

Trypanosome RNA Editing Mediator Complex proteins have distinct functions in gRNA utilization

Rachel M. Simpson¹, Andrew E. Bruno², Runpu Chen³, Kaylen Lott¹, Brianna L. Tylec¹, Jonathan E. Bard⁴, Yijun Sun^{1,2}, Michael J. Buck⁵ and Laurie K. Read^{1,*}

¹Department of Microbiology and Immunology, University at Buffalo Jacobs School of Medicine and Biomedical Sciences, 3435 Main Street, Buffalo, NY 14214, USA, ²Center for Computational Research, University at Buffalo, 701 Ellicott St., Buffalo, NY 14203, USA, ³Department of Computer Science and Engineering, New York State Center of Excellence in Bioinformatics and Life Sciences, 701 Ellicott St., Buffalo, NY 14203, USA, ⁴Genomics and Bioinformatics Core, University at Buffalo, 701 Ellicott St., Buffalo, NY 14203, USA and ⁵Department of Biochemistry, University at Buffalo Jacobs School of Medicine and Biomedical Sciences, 701 Ellicott St., Buffalo, NY 14203, USA

Received February 08, 2017; Revised May 01, 2017; Editorial Decision May 08, 2017; Accepted May 10, 2017

ABSTRACT

Uridine insertion/deletion RNA editing is an essential process in kinetoplastid parasites whereby mitochondrial mRNAs are modified through the specific insertion and deletion of uridines to generate functional open reading frames, many of which encode components of the mitochondrial respiratory chain. The roles of numerous non-enzymatic editing factors have remained opaque given the limitations of conventional methods to interrogate the order and mechanism by which editing progresses and thus roles of individual proteins. Here, we examined whole populations of partially edited sequences using high throughput sequencing and a novel bioinformatic platform, the Trypanosome RNA Editing Alignment Tool (TREAT), to elucidate the roles of three proteins in the RNA Editing Mediator Complex (REMC). We determined that the factors examined function in the progression of editing through a gRNA; however, they have distinct roles and REMC is likely heterogeneous in composition. We provide the first evidence that editing can proceed through numerous paths within a single gRNA and that non-linear modifications are essential, generating commonly observed junction regions. Our data support a model in which RNA editing is executed via multiple paths that necessitate successive re-modification of junction regions facilitated, in part, by the REMC variant containing TbRGG2 and MRB8180.

INTRODUCTION

Uridine (U) insertion/deletion RNA editing is an essential process that occurs in the mitochondria of eukaryotes of the Class Kinetoplastea, whose name derives from their compact mitochondrial genome, or kinetoplast (1–3). In *Trypanosoma brucei*, a parasitic kinetoplastid that causes Human African Trypanosomiasis, the kinetoplast comprises a concatenated network of approximately fifty ~22 kb maxicircles and several thousand ~1 kb minicircles (4). The maxicircles contain 18 protein-coding genes, 12 of which require modification by U insertion/deletion RNA editing to generate their open reading frames (ORFs), and these are thus referred to as cryptogenes (2–5). The minicircles encode small non-coding RNAs, including the 50–70 bp guide RNAs (gRNAs) that serve as trans-acting templates to direct RNA editing (4,6–8). mRNAs encoded by nine of the cryptogenes are edited throughout their lengths and require the utilization of multiple gRNAs, and these are termed ‘pan-edited’ (reviewed in (9) and (5)). Three others are ‘minimally edited’, requiring only one or two gRNAs to modify a small region (reviewed in (9) and (5)). Editing generates translatable ORFs for proteins involved in mitochondrial bioenergetics and a ribosomal protein, thus accounting for the essential nature of the process.

Editing initiates with the first gRNA anchoring to the 3′ end of the mRNA, forming a partial duplex with a 3′ never-edited region that does not require modification (10). This duplexed anchor region forms through canonical Watson–Crick base pairing (10). The remainder of the gRNA contains two components: the guiding region and the poly-U tail. The guiding region initially contains a double hairpin loop structure, as it is not complementary to the pre-edited mRNA (11–15). Following completion of editing, the gRNA is complementary to the fully edited mRNA through a combination of Watson–Crick and G–

*To whom correspondence should be addressed. Tel: +1 716 829 3307; Fax: +1 716 829 3534; Email: lread@buffalo.edu

U basepairing (10,16–18). The poly-U tail is hypothesized to help stabilize the mRNA/gRNA interaction (12,19–21). Although it is unknown the order in which editing modifies the region guided by a single gRNA, the subsequent gRNA requires previously edited mRNA sequence to form an anchor duplex, thus ensuring the overall 3' to 5' progression of editing (10,17,18).

In the steady state mitochondrial RNA population, the majority of transcripts are partially edited, with only a very small number of fully edited mRNAs detectable (17,22–24). Over 90% of these partially edited mRNAs contain regions of non-canonical editing at the 5' leading edge of the fully edited region, termed junctions. Junctions contain edited sequence that does not match the canonical, fully edited sequence, and they are hypothesized to represent regions of active editing that are ultimately re-edited to the correct sequence (17). Junctions likely arise through mis-pairing between cognate gRNA/mRNA pairs, or potentially through utilization of a non-cognate gRNA. Some junctions may also represent alternative edited sequences that could diversify the ORFs generated from a single DNA sequence (6,17,23,25–27). Because even sites that are not modified in the fully edited mRNA sequence can have U's added or deleted in a junction, any location between two non-U nucleotides is designated as an Editing Site (ES; Table 1) (17,23).

Successful execution of U insertion/deletion RNA editing requires both enzymatic and non-enzymatic factors (reviewed in (2) and (3)). The enzymatic factors are contained within the RNA Editing Core Complex (RECC), also termed the 20S editosome (2,28–39). Three distinct RECC variants catalyze general U insertions, general U deletions, or U insertions for a single minimally edited transcript, COII (38,39). However, very little is known about the order in which the RECCs are recruited and whether they act processively or distributively through a region guided by a given gRNA (reviewed in (40)). Additionally, over thirty non-enzymatic factors have been implicated in the RNA editing process (reviewed in (2) and (3)). The organization and function of these proteins is not well understood and likely involves the dynamic coordination of various and likely heterogeneous protein complexes. The current model holds that the RNA editing holoenzyme comprises RECC together with the RNA Editing Substrate Binding Complex (RESC; a.k.a. Mitochondrial RNA Binding Complex 1 or MRB1) (2,3). Within RESC are two subcomplexes, the Guide RNA Binding Complex (GRBC; a.k.a. MRB1 Core) and the less well defined RNA Editing Mediator Complex (REMC; a.k.a. TbrGG2 subcomplex) (reviewed in (2) and (3)). The GRBC contains the only proteins required to stabilize gRNAs, the GAP1/2 heterotetramer, and several proteins thought to be involved in initiation of editing (41–45). GAP1/2 proteins are also present in complexes outside of GRBC, including some that likely impact holoenzyme assembly ((46–49); N. McAdams and L. Read, unpublished). Although the composition of REMC is incompletely defined, the presence of the TbrGG2 protein has been the defining element of REMC, and MRB8170 (a.k.a. REMC5), MRB4160 (a.k.a. REMC5a), and MRB8180 (a.k.a. REMC4) were placed in REMC primarily due to

their direct, RNA-independent interactions with TbrGG2 (43,50–52, reviewed in (2,3)).

In this study, we aimed to better define the function and organization of the REMC complex in U insertion/deletion RNA editing. We focus here on TbrGG2 and three of its binding partners, MRB8170, MRB4160 and MRB8180. We previously showed that TbrGG2 plays a role in the 3' to 5' progression of editing in pan-edited RNAs (22). However, our previous studies using conventional sequencing were insufficiently robust to reveal whether this effect was manifest at the level of progression through a single gRNA-defined region or at the level of gRNA exchange. TbrGG2 binds both MRB8170 and MRB4160 in yeast two-hybrid assays (46). MRB8170 and MRB4160 are paralogues with 77% identity and partially redundant function, and thus they are often notated MRB8170/4160 in RNAi studies where cells are routinely depleted of both proteins to study their function as we do in this study (50,52,53). RNAi demonstrated that MRB8170/4160 play a role in the editing of both pan-edited and minimally edited transcripts and may affect the stability of some mitochondrial mRNAs (50,52,53). MRB8180 also interacts with TbrGG2 via yeast two-hybrid assay (46). It is abundant in immunoprecipitates of MRB8170 and MRB4160 (50,52), and it was reported as a member of the REMC with some transcript specific effects on editing (46,50,52). TbrGG2, MRB8170, and MRB4160 are all RNA binding proteins, and TbrGG2 also exhibits both RNA melting and annealing properties (22,51–54). To refine our understanding of the REMC subcomplex and its role in mediating the progression of editing, we explored the function and interdependence of these proteins using a combination of biochemical techniques and high throughput analysis of partially edited sequences. The RNA editing process requires the precise coordination of protein-protein, protein-RNA and RNA-RNA interactions and is thought to entail the dynamic remodeling of the mRNA/gRNA duplex. Given the effect of TbrGG2 on the progression of editing and the capacity of TbrGG2, MRB8170 and MRB4160 to bind and, in the case of TbrGG2, remodel RNA structure, we hypothesized that these proteins act as remodeling factors to facilitate the progression of editing. This study provides strong evidence that TbrGG2 and MRB8180 are required for resolution of editing within a gRNA-defined region, and indicates that the REMC proteins analyzed here have distinct functions. Our data point to a REMC that has greater functional, and potentially physical, heterogeneity than previously anticipated.

MATERIALS AND METHODS

RNAi cell line construction and culture

The RNAi cell lines in this paper all derive from procyclic form 29-13 *T. brucei brucei* cells. The TbrGG2, MRB8170/4160 and GAP1 RNAi cell lines were previously described (41,51,52). The MRB8180 RNAi construct was made by amplifying a 520 nucleotide region of the MRB8180 ORF with primers GAAAGCTTCTACGACCAACGAAAACGGT, forward primer with HindIII site underlined, and GAGGATCCCTTTACAAGGTTGCGGGGTA, reverse primer with BamHI site underlined. The resulting

Table 1. Glossary of abbreviations

Abbrev.	Full term	Definition
ES	Editing Site	Any space between two non-T nucleotides (cDNA) has the potential to be edited at the RNA level and is termed an Editing Site. These are numbered from 3' to 5' following the direction of editing.
ESS	Editing Stop Site	Moving 3' to 5', the Editing Stop Site (ESS) is the final (5' most) ES that matches the canonical fully edited sequence correctly. All ESs 3' of the ESS match the canonical fully edited sequence.
EPS	Exacerbated Pause Site	An Exacerbated Pause Site is defined as an Editing Stop Site at which the total number of sequences sharing this Editing Stop Site in a sample has increased significantly compared to the uninduced control samples ($P < 0.05$, $q < 0.05$) and thus is considered a site where editing stalls more in the sample.
JSS	Junction Start Site	The first ES moving 3' to 5' that does not match the canonical fully edited sequence correctly (can match pre-edited or mis-edited)
JES	Junction End Site	The 5' most editing site with any editing action, whether canonical or mis-edited.
JL	Junction Length	The number of editing sites contained within a junction including both the JSS and JES (e.g. a junction arising after ESS15 with a JES at ES20 would have a junction length of 5).

amplicon fragment was digested with BamHI and HindIII, and an internal HindIII cut site resulting in a 289-bp fragment that was cloned into the p2T7-177 vector containing phleomycin resistance (55). The construct was transfected and cell lines were cultured in selective media containing phleomycin (2.5 $\mu\text{g/ml}$), hygromycin (50 $\mu\text{g/ml}$) and G418 (15 $\mu\text{g/ml}$), as previously described (43). Growth curves were performed on pseudoclonal cells uninduced and induced with 2.5 $\mu\text{g/ml}$ of tetracycline in triplicate (43).

Quantitative real time PCR analysis

MRB8180 RNAi cells were grown either uninduced or induced with 2.5 $\mu\text{g/ml}$ tetracycline for three days. Cells were harvested and RNA extracted and DNase treated as previously described (43). DNase treated RNA was tested for purity using a NanoDrop 1000 (Thermo Scientific). 1 μg of DNased RNA was run on a 1.2% TBE gel to confirm intactness of ribosomal RNA. cDNA was generated using random hexamer primers in the Taq-man Reverse Transcriptase Kit and a no reverse transcriptase sample was included for each replicate as a control for DNA contamination (Applied Biosystems, Thermo Fisher). Quantitative real time RT-PCR to detect levels of MRB8180 mRNA was performed using the following primers (forward primer: CTTTATCGTCGGACTGTAGG, reverse primer: ATCGCTGTTATCCGTAGAGA) which amplify a region of 93nt in Tb927.8.8180 and its nearly identical paralogue Tb927.4.4150, which is also targeted by the RNAi. Specificity of primers were confirmed by examination of the melt curve for a single peak with separation of product on a 1.2% TAE gel to confirm size and single product. Primers used to detect other target mRNAs in RNAi lines were as described previously for TbRGG2 (51), MRB8170 (52), MRB4160 (52) and GAP1 (forward primer: AACGTATTGCGGATGCTTAC, reverse primer: GAAC-CACACGCTCACAAACAG). Detection of pre-edited and fully edited versions of edited RNAs, never edited RNAs, and precursor RNAs, primers and qRT-PCR was performed as described, normalizing to 18S rRNA (43). Analysis of qRT-PCR results was done using BioRad iQ5 software. All results shown reflect two biological replicates, each with three technical replicates of the qRT-PCR reaction. Shown is the average fold change of each transcript with the

standard deviation in the error bars (43). Further details are provided in the MIQE checklist in Supplementary Table S1.

