with the region (Fig. 12). In contrast, E87A mutation in TbREL2 N-terminal model, the aligned counterpart of E119 in *Tb*REL1 coordinate, perturbed the associated positive patch but did not disrupt the pattern significantly (Fig. 12). Based on the effect of alanine scanning on these positively charged residues we predict that Loop#3 can contribute to negatively charged RNA substrate recognition, and along with Loop#1 and Loop#4, may form an RNA recognition motif in TbREL1 N-terminal structure.

Figure 13 shows that triple mutations of E30A-D32A-E277A and E49A-D50A-E51A significantly disrupted the large negative contours on the surface of *Tb*REL2 N-terminal model. Therefore, we predict that E30, D32, E49, D50, E51, and E277 could play a role in protein-protein interaction in *Tb*REL2 protein. Comparing our putative proposed RNA binding Loop#1, Loop#3 and Loop#4 in *Tb*REL1 coordinate with their structurally conserved counterparts in *Tb*REL2 model shows a higher abundance of positively charged residues in *Tb*REL2 protein. Subsequently, substituting the positively charged residues of Loop#1, Loop#3, and Loop#4 in *Tb*REL2 N-terminal model with alanine disrupted the respective positive patches and, in the case of R27A-R75A-K78A-R79A in Loop#4, increased the neighboring negative charges (Fig. 14). Accordingly, we can predict Loop#1, Loop#3, and Loop#4 as RNA recognition motifs of *Tb*REL2 protein.

Figure 15 illustrates the surface charged residues of *Tb*REL1 and *Tb*REL2 N-terminal coordinates whose mutations disrupted their respective electrostatic charges, and Table 3 summarizes those residues. The abundance of positively charged residues on the structurally conserved loops of *Tb*REL2 model and the effect of mutagenesis on the electrostatic patches associated with those loops suggests that *Tb*REL2 catalytic domain have possibly different affinity and/or specificity for its RNA substrate compared to *Tb*REL1.

Our data are consistent with the experimental observation of Huang and colleagues [22] who have demonstrated that *Tb*REL1 requires no guide RNA that pairs with a pre-edited mRNA transcript and that holds the ends of the mRNA fragments to complete the ligation process. In addition, Huang and coworkers [22] have shown that *Tb*REL2 specifically requires a cognate gRNA to forward the ligation step. One may think that the abundance of long-range electropositive patches on the surface of *Tb*REL2 may help the enzyme to specify the fully

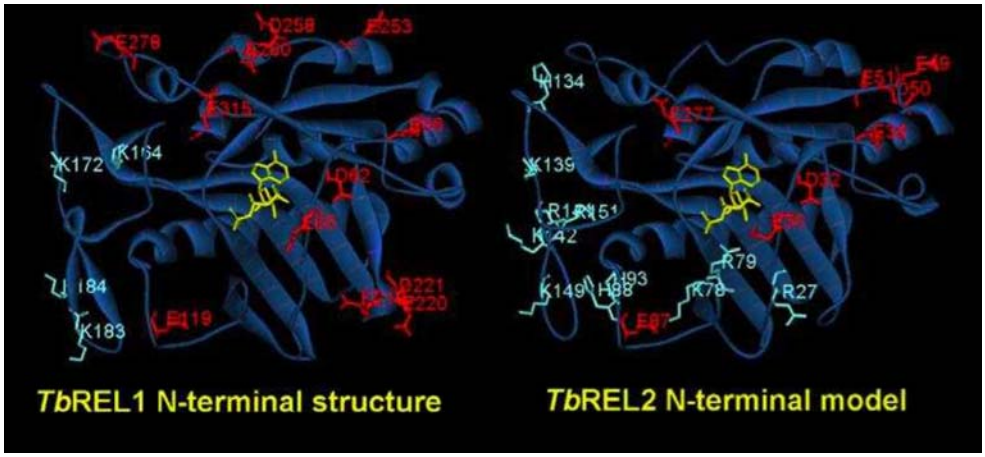

Fig. 15 The positively and negatively charged residues whose mutation disrupted the electrostatic charges associated with these residues. The mutated residues found to cancel their corresponding electrostatic charges are illustrated in *Tb*REL1 and *Tb*REL2 coordinates. The negatively charged residues are shown in red sticks and the positively charged residues in blue. ATP is displayed in stick and color-coded as yellow

matched gRNA-mRNA duplex. Therefore, based on our results that predict the roles for the surface charged residues of *Tb*REL1 and *Tb*REL2, future experimental investigations should help to determine the RNA substrate specificities for these proteins.

Analysis of electrostatic similarity between *Tb*REL1 and *Tb*REL2

To further establish whether the protein charge distribution and electrostatic isopotential surface of *Tb*REL1 and *Tb*REL2 are significantly different from one another, electrostatic properties of an ensemble of 42 modeled *Tb*REL2 structures was compared with electrostatic potential of *Tb*REL1 crystal structure (PDB: 1XDN) using PIPSA. The results of PIPSA supports our hypothesis that the two ligases differ in their electrostatic potential surfaces. Figure 16 shows that the models of *Tb*REL2 fall into one cluster. On the other hand, the the electrostatic potential of crystal structure of *Tb*REL1 (PDB:1XDN) is an outlier compared to that of other model coordinates (see Supplementary Figure 2).

Table 3 The residues whose mutation disrupted the electrostatic patterns in *Tb*REL1 and *Tb*REL2 N-terminal domains

*Tb*REL1 N-terminal domain	*Tb*REL2 N-terminal domain
E278A	E87A[a]
D258A[a]-E260A[a]	E30A[a]-D32A-E277A[a]
E60A[a]-D62A-E66A[a]	E49A-D50A-E51A[a]
E219A-E220A-D221A	R27A-R75A[a]-K78A[a]-R79A[a]
E253A-E315A[a]	H134A[a]-K139A[a]-R141A-K142A[a]
K164A[a]-K172A[a]	H88A-H93A[a]-K149A-R151A[a]
K183A-K184A	

The residues superscripted with 'a' are conserved in the multiple sequence alignment

Conclusions

In this study, we analyzed the electrostatic properties of KREL1 and KREL2 that catalyze the ligation step of RNA editing in trypanosomatids. In their N-terminal domain sequences, KRELs have five conserved motifs that belong to the covalent nucleotidyl transferase superfamily. Although *Tb*REL1 and *Tb*REL2 have 41% sequence identity and 61% sequence similarity, they exhibit different biochemical properties. Experimental assays have shown that the loss of KREL1 inhibits RNA editing process and causes the death of the parasite, whereas KREL2 loss has no effect on RNA editing. Previous study [26] proposed that KREL1 proteins participate in U-deletion, whereas KREL2 proteins contribute to U-insertion. We hypothesized that KREL1 and KREL2 have different physical-chemical properties. To test our hypothesis, we studied the theoretical charge distribution of KREL1, KREL2 and their binding partner protein sequences. Furthermore, we studied the long-range electrostatic potential energies of KRELs at structure level. The recently solved structure of *Tb*REL1 N-terminal domain enabled us to model *Tb*REL2 N-terminal domain. Therefore, in the present study, we modeled the structure of *Tb*REL2 N-terminal domain, and we used it as a working model. At the sequence level, we plotted the theoretical titration curves for KRELs and their interacting partner proteins against the pH values. The results showed that there is an acid-base complementarity supporting the previously proposed insertion and deletion RNA editing subcomplexes in trypanosomatids [26]. Our analysis showed that the acidic proteins interacted with basic