Generation of recombinant MBP-MRB8180 and anti-MRB8180 antibody

The MRB8180 open reading frame was amplified using forward primer AAGGATTTTCAGAATTCATGTACCGTC TTTATCGTCG and reverse primer CGACTCTAGAGG ATCCCTAAACTGCTGTGGCCAG, and cloned into the pMalC2 vector (NEB) using the Infusion cloning kit (Clontech). The resulting construct was expressed in BL21 *Escherichia coli* cells (New England Biolabs). Cells were grown in liquid LB with selective ampicillin (0.1 mg/ml) at 37°C to a density of OD₆₀₀ 0.4, at which point the temperature was reduced to 18°C and at OD₆₀₀ 0.6, cells were induced with 0.3 mM IPTG and grown overnight. Cells were harvested by centrifugation, resuspended in lysis buffer (10 mM Tris pH 8.0, 200 mM NaCl, 3 mM BME, 0.1 mM PMSF, 100 μl leupeptin, 100 μl pepstatin), and lysed by sonication (60% power, six pulses for 30 s each with 30 s breaks, Fisher Scientific Sonic Dismembrator model 500). Lysate was cleared by centrifugation, the NaCl concentration of the cleared lysate was brought up to 600 mM, and polyethyleneimine was added to a final 0.1% to remove nucleic acids. Polyethyleneimine treated lysate was cleared by centrifugation and incubated with amylose beads (New England Biolabs), rocking overnight. Beads were then washed with 850 ml of wash buffer (10 mM Tris pH 8.0, 200 mM NaCl, 3 mM BME, 0.1 mM PMSF), and protein was eluted from the beads with wash buffer containing 30 mM maltose, incubated for 1 h at room temperature. The concentration of the recombinant protein was determined by comparison with a BSA standard curve. Protein was stored in elution buffer at -80°C after snap freezing with liquid nitrogen. This recombinant protein was used to produce antibodies against MRB8180 in rabbits (Bethyl Laboratories), and in the affinity purification of antibody from the rabbit sera. The specificity of this antibody was confirmed by probing an RNAi cell line targeting MRB8180 and levels of MRB8180 mRNA in that sample were confirmed by qRT-PCR.

***In vitro* RNA crosslinking**

In vitro RNA crosslinking was performed as previously described (48,52). Briefly, RNA fragments of A6 mRNA (A6U5, described in (56)) and A6 gRNA (gA6[14]NX, described in (56)) were internally labeled with [α - 32 P]-UTP during *in vitro* transcription using the T7 Megascript Kit (Ambion). Five fmol of RNA was incubated with 1.2 μ g of recombinant protein at room temperature for 20 min, followed by exposure to UV radiation. RNA was digested using RNase A, and proteins separated by SDS-PAGE. Protein was visualized by Coomassie stain. The gel was then dried and label transfer visualized with phosphorimaging. The data shown is one of two technical replicates whose results are consistent.

Western blot analysis

Cells were grown either uninduced or induced with tetracycline (5 μ g/ml for MRB8180 RNAi; 2.5 μ g/ml all others), for three days. Cells were harvested, resuspended in SDS-PAGE sample buffer to 1×10^6 cells/ μ l, and lysed at 95°C for 15 min. Equal cell equivalents were separated by SDS-PAGE. Protein was transferred to a nitrocellulose membrane and probed with primary antibody for TbRGG2 (51), MRB8170 (52), GAP1 (43) and MRB8180 (Bethyl Laboratory, described above), using anti-Hsp70 as a loading control (Hsp70 antibody was a generous gift from Jay Bangs, University at Buffalo). Antibody was visualized using HRP conjugated secondary antibody (EMD Millipore) and imaged using a ChemiDoc MP (BioRad). The intensity of the bands was quantified using ImageLab (BioRad). Each sample was repeated with three biological replicates, and two technical replicates were performed for each biological replicate for each antibody. Representative blots are shown.

Determination of cellular protein abundance

The abundance of cellular proteins was estimated as previously described (57). Briefly, recombinant GST-TbRGG2 (51), GST-MRB8170 (52), and MBP-MRB8180 (above) were quantified using a BSA standard curve. Recombinant proteins were then titrated such that the bands were detectable by western blot, and a known quantity of cells was also detectable with the intensity of its band falling within the standard curve of the recombinant proteins. The number of molecules per cell was calculated using the ng of protein in the recombinant protein curve and known number of cells in the test sample. Determination for each protein was done with two biological replicates.

Preparation of partially edited RNA library

RNA was extracted from RNAi lines either uninduced or induced with tetracycline (5 μ g/ml for MRB8180 RNAi, 2.5 μ g/ml for all others) for 3 days (all samples except TbRGG2 RNAi replicate 2 which was induced for 2 days). RNA extraction and library preparation were performed as described (23). Briefly, cDNA was generated from DNase treated whole cell RNA using gene specific primers (23). These cDNA samples were PCR amplified within the linear

range of the PCR assay to maintain the relative abundance of unique fragments, and the amplicons were sequenced using paired-end Illumina MiSeq and paired (23). The number of fragments obtained (total and unique) in each sample are detailed in Supplementary Table S2. To normalize the number of reads in each sample, the total (de-collapsed) number of fragments that have no non-T mismatches (standard alignments in Table S2) are normalized to 100 000 reads. Thus, each unique sample is scaled such that their relative abundance can be compared via their normalized counts (23). The sequencing data has been deposited in Sequence Read Archive, accession number SRP097727.

Sequence alignment using Trypanosome RNA Editing Alignment Tool (TREAT)

Trypanosome RNA Editing Alignment Tool (TREAT) is a special purpose multiple sequence aligner designed to analyze Uridine insertion/deletion RNA editing. TREAT consists of a command-line alignment tool along with a built-in web server providing a robust web-based interface for searching, viewing and analyzing the alignment results. TREAT is written in Go and freely available under the GPLv3 license at <https://github.com/ubccr/treat>. The initial releases of TREAT up to version 0.0.2 are further described here (23). The new release of TREAT v0.0.3 used in this analysis includes several new features summarized below.

TREAT assumes paired-end sequencing reads have been preprocessed with tools like PEAR for merging paired-end reads and fastx-collapser for collapsing unique sequences. Primer regions are often but not always removed in the pre-processing step. The new release of TREAT adds an offset parameter to the align command. This feature allows the user to adjust the editing site numbering accounting for primer regions that were stripped off during the preprocessing step. The command line tool included in TREAT allows the user to perform searches and export the raw data in various formats. Included in the new release is the ability to export the raw sequences in FASTA format. Notable changes to the web interface in TREAT include the addition of a bubble chart comparing the number of uridine insertions at various editing sites and no longer counting sequences flagged as alternatively edited in histograms on the overview page.

Determination of significant increase of pre-edited transcripts

The number of pre-edited transcripts found in each sample was determined by TREAT analysis (23). The number of pre-edited transcripts in each of the four uninduced samples in a given replicate was compared, and the averages (mean) and standard deviations were calculated. Any induced samples that were increased more than two standard deviations above the average of the uninduced samples in a given replicate were considered significantly increased. RNAi of a given protein was only considered to cause a significant increase in pre-edited transcripts if both replicates showed a significant increase by this method.

Determination of exacerbated pause sites and junction lengths

We developed a hypothesis testing to determine whether the number of sequences with an editing site in induced samples is significantly different from the uninduced number of sequences at the same editing site. The same determination is used for junction lengths. For the i th editing site s_i , let $[n_{i,1}, n_{i,2}, \dots, n_{i,r}]$ and c_i be the observed uninduced numbers and induced number, respectively. By assuming that the uninduced numbers of editing site s_i follow a normal distribution $N(\mu_i, \sigma_i^2)$, we design the following hypothesis test:

$$\begin{aligned} H_0: c_i \text{ is a sample from } N(\mu_i, \sigma_i^2) \\ H_1: c_i \text{ is not a sample from } N(\mu_i, \sigma_i^2) \end{aligned} \quad (1)$$

The mean μ_i and the standard deviation σ_i of the normal distribution can be estimated using the maximum-likelihood estimation:

$$\begin{aligned} \hat{\mu}_i &= \frac{1}{m} \sum_{j=1}^m n_{i,j} \\ \hat{\sigma}_i^2 &= \frac{1}{m-1} \sum_{j=1}^m (n_{i,j} - \hat{\mu}_i)^2 \end{aligned} \quad (2)$$

The P -value of rejecting H_0 for s_i is the probability to obtain the same or more extreme observation under the null model compared to the actual observation c_i computed as

$$p_i = \text{Prob}(|n - \hat{\mu}_i| \geq |c_i - \hat{\mu}_i|) = \begin{cases} 2\Phi\left(\frac{c_i - \hat{\mu}_i}{\hat{\sigma}_i}\right) / \hat{\sigma}_i, & \text{if } c_i \leq \hat{\mu}_i \\ 2 - 2\Phi\left(\frac{c_i - \hat{\mu}_i}{\hat{\sigma}_i}\right) / \hat{\sigma}_i, & \text{if } c_i > \hat{\mu}_i \end{cases} \quad (3)$$

where $n \sim N(\hat{\mu}_i, \hat{\sigma}_i^2)$ and $\Phi(\cdot)$ is the cumulative distribution function of standard Normal distribution $N(0, 1)$. Since the above analysis involves multiple comparisons, it is necessary to control the false discovery rate. To this end, the Benjamini–Hochberg correction (58) is performed to compute the adjusted P -values (i.e. q -values). With a predefined threshold γ , the sites with a q -value smaller than or equal to γ are identified to have a statistically significant difference in the induced number and uninduced numbers, and the selected sites have a false discovery rate not exceeding γ . ESSs were considered EPSs when the $P < 0.05$ and $q < 0.05$ in both biological replicates. Exact values for the normalized counts of sequences sharing each ESS in both replicates and the associated P and q values are shown in Supplementary Tables S3 and S4.

Calculation of correlation of EPSs with editing characteristics

For both RPS12 and ND7-5', editing characteristics were determined for each editing site (5' nucleotide, 3' nucleotide, editing action at the site, and editing action 5' of the site). The number of sites with a given characteristic was determined for EPSs in each sample and for non-EPSs. A Fisher's Exact test was performed in R to determine whether the overlap of each characteristic with a sample's EPS was significant (Supplementary Table S5) as described previously (23). Code showing this analysis can be found in the TREAT github repository.

Examination of junction sequences and editing paths

In each sample, the number of sequences generated at each ESS with junction of length 0, 1–10, 11–50 and > 50 ES long was determined. The average number of sequences was determined for each induced RNAi cell line across both replicates and across all eight uninduced samples, and the values were plotted in R.

The major junction sequences found within gRNA-1 were determined from the database of all sequences found in all samples across both biological replicates by limiting analysis to those sequences whose ESS was found within gRNA-1 (ESS9 to ESS22). The average number of times this sequence was retrieved in each sample was determined by taking the sum of the unique sequence divided by the sample number ($n = 2$ for induced RNAi samples; $n = 8$ for all uninduced samples taken together). The sequences were then probed to determine which had one, two and three modifications beyond correct canonical editing to ES15 using a series of code testing for exact character matching. See Supplementary Tables S6, S7 and S8 for all sequences containing 1, 2 and 3 modifications, respectively.

The number of long junction sequences at EPS in MRB8170/4160 RNAi and TbRGG2 RNAi for RPS12 were determined by limiting these ESSs to sequences with long junctions (greater than 50 ES). These long junctions were then probed to determine whether the majority middle sequence (ES20 to ES48) matched fully edited as seen in potential alternatively edited transcripts or pre-edited sequence. The average number of sequences in both categories was calculated across the two biological replicates and compared graphically between individual uninduced and induced samples.

Code for all R analysis used in this paper can be found through the TREAT github repository.

RESULTS

MRB8180 is an RNA binding protein that is essential for cell growth and pan-editing

Of the proposed REMC components examined in this paper, MRB8180 is the least well characterized. To begin to address MRB8180 function, we depleted procyclic form *T. brucei* cells of this protein using tetracycline-regulated RNAi. We monitored cell growth over 15 days, and demonstrated that MRB8180 is essential for optimal growth (Figure 1A), a result that was not as apparent after a published 8-day growth curve (50). MRB8180 depletion begins to show an effect on growth at day 6, one day later than depletion of MRB8170/4160 and two days later than loss of TbRGG2 (52,54). To determine whether MRB8180 plays a role in RNA editing, we analyzed levels of pre-edited, fully edited, and never edited transcripts, as well as precursor transcripts spanning two genes, for changes in abundance after MRB8180 depletion using qRT-PCR. Depletion of MRB8180 reduced editing of three of four pan-edited mRNAs tested, but did not affect editing of minimally edited, never edited, or precursor mRNAs (Figure 1B). Interestingly, this result is reminiscent of the effect of TbRGG2 depletion, although TbRGG2 also impacts ND7 mRNA editing (54). We observed only a modest increase in pre-

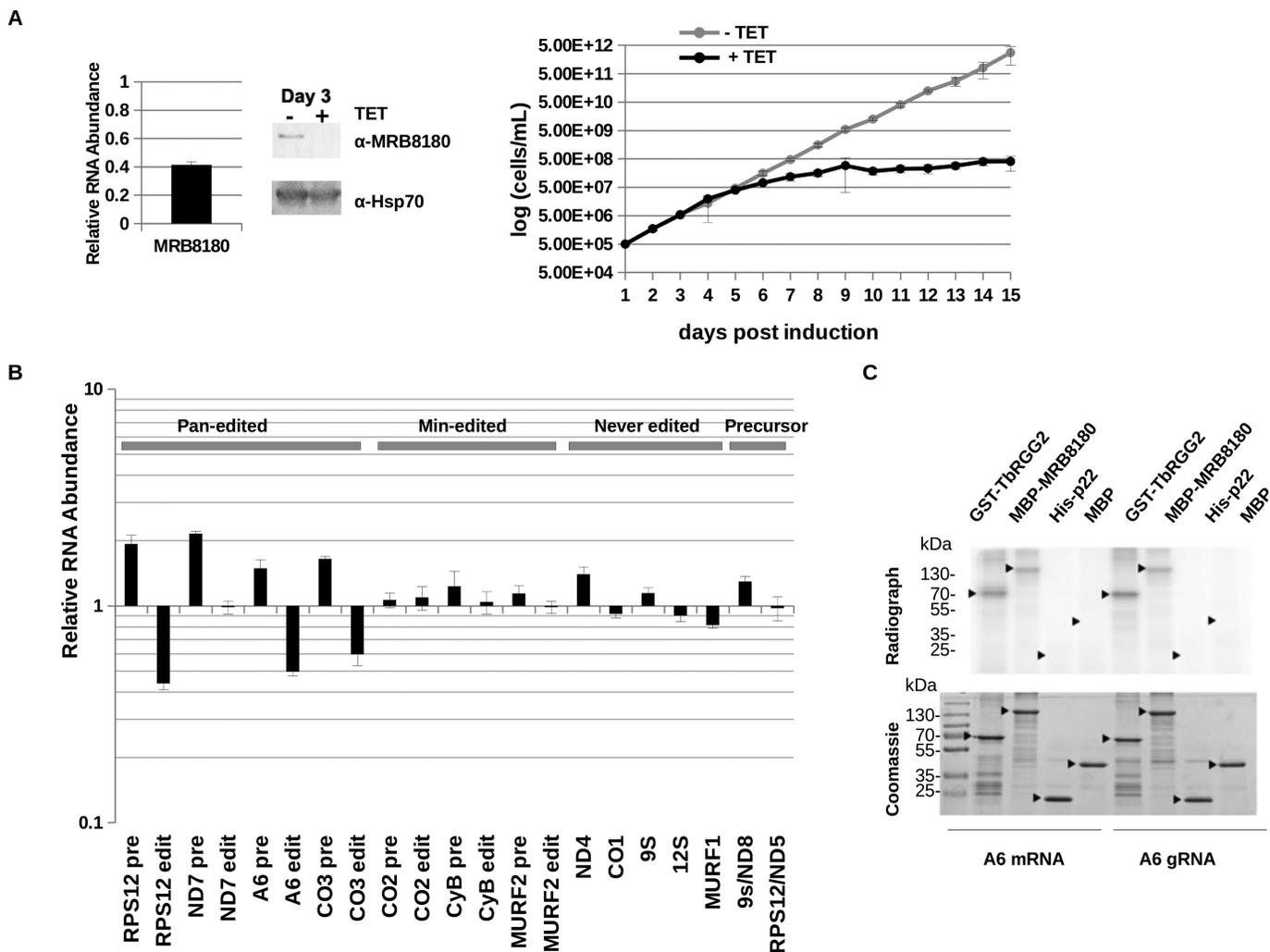


Figure 1. MRB8180 is an essential RNA binding protein. (A) Procytic form *T. brucei* cells containing a stable RNAi construct targeting MRB8180 were grown either uninduced or induced with tetracycline for three days and level of depletion of MRB8180 was determined by qRT-PCR (two biological replicates, $n = 3$ for each) and western blot analysis (left). Relative abundance of MRB8180 RNA in the induced versus uninduced RNAi cells is shown in the y-axis. Growth of MRB8180 RNAi cells either uninduced (grey) or induced (black) was monitored over 16 days in triplicate (right). (B) Levels of maxicircle transcripts in MRB8180 RNAi cells either uninduced or induced as in A for three days were determined using qRT-PCR (two biological replicates, $n = 3$ for each). For transcripts that require editing, both pre-edited (pre) and fully edited (edit) transcripts were measured. Relative abundance of each transcript in induced versus uninduced RNAi cells is shown. (C) *In vitro* UV crosslinking of a body labeled fragment of A6 mRNA and a body labeled A6 gRNA to recombinant MBP-MRB8180 was performed using GST-TbRGG2 as a positive control and his-p22 and MBP as negative controls. The radiograph (top) and Coomassie stained gel (bottom) are both shown; arrowheads indicate positions of recombinant proteins.

edited mRNAs upon MRB8180 depletion, and detected this even in a case in which edited mRNA is not decreased (ND7, Figure 1B). The lack of both substantial pre-edited mRNA accumulation and impact on minimally edited mRNAs, similar to TbRGG2, suggests that MRB8180 plays a role in the progression of editing (22). Given that both TbRGG2 and MRB8170/4160 are RNA binding proteins capable of associating with mRNA and gRNA, we next tested whether MRB8180 is also an RNA binding protein. In UV cross-linking assays with recombinant MRB8180 and body-labeled RNAs, we demonstrated that MRB8180 binds both mRNA and gRNA (Figure 1C). Thus, we conclude that MRB8180 is an essential RNA binding protein that likely facilitates the progression of RNA editing.

REMC exhibits heterogeneity

Multiple proteins that form complexes in *T. brucei*, including the gRNA stabilizing GAP1/2 heterotetramer, display mutual dependence upon one another for stability, such that RNAi of any integral protein results in the complete degradation of the other protein(s) (59–61). Complexes that fit this description frequently act as a single unit within the cell. As TbRGG2, MRB8180, and MRB8170/4160 are binding partners designated as members of REMC (46,50–52), we asked whether they are dependent upon one another for stability (Figure 2 A). If so, it would suggest that REMC is a distinct subparticle within RESC, and that it likely operates to facilitate a single function. However, a lack of stability dependence would suggest that this subcomplex is more transient or heterogeneous than

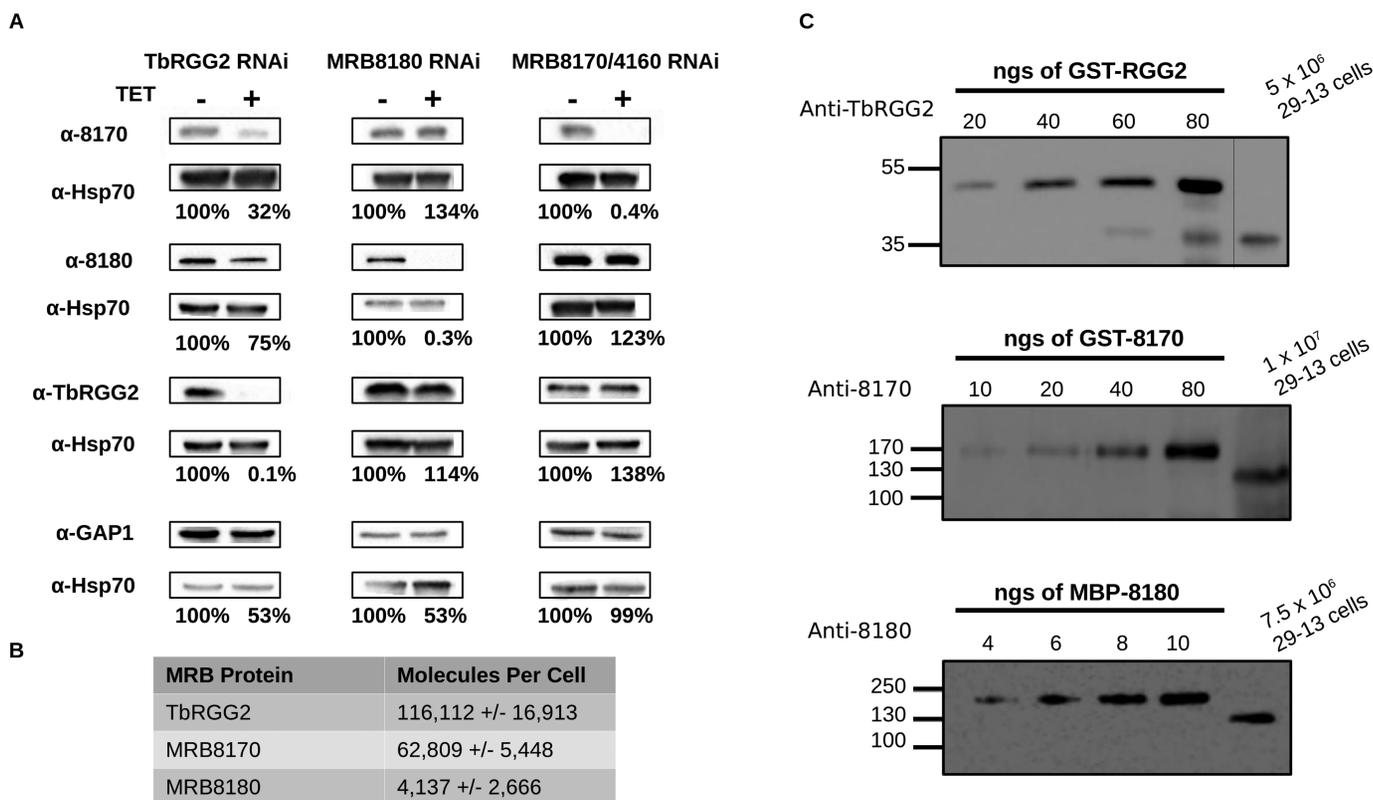


Figure 2. Mutual dependence and abundance of REMC factors. (A) Stable RNAi cell lines targeting TbRGG2, MRB8180, and MRB8170/4160 were induced with tetracycline for 3 days, cells were harvested, and whole cell lysate was probed for each protein using protein specific antibodies, including GAP1 (non-REMC control). The levels of protein in each sample was compared to Hsp70 as a loading control and the quantification of uninduced and induced samples is shown as a percent below each western blot. (B) The number of molecules per cell of each protein was estimated by comparing the intensity of bands generated by protein specific antibodies in a standard curve of known quantities of recombinant protein to a known number of procyclic form, strain 29-13 cells ($n = 2$). An example is shown in (C).

previously described. To address these possibilities, we induced cells containing the established RNAi constructs for depletion of TbRGG2 (51,54), MRB8170/4160 (52) and MRB8180 (Figure 1) for 3 days and analyzed whole cell lysates for the abundance of each REMC factor by western blot, using GAP1 as a control. While loss of TbRGG2 caused a partial destabilization of both MRB8170 and MRB8180, the loss of either MRB8170 or MRB8180 did not affect the stability of TbRGG2 or one another (Figure 2A). GAP1 levels were modestly affected by depletion of either TbRGG2 or MRB8180. To further address whether TbRGG2, MRB8170 or MRB8180 form stoichiometric complexes *in vivo*, we next determined the relative number of molecules per cell of each of the three proteins. To that end, we compared the amount of each protein in a known number of cells to titrations of recombinant proteins (57). Although it is possible that some antibody binding can be masked by the presence of other proteins in the whole cell lysate, the fact that MRB8180 and MRB8170 are essentially identical in size (103 and 100 kDa, respectively), indicates that any potential masking will be comparable for both proteins. We found that TbRGG2 is the most abundant of the three REMC factors, twice as abundant as MRB8170 and 30 times that of MRB8180 (Figure 2B and C). Due to limitations in antibody specificity, we cannot determine the abundance of

the MRB8170 paralog, MRB4160. If MRB4160 is equally abundant as MRB8170 it is possible that TbRGG2 and the combined MRB8170/4160 are of equal abundance and that all TbRGG2 molecules in the cell could be bound to one of the two paralogs. In contrast, as there are 30 times as many TbRGG2 molecules as MRB8180 molecules in the cell, we conclude that MRB8180 can at most be bound to a subset of TbRGG2. The lack of absolute stability dependence and the differing relative abundances of these proteins suggest that REMC does not act as a single particle with a collective function, but instead suggest that REMC is more likely to be a heterogeneous or dynamically forming complex within RESC.