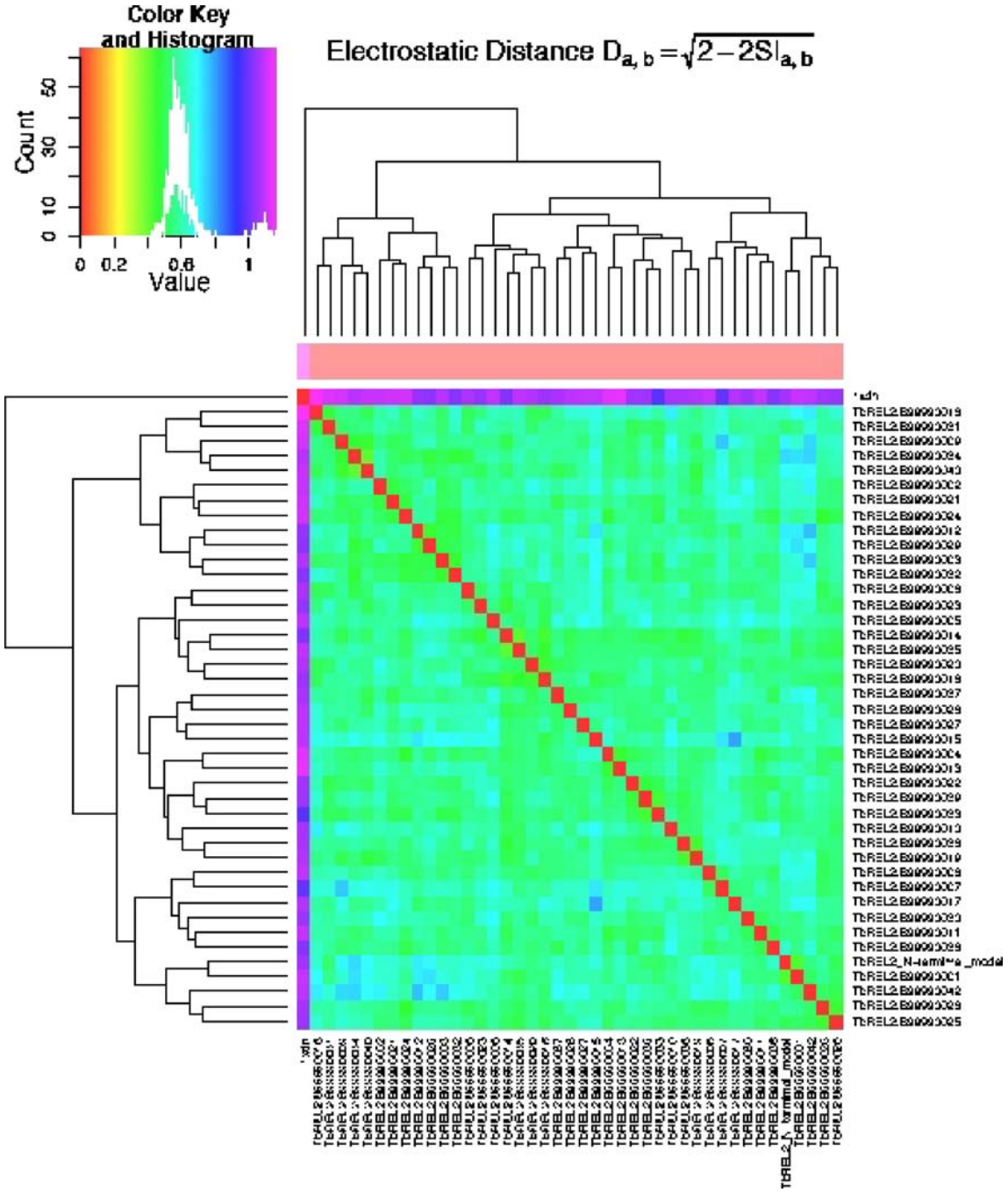

Fig. 16 Protein interaction property similarity analysis (PIPSA) heat map of the *Tb*REL2 models and *Tb*REL1 structure. The dendograms show the cluster levels of coordinates. The electrostatic distance is Hodgkin similarity index [62]. The distances are color-coded from similar (red) to dissimilar (magenta)

proteins in each subcomplex. We then anticipated that the charge distribution is related to the electrostatic properties of KRELs, and that KRELs differ in their electrostatic potential energies. To examine our anticipation, we solved the finite-difference values for electrostatic potential energies of available coordinate of *Tb*REL1 structure and *Tb*REL2 model. Intriguingly, we found that *Tb*REL1 and *Tb*REL2 show clear differences in their electrostatic patterns. While long-range negative charges predominantly cover *Tb*REL1 catalytic domain, positive charges encompass *Tb*REL2 N-terminal structure. Electrostatic similarity analysis clustered the models of *Tb*REL2 from the crystal structure of *Tb*REL1, supporting our hypothesis that the two ligases exhibit different electrostatic properties. To predict the residues contributing to electrostatic pattern distribution, we employed charged-to-alanine scanning and