REMC factors affect the progression of editing

Given the evidence for heterogeneity of REMC architecture, we next asked whether TbRGG2, MRB8180 and MRB8170/4160 are functionally distinct. To examine the effects of these proteins on RNA editing in detail, we employed paired-end Illumina MiSeq to sequence the complete population of pre-, fully- and partially-edited RPS12 and ND7-5' transcripts from cells replete with and depleted of TbRGG2, MRB8180, MRB8170/4160 and GAP1 (23). Primers to amplify these transcripts are located in the 3' and 5' never edited regions of RPS12 and 5' edited domain

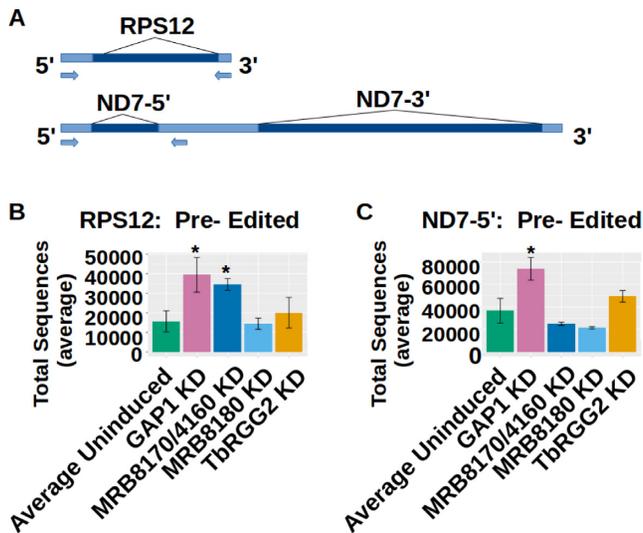


Figure 3. Effect of REMC factor depletion on pre-edited mRNA abundances. (A) Schematics show the locations of primers used to amplify the full populations of pre-edited, partially edited, and fully edited RPS12 and ND7-5' transcripts (not to scale). Note that editing of ND7-5' can initiate prior to completion of editing of the 3' domain. (B and C) The number of pre-edited transcripts was determined in each replicate of both transcripts for induced and uninduced cells. Asterisks indicate a significant increase in both replicates; error bars indicate standard deviations.

of ND7 (Figure 3A). Transcripts were amplified in a linear manner from two biological replicates for each RNAi cell line, and the reads from these data were paired and normalized to 100 000 total de-collapsed reads as previously described, such that all comparisons made in this paper examine the relative abundance of individual sequences (23). For each partially edited sequence, we used our previously described TREAT algorithm to determine the extent of canonical editing and both the length and sequence of the junction regions (Table 1) (23). In these analyses, the 5' most site of contiguous canonical editing is termed the Editing Stop Site (ESS) (Table 1). Just 5' of the majority of ESSs, partially edited RPS12 and ND7-5' mRNAs contain sequences termed junctions that are edited to a sequence matching neither pre-edited nor canonically edited, which are hypothesized to be critical intermediates in the editing process (Table 1) (17,22,23). TREAT defines the junction region as extending from the 3' most editing site that does not correctly match canonical fully edited sequence, termed the junction start site (JSS), to the 5' most editing site beyond the ESS that shows any editing modification, defined as the Junction End Site (JES), enabling us to define the junction length and sequence (Table 1) (23). TREAT also quantifies pre-edited mRNAs, and accumulation of this class of mRNA upon knockdown of a given protein is consistent with a role for that protein in editing initiation (43,44). To confirm qRT-PCR data suggesting that REMC proteins have a minimal effect on editing initiation, we first determined whether depletion of REMC factors caused a significant increase in pre-edited RPS12 or ND7-5' mRNAs using high throughput sequencing. The GAP1 RNAi cell line was used as a control, as GAP1 depletion impairs gRNA stability and editing initiation (41,42). As

expected, GAP1 depletion significantly increased the levels of pre-edited transcripts ($P < 0.05$) for both RPS12 and ND7-5' (Figure 3B,C). In contrast, neither TbrGG2 nor MRB8180 depletion caused a significant increase in the levels of either pre-edited transcript, indicating that these proteins do not play a significant role in editing initiation. Unlike the other REMC factors, MRB8170/4160 depletion resulted in a significant increase in pre-edited RPS12 mRNA, thereby indicating a distinct function for MRB8170/4160, likely at the step of editing initiation. The effect of MRB8170/4160 on pre-edited mRNA abundance was not mirrored in ND7-5' transcripts, however. These results suggest that MRB8170/4160 affects progression of ND7-5' editing, and demonstrates transcript specific functions for this protein. The absence of MRB8170/4160 impact on pre-edited ND7-5' transcripts may reflect the fact that we are measuring editing only in the 5' editing domain of the complete ND7 mRNA, and we cannot rule out that MRB8170/4160 functions in initiation of the 3' domain of ND7 (Figure 3A). Collectively, high throughput sequencing data analyzed using TREAT confirm that the primary roles of TbrGG2 and MRB8180 are in mediating the 3' to 5' progression of editing, rather than in editing initiation. They further suggest that MRB8170/4160 has diverse, transcript specific functions in editing initiation and progression. These findings support a model wherein these REMC proteins are functionally heterogeneous, despite being binding partners.

REMC factors play distinct roles in facilitating editing through a gRNA-defined region

Having confirmed that TbrGG2, MRB8180 and MRB8170/4160 are required for the progression of editing, we next asked whether they affect gRNA exchange or the progression of editing through a gRNA-defined region. To this end, we utilized high throughput sequencing and TREAT analysis to determine the sites at which the 3' to 5' progression of canonical editing becomes stalled upon loss of each of these proteins by quantifying the ESSs that increase significantly upon depletion of TbrGG2, MRB8180, or MRB8170/4160. The significantly increased ESSs ($P < 0.05$, $q < 0.05$, Tables S3 and S4) are termed Exacerbated Pause Sites (EPS; Table 1) and comprise a heterogeneous population of sequences that share the same 5'-most site of contiguous canonical editing, but whose junction sequences can vary. As a control, we first analyzed the EPSs arising through depletion of GAP1, as this protein's role in stabilizing gRNAs implies that its loss should cause a defect in gRNA exchange in those mRNAs that have entered the editing pathway. We observed that, in both ND7-5' and RPS12, the EPSs in GAP1 depleted cells are found almost exclusively at the ends of gRNAs (Figure 4, purple diamonds), indicative of a defect in gRNA exchange. As the gRNA sequences used to define the gRNA blocks in Figure 4 were sequenced in a different procyclic *T. brucei* strain, this control also supports that the gRNA-defined blocks are consistent between these two strains (6). Further, the most exacerbated pause sites in GAP1 depleted cells in both transcripts occur at the ends of the regions guided by the first and second gRNAs, consistent with a decreased

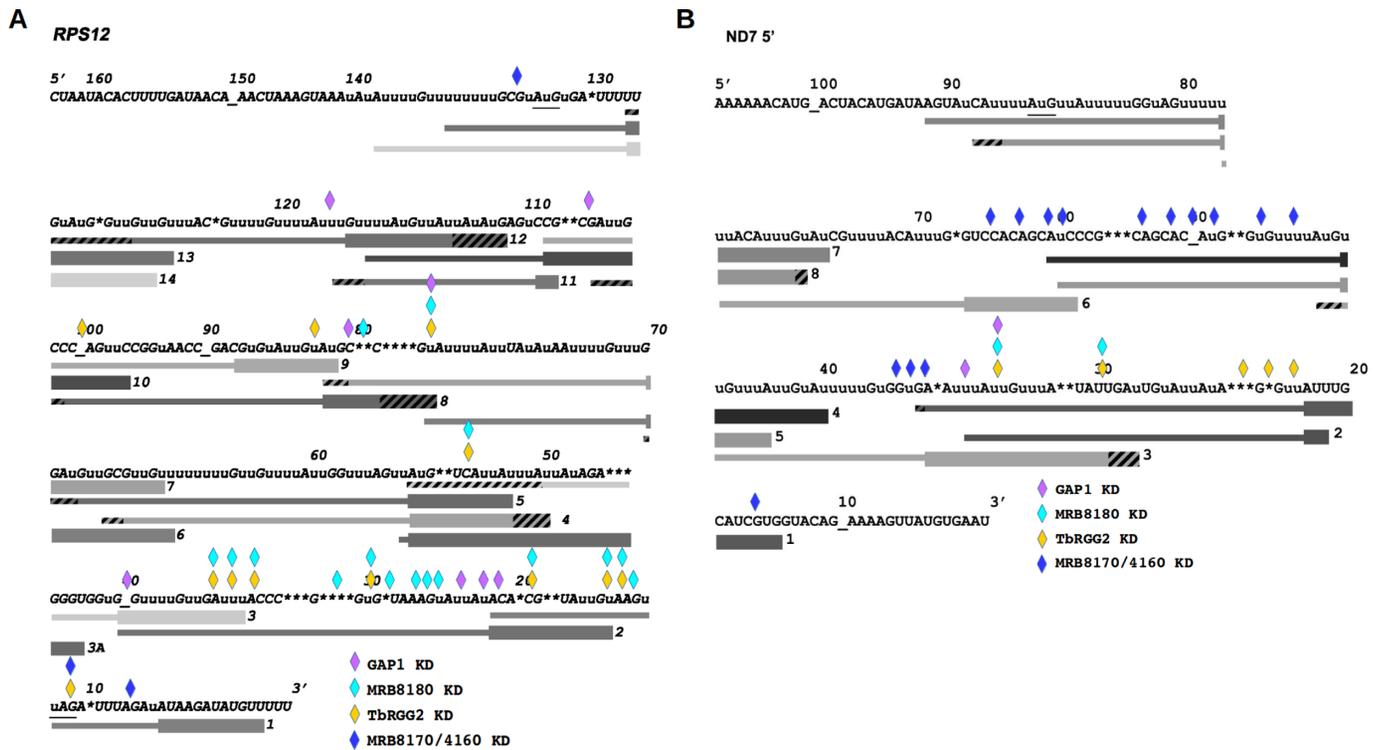


Figure 4. Exacerbated Pause Sites (EPSs) resulting from TbRGG2, MRB8180, MRB8170/4160, and GAP1 knockdown EPSs arising upon depletion of each protein were determined in each of two biological replicates. Diamonds denote the location of EPSs that were present in both replicates for each cell line. Regions guided by specific gRNAs are shown in grey bars below the sequence as reported in Koslowsky *et al.* (6). The gRNAs are numbered at their rightmost side from 3' to 5' in the order of presumed use. The wider segment of the gRNA bar represents the likely anchor region, and the hatched region denotes range of variation of gRNA lengths across members of the same gRNA class. Upper case Us denote genomically encoded uridines, lower case u's denote uridines added in the editing process, and asterisks (*) denote sites where genomic uridines have been deleted in editing. (A) RPS12 mRNA; (B) ND7-5' mRNA.

population of sequences entering the editing process due to impaired initiation. The junction sequences 5' of these EPSs are largely consistent with complete utilization of the 3' gRNA with no editing in the subsequent gRNA (i.e., junction length zero or junction length one if the 5' most site is not correctly guided by 3' gRNA, Supplementary Figure S1). To distinguish roles for REMC factors in gRNA exchange vs. progression through a gRNA-defined region, we next analyzed EPSs in RPS12 and ND7-5' mRNAs in cells depleted of TbRGG2, MRB8180, or MRB8170/4160. These analyses clearly demonstrate that depletion of TbRGG2, MRB8180 or MRB8170/4160 causes pauses in canonical editing to arise in the middle of regions guided by a single gRNA in both transcripts (Figure 4; yellow, cyan, and blue diamonds). The degree of exacerbation at each EPS is shown in Supplementary Figure S2. The occurrence of EPSs throughout the lengths of gRNAs demonstrates that TbRGG2, MRB8180 and MRB8170/4160 are required for the progression of editing through a gRNA-defined region.

To begin to understand the similarities or distinctions between TbRGG2, MRB8180, and MRB8170/4160 functions in editing progression, we first analyzed the overlap in the locations of EPSs arising due to the depletion of each of these three REMC factors. Visual inspection showed that TbRGG2 and MRB8180 RNAi lines display the most apparent similarity (compare yellow and cyan diamonds, Fig-

ure 4). To define how likely it is that any observed overlap is due to chance, we determined the significance of the overlap across each protein pair. These analyses revealed that only the overlap between the EPSs in TbRGG2 and MRB8180 depleted cells is significant (RPS12: $P = 5.65e-8$, ND7-5': $P = 0.004$). This suggests that TbRGG2 and MRB8180 are required for editing at many of the same locations, possibly in complex together, but that the trio of REMC proteins rarely, if ever, collaborates to facilitate editing at a single location. It is important to note that, despite the partial loss of MRB8180 and MRB8170/4160 upon depletion of TbRGG2, the editing defects apparent when TbRGG2 is depleted are distinct from those arising when the other proteins are depleted. This includes sites where only TbRGG2 loss causes increased pausing as well as sites that are only increased when the other proteins are depleted but which are not affected by loss of TbRGG2 (Figure 4 and see below). Thus, the phenotype observed upon depletion of TbRGG2 is not merely due to the secondary degradation of MRB8180 and MRB8170/4160, but instead reflects the functionality of TbRGG2.

Analysis of EPSs in REMC factor knockdown lines also reveals transcript specific differences in the functions of MRB8180 and MRB8170/4160. Consistent with qRT-PCR showing minimal effect of MRB8180 on ND7-5' editing (Figure 1C), we observed very little effect of MRB8180 depletion on the progression of ND7-5' editing (only two

EPSs, Figure 4B), despite the extensive effect of MRB8180 on RPS12 mRNA editing. Conversely, MRB8170/4160 depletion caused exacerbated pausing at numerous sites across ND7-5', but far fewer effects on RPS12 editing (only three EPSs, Figure 4A). Overall, we conclude from these data that TbrGG2, MRB8180 and MRB8170/4160 play distinct, but partially overlapping, roles in mediating the progression of editing through a gRNA-defined region.