conducted a series of structure-based mutagenesis *in silico*. Deng and colleagues [25] have proposed that Loop#1 and Loop#2 of *Tb*REL1 N-terminal domain may interact with RNA substrate and proteins respectively. Our computational analysis confirmed Deng and coworker's prediction. We also found that that conserved Loop#3 and Loop#4 play a role in RNA substrate recognition. We anticipate Loop#1, Loop#3, and Loop#4 as *RNA recognition motifs* of KRELs. The highly positive patch on the surface of *Tb*REL2 N-terminal model suggests higher affinity for RNA recognition compared to *Tb*REL1 structure, supporting the experimental results that were previously reported in the literature [22]. Site-directed mutagenesis assay can validate the effect of mutating the surface exposed, charged residues of KREL1 and KREL2 that we have reported in this study (Table 3) on RNA or protein binding affinity. Such experimental validation will give a significant insight into the physical-chemical differences between KREL1 and KREL2 proteins and will help to understand the function of KREL1 and KREL2 in greater details.

Acknowledgements We thank the members of the Salavati laboratory for their discussion and critical reading of the manuscript. We also thank Prof. Enrico O. Purisima (National Research Council of Canada) for valuable comments. Research at the Institute of Parasitology is supported by the Centre for Host-Parasite Interactions and Le Fonds québécois de la recherche sur la nature et les technologies (FQRNT), Quebec. This research is supported by the National Sciences and Engineering Research Council of Canada (NSERC) grant #328186 and Leaders Opportunity Fund (LOF) grant #12573 to R.S.