TbrGG2 and MRB8180 are important for accurate resolution of uridine deletion sites

Next, we asked whether EPSs in each of the three knock-down lines arise due to a common challenge in the editing process. We tested whether the locations of EPSs correlate with specific 5' or 3' nucleotides, editing requirement at the EPS, and editing requirement immediately 5' at the next site to be edited. The only significant correlation was that EPSs in TbrGG2 and MRB8180 knock-downs are enriched at sites immediately 3' of sites that require deletion action (TbrGG2: RPS12 $P = 0.04$, ND7-5' $P = 0.01$; MRB8180: RPS12 $P = 0.03$; remaining P -values Supplementary Table S5). One possible explanation of this result is that depletion of these factors leads a failure of the deletion RECC to productively associate with the mRNA/gRNA duplex, which would be expected to correlate with skipped deletions within junction regions. To address this possibility, we examined the junction sequences at EPSs in RPS12 that are 3' of sites requiring deletion and determined whether the proportion of junctions that skipped the deletion entirely increased when TbrGG2 or MRB8180 was depleted. We compared these with the proportion of sites having an incomplete deletion or inappropriate addition. We determined that the majority of these sites in both uninduced cells and cells induced for depletion of TbrGG2 or MRB8180 contained approximately equal levels of skipped deletions and partial deletions, with aberrant additions being a rare phenomenon (Supplementary Figure S3), consistent with findings in wild type cells (23). Additionally, we did not observe any large or consistent shift in the frequency with which junctions containing each of these three actions arise after depletion of MRB8180 or TbrGG2 (Supplementary Figure S3), suggesting that these REMC factors do not play a significant role in recruitment of the deletion RECC to the mRNA/gRNA duplex. Rather, because the majority of these junction intermediates would be predicted to form structures with a one or two nucleotide U bulge in the mRNA, even after a partial deletion, our data suggest that TbrGG2 and MRB8180 facilitate the ability of the editing machinery to resolve this incomplete duplex, potentially through modulation of RNA structure (22).

TbrGG2 and MRB8180 promote editing progression through sites that typically lack junctions

In a previous study (22), depletion of TbrGG2 resulted in an accumulation of sequences with no junctions and with short junctions (1–5 editing sites), as well as a concomitant loss of longer junctions (> 10 ES), suggesting that TbrGG2 promotes junction formation during editing. As this study was based on a small sample size ($n < 50$) derived from con-

ventional sequencing, we first asked whether our more comprehensive data confirm this observation. We quantified the number of sequences with each junction length across all possible ESS in both our induced and uninduced samples and determined which junction lengths changed significantly when REMC factors are depleted. We observed that sequences with no junction (junction length zero) increased significantly upon TbrGG2 depletion in both RPS12 and ND7-5' mRNAs (Figure 5A and B). Sequences containing some short junctions were also increased, although different length junctions were increased between the two transcripts (Figure 5A and B). Thus, we confirmed that TbrGG2 depletion causes an accumulation of sequences with no and short junctions. Short junction lengths are differentially affected across transcripts, suggesting that short junctions are not universally regulated by length.

The accumulation of sequences lacking junctions upon TbrGG2 knockdown does not, by itself, answer the question of whether TbrGG2 promotes junction formation. That is, does the increase in sequences with no junctions reflect a global loss of junctions across all or most editing sites, suggesting a fundamental role for TbrGG2 in junction formation? Or, is the increase in junction zero sequences limited to specific ESSs, suggesting a more localized and specific effect of TbrGG2 depletion? To address this question, we first determined whether the number of sequences with no junction increases significantly at each ESS, examining only those sequences with a zero length junction. We determined that only a subset of ESSs in both RPS12 and ND7-5' had a significant increase in sequences with zero length junctions upon the TbrGG2 knockdown (14 of 68 sites in RPS12; 11 of 38 sites in ND7-5'; Figure 5C and Supplementary Figure S4, black circles), indicating a more localized effect. The majority of the sites at which we detect a significantly increased number of junction zero sequences upon TbrGG2 knockdown (Figure 5C and S4, black circle) are also EPSs in TbrGG2 RNAi samples (Figure 5C and S4, white circles). To ensure that these sites truly account for the increase in the population of junction zero sequences seen in Figure 5A and B, we determined whether these significantly increased sites accounted for a greater percentage of all junction zero sequences in the induced compared to uninduced samples. For both RPS12 and ND7-5', the junction zero sequences that were significantly increased in the TbrGG2 knockdown comprised a greater percentage of all junction sequences in each transcript, growing from ~66% (5675/8538) to ~91% (14716/16258) of all junction zero sequences, and thus they are responsible for the significant increase in junction zero sequences shown in Figure 5A and B. Finally, we asked whether the increase in junction zero sequences at these sites leads to a corresponding loss of the junctions that intrinsically form at these sites. We graphed the number of sequences with no junction, and short (1 – 10 ES), medium (11 – 50 ES), and long (> 50 ES) junctions arising at each ESS in both transcripts in the uninduced and induced RNAi samples (Figure 5C and S4). We observed only small variations in the proportions of junctions of different lengths at each ESS upon knockdown of TbrGG2, and noted that multiple EPSs had a high proportion of sequences with no junction in both uninduced and induced samples. This is typified in the region specified by gRNA-

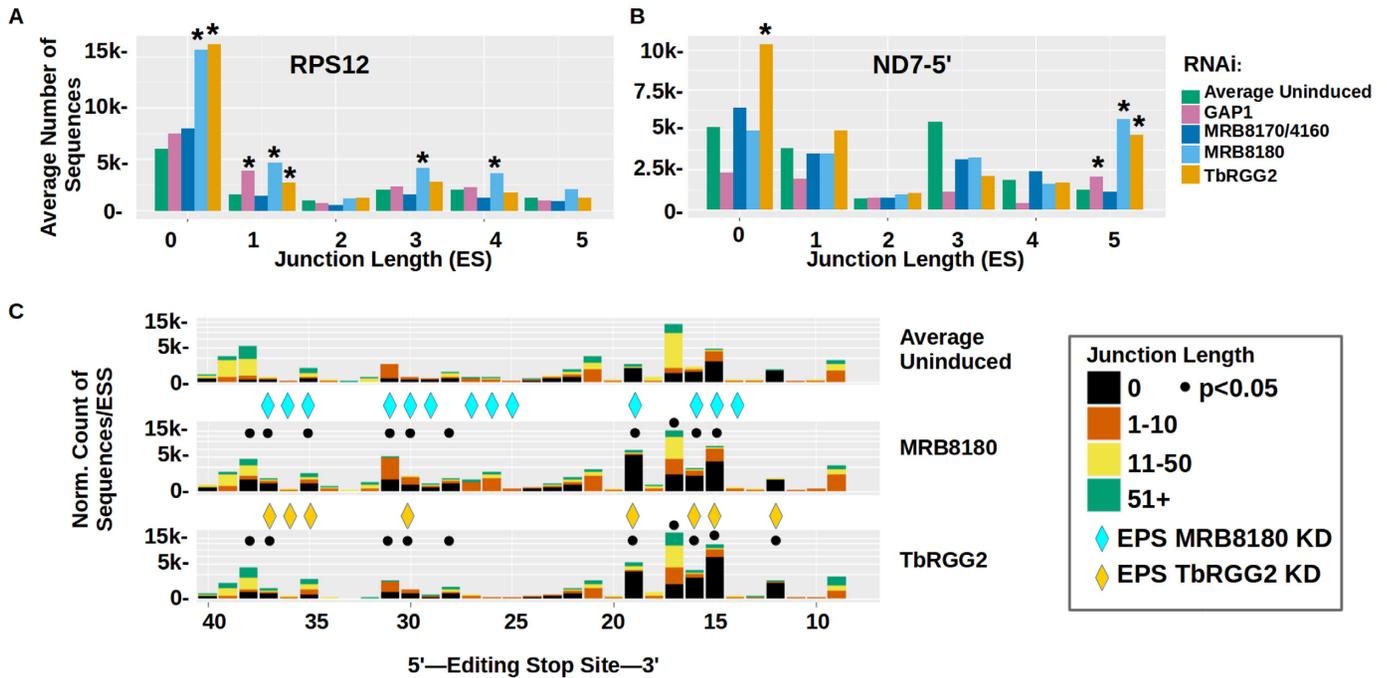


Figure 5. TbRGG2 or MRB8180 depletion causes accumulation of zero and short length junctions at specific locations. The total number of sequences of each junction length was quantified across the whole population of partially edited sequences in each transcript. Junction lengths zero to five from RPS12 (A) and ND7-5' (B) are shown. Each bar is the average number of sequences (normalized counts) across all replicates ($n = 8$ for uninduced, $n = 2$ for each induced). An asterisk denotes junction lengths that were significantly increased in both replicates relative to the uninduced ($P < 0.05$, $q < 0.05$). (C) The average number of sequences with no junction (0), short (1–10 ES), medium (11–50 ES) and long (>50 ES) junctions was determined for sequences arising from each Editing Stop Site (ESS) across RPS12 from ESSs 9–40, which spans the region directed by the first two gRNAs. Filled black circles above the bars indicate those ESSs with a significant ($P < 0.05$, $q < 0.05$) increase in the number of sequences with junction length zero in induced cells compared to the uninduced controls. Colored diamonds, as in Figure 4, are included for reference to denote the location of EPSs for each knockdown.

1 and gRNA-2 of RPS12 shown in Figure 5C. From these data, we conclude that the depletion of TbRGG2 does not cause a global loss of junctions, but rather exacerbates pausing at specific sites that commonly form sequences lacking junctions.

To determine whether REMC components function coordinately with regard to junction formation, we asked whether the depletion of MRB8180 or MRB8170/4160 causes similar editing defects as loss of TbRGG2. As shown in Figure 5A, MRB8180 depletion caused a comparable increase in the number of sequences with junction length zero in RPS12 and, as in TbRGG2 depleted cells, this was localized to specific sites, largely EPSs and sites with intrinsically high levels of junction length zero (Figure 5C and S4). A similar increase in junction zero sequences was not observed in ND7 mRNA, likely due to the overall weak effect of MRB8180 depletion on editing of this mRNA (Figure 5B). MRB8180 depletion also resulted in an increased number of several short junction sequences in RPS12 (Figure 5A). In contrast, MRB8170/4160 depletion did not cause a significant increase in sequences lacking junctions or with junctions between 1 and 5 editing sites in either transcript (Figure 5A and B). The partial overlap of exacerbated junction lengths that we observe upon depletion of MRB8180 and TbRGG2 further supports the model that these proteins are more functionally related than either is to MRB8170/4160. However, the short junctions increasing upon loss of MRB8180, but not TbRGG2, in RPS12 and

the lack of junction zero increasing in ND7-5' for MRB8180 RNAi indicates that these proteins are not identical in function.

Editing within the gRNA-1 guided region can occur by linear or non-linear paths

The data in Figure 5 and S4 reveal a diversity of junction lengths at each ESS, and thus a diversity of partially edited sequences. To more comprehensively define the progression of editing and to compare the impact of TbRGG2 and MRB8180 on this process, we next examined the most abundant sequences in the region of RPS12 mRNA that is guided by gRNA-1. Figure 6 shows these sequences and lists the number of sequences in each cell line (average normalized count), the percent of total gRNA-1 directed sequences made up by each sequence, and the fold increase for a given sequence in each knockdown line compared to the number in the uninduced control. Included in the figure are those sequences contained within the gRNA-1 directed region for which there were at least 100 average normalized counts. In the uninduced samples, ESS15;junction 8 and ESS15;junction 3, are also prominent. The second most abundant sequence in the uninduced samples is ESS19;junction 0, and several other junction length 0 sequences are represented here. The abundance of junction zero sequences at ESSs 15 and 19, together with relatively abundant junction zero sequences at ESSs 16 and