References

1. Schnaufer A, Domingo GJ, Stuart KD (2002) Natural and induced dyskinetoplastid trypanosomatids: how to live without mitochondrial DNA. Int J Parasitol 32(9):1071–1084
2. Matthews KR (2005) The developmental cell biology of Trypanosoma brucei. J Cell Sci 118(2):283–290
3. Benne R (1993) RNA editing: the alteration of protein coding sequences of RNA. Horwood, New York
4. Horvàth A, Berry EA, Maslov DA (2000) Translation of the edited mRNA for Cytochrome b in Trypanosome Mitochondria. Science 287(5458):1639–1640
5. Horváth A, Kingan TG, Maslov DA (2000) Detection of the mitochondrially encoded Cytochrome c oxidase subunit I in the Trypanosomatid Protozoan *Leishmania tarentolae*. Evidence for translation of unedited mRNA in the kinetoplast. J Biol Chem 275 (22):17160–17165
6. Blum B, Bakalara N, Simpson L (1990) A model for RNA editing in kinetoplastid mitochondria: RNA molecules transcribed from maxicircle DNA provide the edited information. Cell 60:189–198
7. Simpson L, Aphazizhev R, Gao G, Kang X (2004) Mitochondrial proteins and complexes in *Leishmania* and *Trypanosoma* involved in U-insertion/deletion RNA editing. RNA 10(2):159–170
8. Lukeš J, Hashimi H, Zíková A (2005) Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. Curr Genet 48(5):277–299
9. Stuart K, Schnaufer K, Ernst NL, Panigrahi AK (2005) Complex management: RNA editing in trypanosomes. Trends Biochem Sci 30(2):97–105
10. Aphasizhev R (2007) RNA editing. Mol Biol 41(2):227–239
11. Aphasizhev R, Aphasizheva I, Nelson RE, Simpson L (2003) A 100-kD complex of two RNA-binding proteins from mitochondria of *Leishmania tarentolae* catalyzes RNA annealing and interacts with several RNA editing components. RNA 9 (1):62–76
12. Vondrušková E, van den Burg J, Zíková A, Ernst NL, Stuart K, Benne R, Lukeš J (2005) RNA interference analyses suggest a transcript-specific regulatory role for mitochondrial RNA-binding Proteins MRP1 and MRP2 in RNA editing and other RNA processing in *Trypanosoma brucei*. J Biol Chem 280(4):2429–2438
13. Schumacher M, Karamooz E, Zíková A, Trantírek L, Lukeš J (2006) Crystal structures of *T. brucei* MRP1/MRP2 guide-RNA binding complex reveal RNA matchmaking mechanism. Cell 126 (4):701–711
14. Panigrahi AK, Gygi SP, Ernst NL, Igo RP Jr, Palazzo SS, Schnaufer A, Weston DS, Carmean N, Salavati R, Aebersold R, Stuart KD (2001) Association of two novel proteins, TbMP52 and TbMP48, with the *Trypanosoma brucei* RNA editing complex. Mol Cell Biol 21(2):380–389
15. Rusché LN, Huang CE, Piller KJ, Hemann M, Wirtz E, Sollner-Webb B (2001) The two RNA ligases of the *Trypanosoma brucei* RNA editing complex: cloning the essential band IV gene and identifying the band V gene. Mol Cell Biol 21(4):979–989
16. McManus MT, Shimamura M, Grams J, Hajduk SL (2001) Identification of candidate mitochondrial RNA editing ligases from *Trypanosoma brucei*. RNA 7(2):167–175
17. Cruz-Reyes J, Zhelonkina AG, Huang CE, Sollner-Webb B (2002) Distinct functions of two RNA ligases in active Trypanosoma brucei RNA editing complexes. Mol Cell Biol 22(13):4652–4660
18. Worthey EA, Schnaufer A, Mian IS, Stuart K, Salavati R (2003) Comparative analysis of editosome proteins in trypanosomatids. Nucl Acids Res 31(22):6392–6408
19. Palazzo SS, Panigrahi AK, Igo RP Jr, Salavati R, Stuart K (2003) Kinetoplastid RNA editing ligases: complex association, characterization, and substrate requirements. Mol Biochem Parasito 127 (2):161–167
20. Gao G, Simpson AM, Kang X, Rogers K, Nebohacova M, Li F, Simpson L (2005) Functional complementation of *Trypanosoma brucei* RNA in vitro editing with recombinant RNA ligase. Proc Natl Acad Sci U S A 102(13):4712–4717
21. Schnaufer A, Panigrahi AK, Panicucci B, Igo RP Jr, Salavati R, Stuart K (2001) An RNA ligase essential for RNA editing and survival of the bloodstream form of *Trypanosoma brucei*. Science 291(5511):2159–2162
22. Huang CE, Cruz-Reyes J, Zhelonkina AG, O'Hearn S, Wirtz E, Sollner-Webb B (2001) Roles for ligases in the RNA editing complex of *Trypanosoma brucei*: band IV is needed for U-deletion and RNA repair. The EMBO J 20(17):4694–4703
23. Drozdz M, Palazzo SS, Salavati R, O'Rear J, Clayton C, Stuart K (2002) *Tb*MP81 is required for RNA editing in *Trypanosoma brucei*. EMBO J 21(7):1791–1799
24. Gao G, Simpson L (2003) Is the Trypanosoma brucei REL1 RNA ligase specific for U-deletion RNA editing, and is the REL2 RNA ligase specific for U-insertion editing? J Biol Chem 278 (30):27570–27574
25. Deng J, Schnaufer A, Salavati R, Stuart KD, Hol GJ (2004) High resolution crystal structure of a key editosome enzyme from *Trypanosoma brucei*: RNA editing ligase 1. J Mol Biol 343 (3):601–613
26. Schnaufer A, Ernst NL, Palazzo SS, O'Rear J, Salavati R, Stuart K (2003) Separate insertion and deletion subcomplexes of the