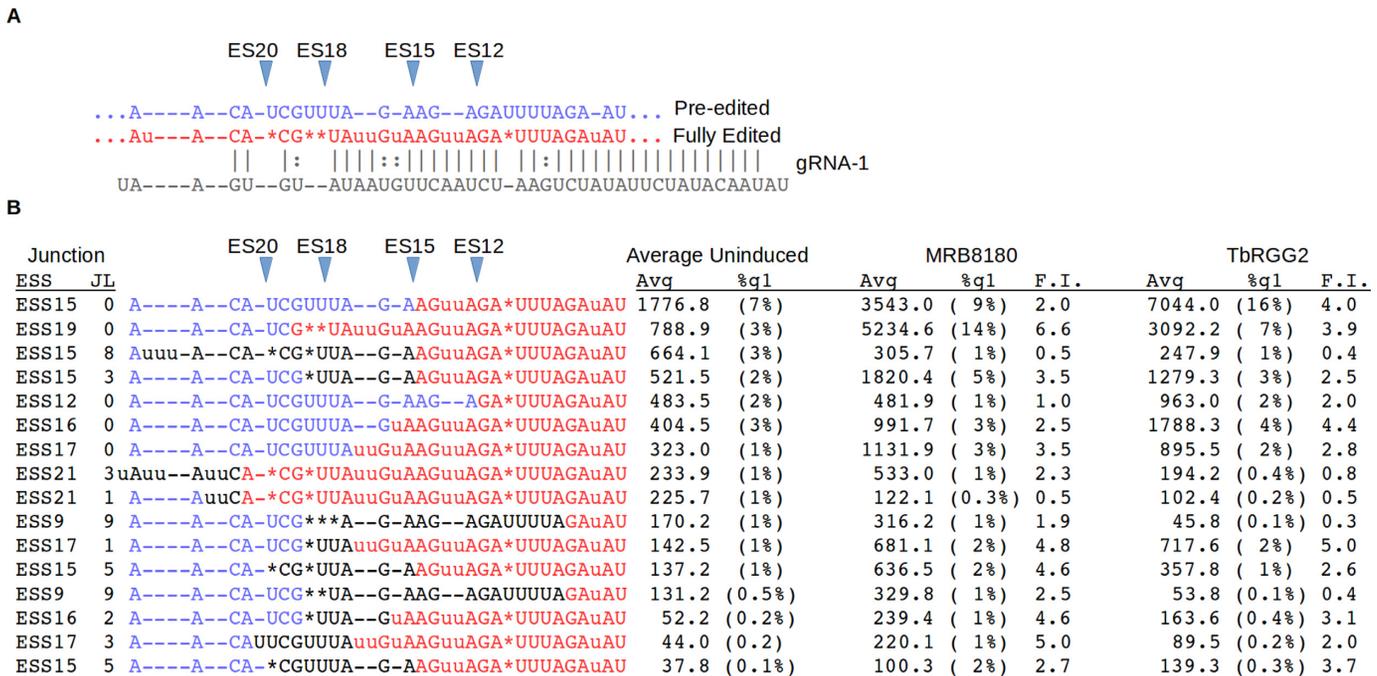


Figure 6. Major intermediate sequences in the region of RPS12 guided by gRNA-1. (A) Schematic shows the sequence of pre-edited mRNA (blue), canonical fully edited mRNA (red) and the first gRNA (gray) as reported by Koslowsky *et al.* (6). Upper case Us indicate U's present in the pre-edited transcript, lower case u's are those added in through RNA editing, and asterisks (*) denote sites where U's have been deleted. (B) Abundant sequences in the gRNA-1 direction region of RPS12 mRNA. Included are any sequences contained within gRNA-1 that were present in >100 average normalized counts in any of the cell lines. The region matching fully edited is shown in red, pre-edited in blue, and the junction region (containing mis-edited sequence) in black. The ESS number and the junction length (JL) are shown to the left of each sequence. To the right is indicated the average number of each sequence detected ($n = 8$ for uninduced, $n = 2$ for induced averages) and the percentage of all sequences whose ESS is within gRNA-1 that is represented by each sequence in the three samples (calculated using the average number of sequences detected and the average number of sequences with an ESS in gRNA-1, denoted %g1). For the MRB8180 and TbRGG2 samples, the fold increase (F.I.) of each sequence compared to the average of uninduced is also indicated.

17 suggests that correct editing can proceed linearly, that is from ES to ES in a 3' to 5' direction, through this region. However, the predominance of ESS15:junction 8 and ESS15:junction 3, wherein the deletion actions at ES18 and ES20 have been performed at least partially but the additions at ES16 and ES17 have not yet been executed, suggests that a non-linear order of editing also often proceeds beyond ESS15, with deletions being executed prior to additions.

TbRGG2 and MRB8180 differentially affect editing path linearity

We next asked whether TbRGG2 or MRB8180 depletion leads to changes in the prominent editing intermediates within the gRNA-1 guided region of RPS12 mRNA. When we examine the changes in the most abundant gRNA-1 directed sequences in MRB8180 and TbRGG2 RNAi induced cells, we find that the majority of the top sequences in the uninduced samples are preserved as major sequences in both knockdown lines (Figure 6). However, the degree to which specific sequences are increased varies between TbRGG2 and MRB8180 depleted cells, and some sequences actually decrease in these cells compared to uninduced (ESS15:junction 8) (Figure 6). Thus, while MRB8180 and TbRGG2 depletion results in an increase of several intrinsic intermediates, the two proteins do not do so uniformly and they are not identical in their impact.

Given that the major sequences in Figure 6 suggest that editing beyond ESS15 can proceed in a linear or non-linear fashion, we next asked whether modifications are observed in specific patterns suggestive of distinct editing paths and whether TbRGG2 or MRB8180 depletion affects paths differentially. Although we cannot definitively say which sequences are directly related as we are examining a steady state population, we know that editing is a multi-step process and thus all sequences exist on a continuum of intermediates that are progressively modified and likely remodified (17,22,62). Given this, we approach this analysis like a phylogenetic classification and build a tree of likely relatedness connecting intermediates that can be generated by single modifications in the sequence. We assume that single modifications are required to move from one sequence to the next and that sequences with successive modifications are likely visible in the steady state population. This assumption is supported by the dual site *in vitro* editing assay published by Alatortsev *et al.* which shows evidence of sequences containing action at only one and also at both sites (63). Thus, it is likely that we will see intermediate sequences containing changes at single editing sites and will be able to connect them stepwise to generate a testable model for editing progression.

To begin to define editing paths, we examined all junction sequences that are canonically edited up to ES15 and whose editing was contained within the region directed by gRNA-1. These sequences represent plausible intermediates

ates of gRNA-1 directed editing. We quantified the number of sequences with modifications at 1, 2 or 3 ESs beyond ESS15 for all possible ES combinations, regardless of whether these modifications were canonical or mis-edited, as the major junctions in the gRNA-1 directed region displayed evidence of both (Figure 6). Interestingly, we did not observe modifications at all possible combinations of ESs in the uninduced control cells or in the induced TbrGG2 or MRB8180 RNAi cell lines, suggesting that the location and order of modification is non-random (Figs S2–S4). For example, almost no sequences were detected with modifications at ES17 but not at ES16. Similarly, while a very small number of sequences have a mis-edit at ES21, no sequences were observed with mis-editing at this site alone in any samples (Table S6). By examining the combinations of ESs having 1, 2 or 3 modifications beyond ES15 (Figure 7A–D), we discerned three likely editing paths (Figure 8). The majority of sequences appear to be edited either through a linear path (Figure 8, ia–c) or a non-linear path beginning with modifications at ES18 (iia–c). A small number of sequences appear to follow a non-linear path beginning at ES20 (iiia–c). In several cases, there are abundant intermediates that could be generated by more than one proposed path. These are denoted in Figure 8 by dotted lines and labeled with the numbers of both paths (e.g. ES16 and ES18, i/iib). While it is possible that the linear path sequences could be generated by remodeling non-linear products, the fact that very few sequences were detected that could represent bridges between the linear and non-linear paths suggests that this is unlikely without a complete undoing of modifications at multiple sites. Together, these data support a model in which editing can proceed by multiple paths through a single gRNA-directed region and can be either linear (8-i) or non-linear (8-ii and iii).

To gain insight into the functions of TbrGG2 and MRB8180 in editing progression, we next asked whether these factors preferentially affect sequences in specific editing paths by comparing the percentage of sequences in each path with one, two and three modifications (Figures 7 and 8). The percent of each group (single, double or triple modification) observed in each cell line is indicated in the boxes below the diagram in Figure 8. The sharpest contrast between the uninduced control and TbrGG2 and MRB8180 knockdown samples is in sequences with three modifications (Figure 8; far right). Compared to the uninduced control, depletion of either TbrGG2 or MRB8180 results in a sharp increase in the product proposed to arise via a linear path, (ic) (57% uninduced versus 92% induced), and there is a corresponding decrease in the proposed non-linear product (iic) (39% uninduced versus 5% induced). This suggests that in the absence of TbrGG2 or MRB8180, more sequences are proceeding through the linear path of editing.

To determine whether TbrGG2 and MRB8180 are causing this shift in an identical way, we traced the buildup of linear products starting with the first modification (Figure 8, far left). Here, we observe that the uninduced cells modify either ES16 or ES18 with approximately equal frequency (ia, iia). However, in both the MRB8180 and TbrGG2 depleted cells, we see a shift towards either ES16 or ES18 and in opposing ways. MRB8180 depletion leads to an increase in sequences harboring modification only

at ES18 (iia) while TbrGG2 depleted cells exhibit an increase in sequences modified only at ES16 (ia). In addition, the TbrGG2 RNAi causes a much larger increase in the number of ESS15:junction 0 sequences than MRB8180 RNAi (Figure 6). When examining the sequence populations having two modifications beyond ES15, we observe that MRB8180 RNAi results in a slight increase in the ES18 derived path (iib), observable in sequences with two modifications, particularly ES18 and ES20 (19% uninduced versus 26% induced; Figure 7C, second sequence). In contrast, TbrGG2 RNAi results in a continued shift to the linear path (ib) as indicated by 46% of uninduced samples entering this path versus 56% of TbrGG2 knockdown cells. This increase coincides with a concomitant decrease in the non-linear path in TbrGG2 depletion, from 35% in uninduced samples versus 26% in TbrGG2 knockdowns. These differences suggest that while both MRB8180 and TbrGG2 cause can shift toward the linear path of editing, TbrGG2 does so more severely and earlier in the editing of the mRNA at and beyond ES15. Although it is possible that the increase in the linear path could be due to a slowing of this path, the decrease in the number of sequences in the non-linear path (e.g. ES18, 20, 23; Figure 7D) in both the TbrGG2 and MRB8180 RNAi samples supports a model in which the non-linear path is decreased, leaving primarily the linear path available upon depletion of TbrGG2 or MRB8180.

TbrGG2 and MRB8170/4160 depletion disrupts sequential progression of editing

Having shown that TbrGG2 and MRB8180 have common functionalities, we next wanted to explore whether the same is true at any level for binding partners MRB8170/4160 and TbrGG2. Throughout RPS12 and ND7-5' mRNAs, we observed only one ESS at which both MRB8170/4160 and TbrGG2 depletion caused exacerbated pausing. This common ESS is RPS12 ESS12, which is found early within the first gRNA (Figure 4A). This ESS site contains a large percentage of sequences with junction length zero, even in uninduced cells (Figure 5), and both TbrGG2 and MRB8170/4160 depletion result in a substantial increase in the number of junction length zero containing sequences at this site (Supplementary Figure S2A, black). Interestingly, we also observe an increase in the number of sequences with very long junctions (>50 ES) at RPS12 ESS12 in both TbrGG2 and MRB8170/4160 RNAi samples (Supplementary Figure S2A, green). A similar increase in the number of sequences with junctions >50 ES is observed at ESS9 upon depletion of MRB8170/4160, as well as at a few other EPSs in the TbrGG2 RNAi sample (ES15, ES16 and ES19; Supplementary Figure S2A). Indeed, of those sequences containing junctions, sequences with junctions >50 ES were among the most abundant at some of these EPSs.

Previously, we reported that in wild type cells long junctions (>50 ES) can represent alternatively edited sequences that are potentially translatable (23). This type of sequence displayed a bipartite structure wherein a small region of conserved mis-editing is followed 5' by an extended region matching canonical fully edited sequence. Editing is then either completed to the 5' end of the transcript or there is a