Trypanosoma brucei RNA editing complex. Mol Cell 12(2):307–319
27. Murzin AG (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. EMBO J 12(3):861–867
28. Doherty AJ, Suh SW (2000) Structural and mechanistic conservation in DNA ligases. Nucleic Acids Res 28(21):4051–4058
29. Amaro RE, Swift RV, McCammon JA (2007) Functional and structural insights revealed by molecular dynamics simulations of an essential RNA editing ligase in Trypanosoma brucei. PLoS Negl Trop Dis 1(2):e68
30. Kang X, Gao G, Rogers K, Falick AM, Zhou S, Simpson L (2006) Reconstitution of full-round uridine-deletion RNA editing with three recombinant proteins. Proc National Academy of Sciences 103(38):13944–13949
31. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28(1):235–242
32. Henderson LJ (1908) Concerning the relationship between the strength of acids and their capacity to preserve neutrality. Am J Physiol 21(4):173–179
33. Hasselbalch KA (1917) Die Berechnung der Wasserstoffzahl des Blutes aus der freien und gebunden Kohlensäure desselben, und die Sauerstoffbindung des Blutes als Funktion der Wasserstoffzahl. Biochemische Zeitschrift 78:112–144
34. Olmsted JB (1986) Microtubule-associated proteins. Annu Rev Cell Biol 2(1):421–457
35. Bloom GS, Endow SA (1994) Motor proteins. 1: kinesins. Protein Profile 1(10):1059–1116
36. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389–3402
37. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A 89(22):10915–10919
38. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5(1):113
39. Michener CD, Sokal RR (1957) A quantitative approach to a problem in classification. Evolution 11(2):130–162
40. Sneath PHA, Sokal RB (1973) Numerical taxonomy: the principles and practice of numerical classification. W.H. Freeman, San Francisco
41. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 302(1):205–217
42. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30(14):3059–3066
43. Šali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234(3):779–815
44. Shen MY, Šali A (2006) Statistical potential for assessment and prediction of protein structures. Protein Sci 15(11):2507–2524
45. Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356(6364):83–85
46. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22(12):2577–2637
47. Klapper I, Hagstrom R, Fine R, Sharp K, Honig B (1986) Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: effects of ionic strength and amino-acid modification. Proteins: Structure, Function, Genetics 1(1):47–59
48. Gilson MK, Honig B (1988) Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. Proteins: Structure, Function, Genetics 4(1):7–18
49. Nicholls A, Honig B (1991) A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. J Comp Chem 12(4):435–445
50. Rocchia W, Alexov E, Honig B (2001) Extending the applicability of the nonlinear Poisson-Boltzmann equation: multiple dielectric constants and multivalent ions. J of Phys l Chem B 105(28):6754–6754
51. Halgren TA (1996) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J Comp Chem 17(5 & 6):490–519
52. Halgren TA (1996) Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. J Comp Chem 17(5 & 6):520–552
53. Halgren TA (1996) Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. J Comp Chem 17(5 & 6):553–586
54. Halgren TA, Nachbar RB (1996) Merck molecular force field. IV. Conformational energies and geometries for MMFF94. J Comp Chem 17(5 & 6):587–615
55. Halgren TA (1996) Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. J Comp Chem 17(5 & 6):616–641
56. Halgren TA (1999) MMFF VI. MMFF94s option for energy minimization studies. J Comp Chem 20(7):720–729
57. Halgren TA (1999) MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. J Comp Chem 20(7):730–748
58. Blomberg N, Gabdoulline RR, Nilges M, Wade RC (1999) Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. Proteins: Str, Function Genetics 37:379–387
59. Wade RC, Gabdoulline RR, DeRienzo F (2001) Protein interaction property similarity analysis. Int J Quant Chem 83:122–127
60. Wade RC, Gabdoulline RR, DeRienzo F (2007) qPIPSA: relating enzymatic kinetic parameters and interaction fields. BMC Bioinformatics 8:373
61. Richter S, Wenzel A, Stein M, Gabdoulline RR, Wade RC (2008) webPIPSA: a web server for the comparison of protein interaction properties. Nucleic Acid Res,Web server early access, 36: W276ÐW280
62. Hodgkin EE, Richards WG (1987) Molecular similarity based on electrostatic potential and electric field. Int J Quant Chem Quant Biol Symp 14:105–110
63. Kortemme T, Kim DE, Baker D (2004) Computational alanine scanning of protein-protein interfaces. Sci STKE 2004(219):pl2