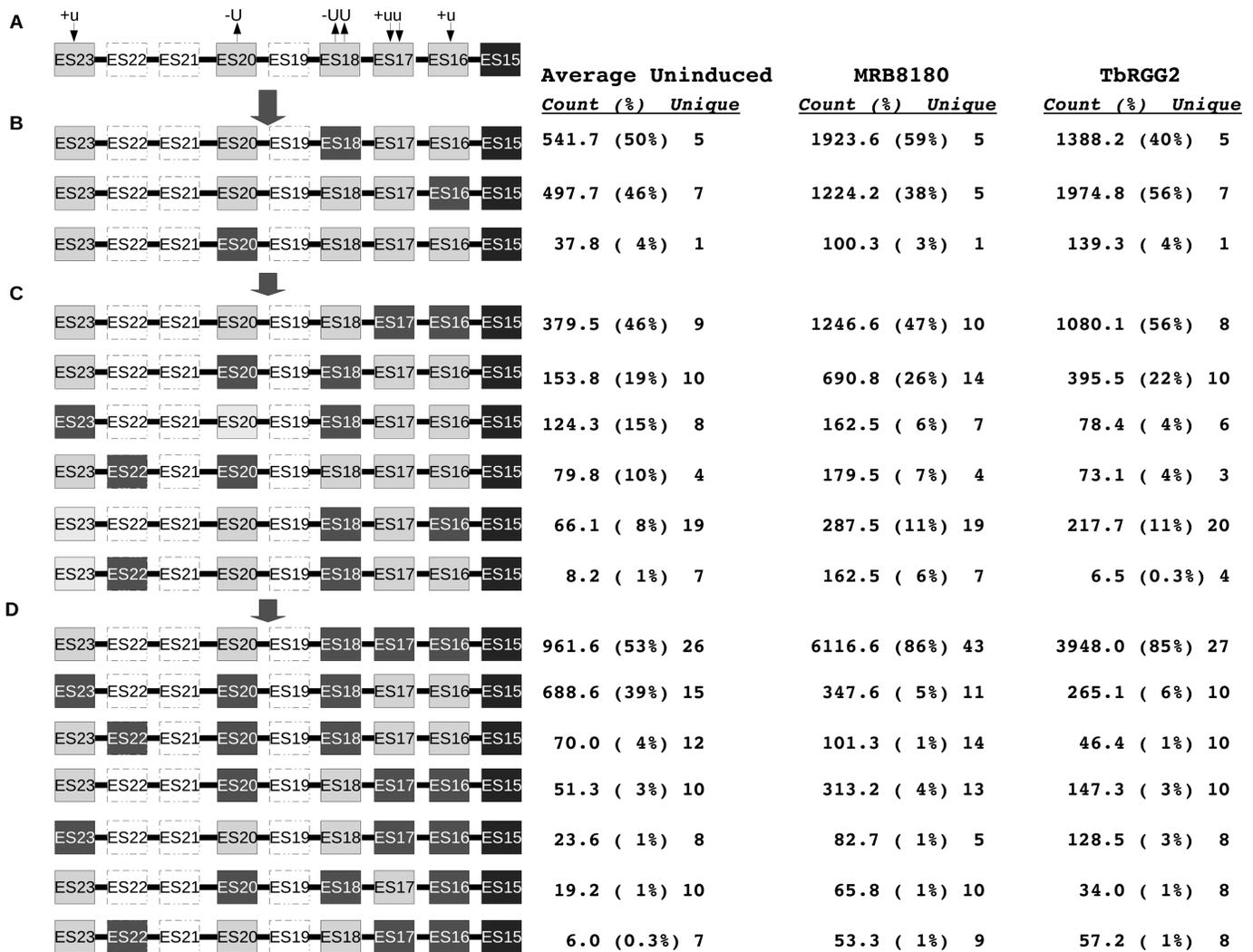


Figure 7. Resolution of editing in gRNA-1 occurs by linear and non-linear modifications. (A) Schematic shows each potential Editing Site (ES) in the region of RPS12 guided by gRNA-1 and 5' of ESS15. ESS15 is black, denoting sequences examined here are constrained to be fully edited up to, and including, ESS15. Light grey blocks indicate ESs that require modification to generate the canonical fully edited sequence, and the number of Us that need to be added to (+u) or deleted from (-U) the sequence at these ESs are shown above each respective block. ESs that do not require modification to generate the canonical fully edited sequence, but can be modified, are shown as white boxes with a dotted border. Boxes shaded dark grey in B-E show locations where editing has modified the sequence, be it by canonical editing or mis-editing. (B) The number of sequences with only a single ES modified beyond ESS15 were calculated for all possible ESs in the gRNA-1 directed region. The top three sites for subsequent modification are shown in the schematic at the left and represent a potentially heterogeneous population of sequences with canonical editing or mis-editing at this next site. The table to the right shows the average number of sequences with each site of modification (count), the percentage of all the sequences with only a single modification beyond ESS15 (%), and the number of unique sequences (unique) represented in this modification for the uninduced samples ($n = 8$) and the induced MRB8180 and TbRGG2 RNAi samples ($n = 2$). The same determination was made for samples with two modified ESs beyond ESS15 (C), and three ESs modified beyond ESS15 (D). Modification combinations were included if they comprised 1% or more of the total sequences with one, two or three modifications in any of the samples.

second region of mis-editing at the 5' edge of the fully edited region, hypothesized to be the true junction of the alternatively edited sequence (Figure 9A). We first asked whether the increase in junctions >50 ES at ESS9 and ESS12 of RPS12 represents an increase in this type of bipartite editing. However, we discovered that the most abundant sequences harboring junctions >50 ES at RPS12 ESS9 and ESS12 in TbRGG2 and MRB8170/4160 depleted cells contain a different type of long junction. These sequences contain regions of mis-editing far 5' in the transcript in the absence of contiguous 3' editing. Such sequences either ex-

hibit a bipartite nature in which two regions of mis-edited sequence are separated by a long stretch of pre-edited sequence (Figure 9B, top) or contain only the 5' mis-edited region (Figure 9B, bottom). Although sequences of this type at RPS12 ESS9 and ESS12 are found in uninduced control cells, they are markedly increased in both TbRGG2 and MRB8170/4160 RNAi cell lines, but not in cells depleted of MRB8180 or GAP1 (Figure 9C). As the general 3' to 5' progression of editing is highly conserved (10,17,64), this type of editing provides evidence of editing action that is disjointed from the standard region of active editing at the

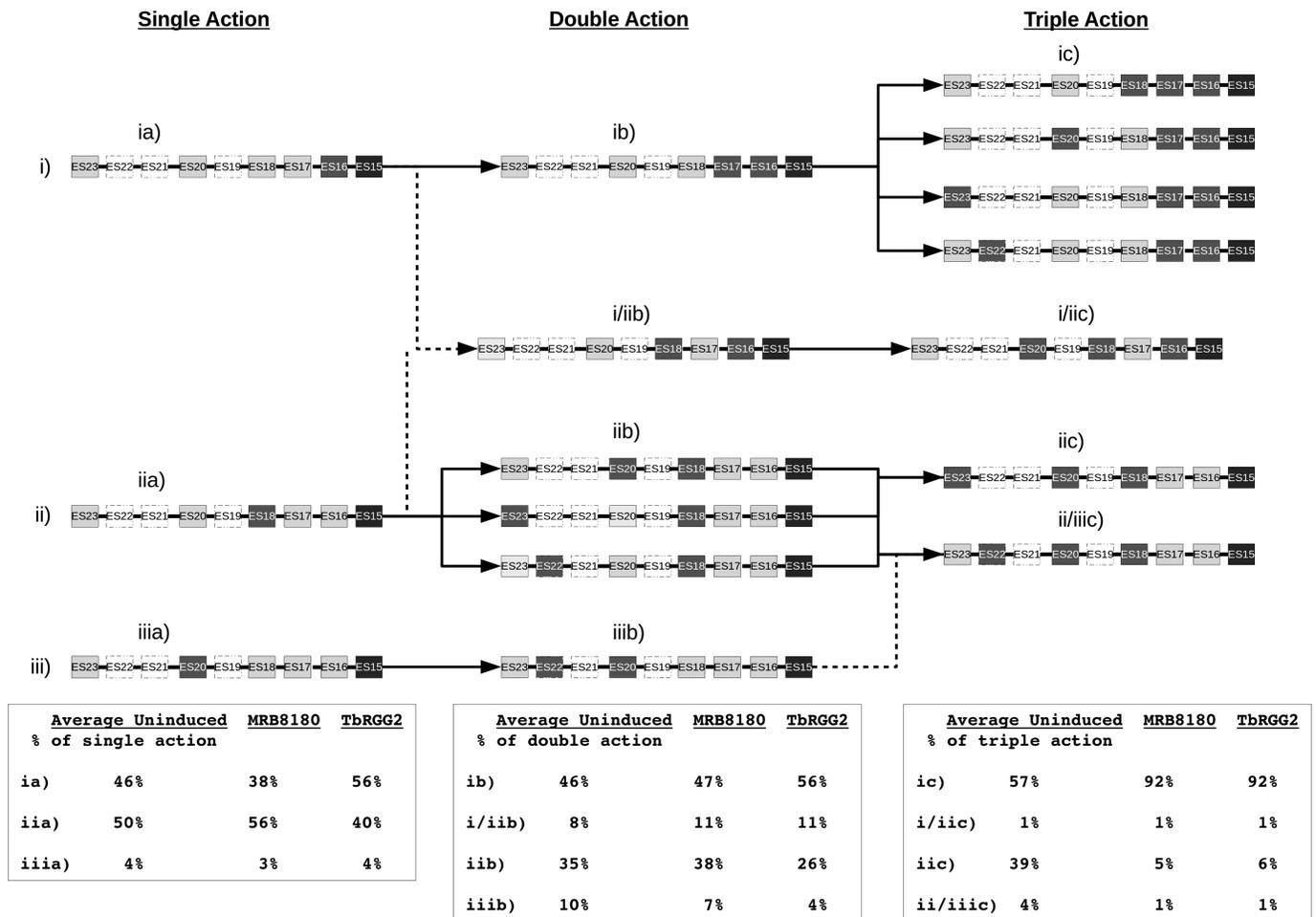


Figure 8. Inferred order of modification suggests editing paths are differentially affected by TbrGG2 and MRB8180. Data from Figure 7B–D were collated to reveal three distinct paths through which editing can proceed beyond ESS15. Shading of boxes designate editing action as in Figure 7. Sequences that could be generated by multiple paths are denoted by dashed lines. Each path is labeled i, ii, or iii at the far left, and subsequent branches for single, double and triple modifications beyond ESS15 are labeled ia, ib, ic, etc. Boxed tables below the diagram indicate the percentage of sequences in each arm for each stage of action (single, double, or triple modification beyond ESS15) in uninduced control and TbrGG2 and MRB8180 knockdown samples. (D) Modification combinations were included if they comprised 1% or more of the total sequences with one, two or three modifications in any of the samples.

3' end of the transcript. This type of sequence could arise through aberrant large-scale mRNA folding or a by a mis-anchored gRNA being used to edit far 5' of what would be the normal region of active editing. As a disjointed region of editing far upstream of the 3' region of active editing appears to be prevented normally, these data suggest that TbrGG2 and MRB8170/4160 are critical for maintenance of the region of active editing.

DISCUSSION

The successful execution of U insertion/deletion RNA editing in trypanosomatids requires numerous enzymatic and non-enzymatic factors (reviewed in (2,3,40)). The current model for non-enzymatic factors includes RESC and its two component subcomplexes, GRBC and REMC. GRBC components are relatively well defined, and some are functionally linked to editing initiation (41–45,50). In contrast, the composition and proposed functions of REMC are not as clear, and have largely relied on functional studies of TbrGG2 and identification of its direct binding partners

(46,50–53). TbrGG2 promotes editing progression and its RNA binding activity is stimulated by MRB8170 (22,53), leading to a model in which REMC as a complex mediates the progression of editing (2,3). What has remained unexplored is whether REMC acts as a single complex with a cohesive function and whether TbrGG2 is truly representative of the function of the complex (2,3). Here, we demonstrate that TbrGG2, MRB8180 and MRB8170/4160 have distinct but overlapping functions in gRNA utilization and editing progression, likely reflecting complex heterogeneity (Figure 10). Our studies further provide insight into the process of editing progression in general, and provide support for the essentiality of mis-edited junction regions in the editing process.

Taken together, the structural and functional analyses of REMC components TbrGG2, MRB8180 and MRB8170/4160 suggest that REMC is heterogeneous in both function and composition. Structurally, we show here that MRB8170 and MRB8180 are partially dependent upon TbrGG2 for stability, but TbrGG2 is not

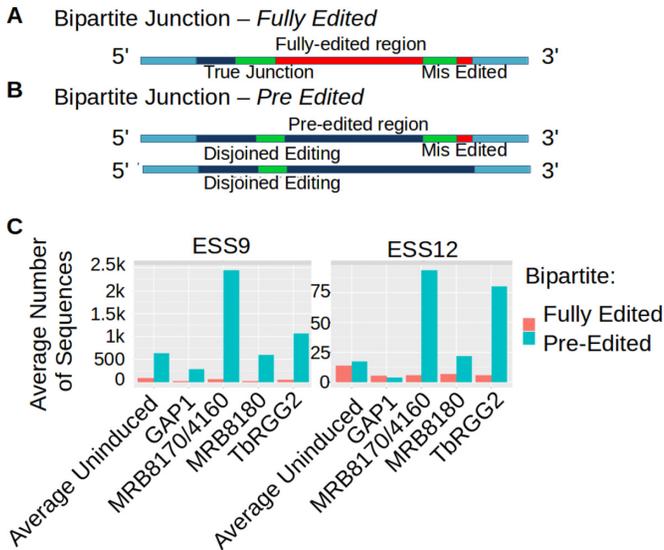


Figure 9. Long junction sequences in RSP12 upon MRB8170/4160 and TbrGG2 knockdown exhibit disjoined regions of editing. (A) Schematic showing a previously reported type of bipartite junction (23) that could generate diversity in ORFs through alternative editing. Red shows the region that is canonically edited, green represents regions of mis-editing, and dark blue represents region matching pre-edited sequence. (B) Schematic shows a long junction that appears to have an edited region that is disjoined from the standard region of active editing at the 3' leading edge of editing. Top schematic shows a bipartite pre-edited junction wherein some canonical editing has been achieved followed by some mis-edited sequence prior to the long region of pre-edited. Bottom schematic shows the bipartite sequences found only at ESS9 of RPS12 and ESS21 of ND7-5' where there does not appear to be standard initiation of editing, only the 5' disjoined editing. (C) The number of RPS12 ESS9 and ESS12 sequences containing junctions >50 ES that are fully edited and bipartite (as in A, orange bars) or pre-edited and bipartite (as in B, green bars) in uninduced control and in each RNAi cell line. ESS9 is exacerbated only in the MRB8170/4160 RNAi and ESS12 is exacerbated in both MRB8170/4160 and TbrGG2 RNAi lines.

dependent on either protein, and neither MRB8180 nor MRB8170 are dependent on each other. We also determined that TbrGG2 is the most abundant of the three proteins, twice that of MRB8170 and thirty times more abundant than MRB8180. This makes it unlikely that all molecules of TbrGG2 are bound to both MRB8180 and MRB8170. These data complement previous studies showing that MRB8170 and MRB8180 both bind to the RRM containing C-terminal region of TbrGG2, and thus may compete for binding to TbrGG2 (51). Moreover, upon glycerol gradient sedimentation, the distribution of MRB8170 and TbrGG2 overlap, but the two clearly do not co-peak (46,47), and MRB8170 reportedly interacts with several non-RESC proteins (53). These findings suggest that some proposed REMC components do not form a single particle but instead associate dynamically with proteins or complexes not currently designated RESC. Functional data also provide evidence for both similarities and differences between TbrGG2, MRB8170/4160 and MRB8180. The large number of EPSs arising upon depletion of each of these three proteins are positioned throughout regions guided by single gRNAs, demonstrating that the proteins are required for the successful progression of editing through a gRNA. Never-

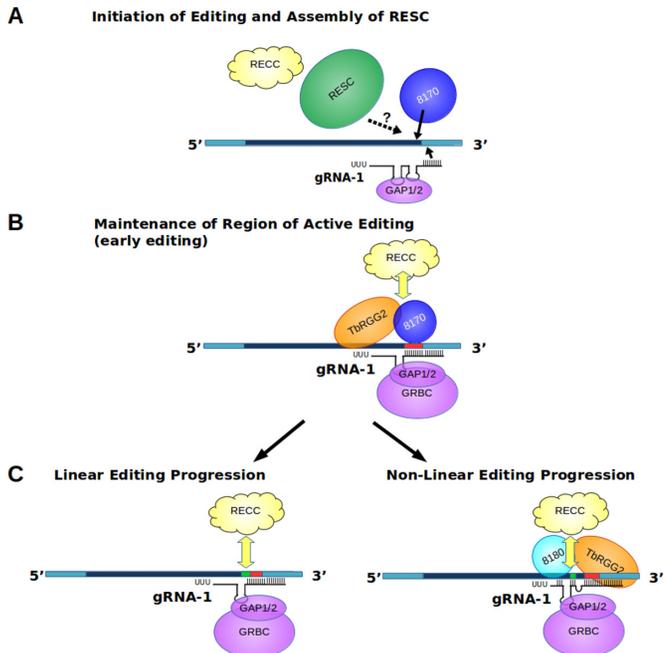


Figure 10. Model of REMC factors' roles in RNA editing. (A) MRB8170/4160 (abbr. 8170) plays a role in initiation and assembly of RESC. (B) MRB8170/4160 and TbrGG2 appear to be required for maintaining the region of active editing early in editing. (C) TbrGG2 and MRB8180 are required for non-linear editing paths.

theless, they do not have identical functions. EPSs caused by TbrGG2 and MRB8180 depletion overlapped significantly, suggesting that these two factors are often needed at the same location and may be in complex at these sites. The lack of overlap with MRB8170/4160 EPSs, suggests that TbrGG2 and MRB8180 are more functionally related to one another than either is to MRB8170/4160. Additionally, TbrGG2 and MRB8180 appear to be required for non-linear editing action (Figure 10C), though in distinct ways. MRB8170/4160 and TbrGG2 appear to be required for maintenance of the region of active editing early in the editing of a transcript (Figure 10B). Both MRB8180 and MRB8170/4160 have transcript-specific effects, with MRB8180 being most important for progression of RPS12 mRNA editing and MRB8170/4160 on ND7-5' progression. Finally, MRB8170/4160 is the only one of the factors examined to substantially affect mRNA editing initiation (Figure 10A). Collectively, these data dispel the notion that TbrGG2, MRB8180 and MRB8170/4160 are found solely in an obligate multimer, and instead support a model wherein REMC factors interact dynamically and may form heterogeneous complex variants that contain subsets of the designated REMC factors.

Examining the function of the TbrGG2 and MRB8180 containing REMCs, we observed that these proteins are required for non-linear editing paths, but they do not appear to affect the overall association of RECC with the mRNA/gRNA duplex. Analysis of partially edited mRNAs within RPS12 gRNA-1 shows that upon depletion of TbrGG2 and MRB8180, the relative proportion of linearly edited sequences increases while the primary non-linear

editing path decreases. The shift from non-linear to linear editing paths likely explains the increase in sequences lacking a junction seen in the TbrGG2 and MRB8180 RNAi cells. For example, the significant increase in no junction sequences at ES15, ES16, ES17 and ES19 in RPS12 reflects the increase in linearly edited intermediates. As both TbrGG2 and MRB8180 are RNA binding proteins and TbrGG2 has RNA melting and annealing capacity, these proteins may facilitate non-linear editing by remodeling the mRNA/gRNA duplex to a productive confirmation. This phenomenon does not appear to be caused by a loss of RECC as the correlation of EPSs arising before incompletely edited deletion sites and the predominance of linearly edited sequences tells us that RECC is still able to associate with the mRNA/gRNA duplex in the absence of TbrGG2 and MRB8180. The observation of sites with incomplete editing is consistent with a recent study by Carnes *et al* showing that single editing sites are likely remodeled during the editing process (62). Further research into the RNA structures that TbrGG2 and MRB8180 bind and remodel will allow us to refine our model for the role these proteins play in promoting non-linear editing modifications.

MRB8170/4160 and TbrGG2 appear to have a common role in the editing process that is distinct from that of MRB8180. When comparing MRB8170/4160 and TbrGG2, we observed only a single site at which depletion of both proteins caused editing to pause, at ESS12 in RPS12 mRNA (Figure 4A). This site in both knockdowns contained an increased number of unusual sequences with disjointed editing, where two regions of editing are intersected by a long region of pre-edited sequence or a lone region of editing far 5' exists (Figure 9). Sequences fitting this pattern at RPS12-ESS12 likely reflect the absence of a *bone fide* TbrGG2 and MRB8170/4160 function because MRB8170/4160 depleted cells also exhibited increases in disjointed editing at RPS12-ESS9 and ND7-5'-ESS21, and TbrGG2 depleted cells display this phenotype at RPS12-ESS15,16, and 19. This curious phenomenon has several plausible explanations. First, sequences with disjointed editing could arise due to an aberrant gRNA inappropriately anchoring far in the 5' region of a transcript. Second, large scale mRNA folding could generate a duplexed region similar in appearance to the *cis*-directed editing of COII, resulting in the recruitment of the COII-specific insertion RECC. As MRB8170/4160 depletion causes this phenomenon to arise further 3' in both the RPS12 and ND7-5' transcripts (ES9 and ES21, respectively) than does the TbrGG2 RNAi, it is likely that MRB8170/4160 is required earlier than, or separately from, TbrGG2 in the assembly of the region of active editing during initiation. This is supported by recent data showing that MRB8170/4160 depletion causes a decrease in the ability of several non-enzymatic RNA editing factors, including TbrGG2, to bind to mRNA (53). MRB8170/4160 depletion may lead to disjointed editing if it is required to bind the mRNA in a way that prevents large scale mRNA folding, or if it acts as a signal for protein or gRNA recruitment such that editing initiation is limited to the 3' most end of the transcript. TbrGG2 exhibits multiple strong and weak interactions with members of REMC, GRBC and RECC

(46,50,51). Loss of TbrGG2 could directly lead to dissolution of RECC or the holoenzyme structure, providing increased substrate for the obscure mechanism that gives rise to the disjointed editing. Thus, MRB8170/4160 and TbrGG2 both appear integral to maintenance of the region of active editing, though it is unclear whether they act in concert or if they do so independently of their interaction or through distinct mechanisms.

Finally, the findings of this study expand our understanding of the the role of junctions in the intrinsic RNA editing process. Junction regions are ubiquitous in the partially edited mRNA population and are observed in kinetoplasts as evolutionarily divergent as the obligate amoebic endosymbiont, *Perkinsela* (17,22,23,26,27,64,65). Junctions have long been hypothesized to be essential intermediates in the editing process. However, it could not be ruled out that they are merely dead-end products that could be subject to regulation or that they may act as a source for genuinely translatable alternative ORFs (17,22-27,64,65). The correlation of the fatal editing defect caused by depletion of TbrGG2 and MRB8180 with the loss of an apparent non-linear editing path provides the most direct evidence to date that the majority of junctions are essential intermediates in the editing process. Similarly, the predominance of linear editing modifications in these RNAi cells suggests that linear modifications alone are insufficient to generate a full length canonically edited mRNA. The primary question this raises is, why are junctions essential? One possibility is that junction formation by non-linear editing could promote the formation of productive mRNA/gRNA duplexes, such as a more complex structure reminiscent of the initial double hairpin loop shown to be bound by RECC components *in vitro* (66). Alternately, the extended complete duplex that would be formed by linear modifications alone could mimic a completely edited region and trigger premature release of the gRNA before the subsequent anchor region has been fully generated. The buildup of ES19:J0 in MRB8180 RNAi cells may represent this kind of phenomenon as this sequence would duplex extensively with gRNA-1, appearing almost identical to the fully edited duplex, but have an incomplete anchor for gRNA-2. By expanding the editing path analysis to encompass longer regions of a transcript and to include more transcripts, we will be able to investigate these hypotheses further and develop a testable model for editing progression through a gRNA.

This study contributes to the knowledge of U insertion/deletion RNA editing by clarifying that elements of the proposed REMC complex are actually heterogeneous in function, and the complex may also have structural heterogeneity. Additionally, our data provide evidence that many junction regions are essential intermediates in the editing process, arising through non-linear editing. We show that REMC factors TbrGG2 and MRB8180 facilitate editing progression through a gRNA-directed region, and this activity appears to be integral to the process of productive non-linear editing. Our data expand the current knowledge of the role of MRB8170/4160 as a critical factor early in editing and provide the first evidence that a low frequency mechanism exists by which editing can occur disjointed from its 3' to 5' directionality. Finally, the methods presented in the paper

can be easily adapted to study many of the enzymatic and non-enzymatic editing factors which could greatly expand our understanding of the delicate interplay of essential editing factors. In sum, our data support a model of RNA editing that is dynamic and complex, requiring substantial plasticity in mRNA/gRNA interactions such that a single transcript is modified and remodified throughout the length of the sequence to generate a final, translatable mRNA.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the University at Buffalo Genomics and Bioinformatics Core, especially Sujith Valiyaparambil, for deep sequencing. We are also grateful to John Panepinto and Natalie McAdams for critical reading of the manuscript, and to past and present Read lab members for helpful discussions.

FUNDING

National Institutes of Health (NIH) [AI061580 to L.K.R., AI125982 to Y.S.]; SUNY Research Foundation (to R.C. and Y.S.). Funding for open access charge: NIH [R01 AI061580].

Conflict of interest statement. None declared.

REFERENCES

- d'Avila Levy, C.M., Boucinha, C., Kostygov, A., Santos, H.L.C., Morelli, K.A., Grybchuk-Ieremenko, A., Duval, L., Votypka, J., Yurchenko, V., Grellier, P. et al. (2015) Exploring the environmental diversity of kinetoplastid flagellates in the high-throughput DNA sequencing era. *Mem. Inst. Oswaldo Cruz*, **110**, 956–965.
- Read, L.K., Lukes, J. and Hashimi, H. (2016) Trypanosome RNA editing: the complexity of getting U in and taking U out. *Wiley Interdiscipl. Rev. RNA*, **7**, 33–51.
- Aphasizheva, I. and Aphasizhev, R. (2016) U-Insertion/deletion mRNA-editing holoenzyme: definition in sight. *Trends Parasitol.*, **32**, 144–156.
- Jensen, R.E. and Englund, P.T. (2012) Network news: the replication of kinetoplast DNA. *Annu. Rev. Microbiol.*, **66**, 473–491.
- Simpson, L., Wang, S.H., Thiemann, O.H., Alfonso, J.D., Maslov, D.A. and Avila, H.A. (1998) U-insertion/deletion Edited Sequence Database. *Nucleic Acids Res.*, **26**, 170–176.
- Koslowsky, D., Sun, Y., Hindenach, J., Theisen, T. and Lucas, J. (2014) The insect-phase gRNA transcriptome in *Trypanosoma brucei*. *Nucleic Acids Res.*, **42**, 1873–1886.
- Suematsu, T., Zhang, L., Aphasizheva, I., Monti, S., Huang, L., Wang, Q., Costello, C.E. and Aphasizhev, R. (2016) Antisense transcripts delimit exonucleolytic activity of the mitochondrial 3' processome to generate guide RNAs. *Mol. Cell*, **61**, 364–378.
- Kirby, L.E., Sun, Y., Judah, D., Nowak, S. and Koslowsky, D. (2016) Analysis of the *Trypanosoma brucei* EATRO 164 bloodstream guide RNA transcriptome. *PLoS Negl. Trop. Dis.*, **10**, e0004793.
- Schnauffer, A., Domingo, G.J. and Stuart, K. (2002) Natural and induced dyskinetoplastic trypanosomatids: how to live without mitochondrial DNA. *Int. J. Parasitol.*, **32**, 1071–1084.
- Maslov, D.A. and Simpson, L. (1992) The polarity of editing within a multiple gRNA-mediated domain is due to formation of anchors for upstream gRNAs by downstream editing. *Cell*, **70**, 459–467.
- Reifur, L. and Koslowsky, D.J. (2008) *Trypanosoma brucei* ATPase subunit 6 mRNA bound to gA6-14 forms a conserved three-helical structure. *RNA (New York, N.Y.)*, **14**, 2195–2211.
- Leung, S.S. and Koslowsky, D.J. (1999) Mapping contacts between gRNA and mRNA in trypanosome RNA editing. *Nucleic Acids Res.*, **27**, 778–787.
- Leung, S.S. and Koslowsky, D.J. (2001) Interactions of mRNAs and gRNAs involved in trypanosome mitochondrial RNA editing: structure probing of an mRNA bound to its cognate gRNA. *RNA (New York, N.Y.)*, **7**, 1803–1816.
- Yu, L.E. and Koslowsky, D.J. (2006) Interactions of mRNAs and gRNAs involved in trypanosome mitochondrial RNA editing: structure probing of a gRNA bound to its cognate mRNA. *RNA (New York, N.Y.)*, **12**, 1050–1060.
- Hermann, T., Schmid, B., Heumann, H. and Goringer, H.U. (1997) A three-dimensional working model for a guide RNA from *Trypanosoma brucei*. *Nucleic Acids Res.*, **25**, 2311–2318.
- Li, F., Herrera, J., Zhou, S., Maslov, D.A. and Simpson, L. (2011) Trypanosome REH1 is an RNA helicase involved with the 3'-5' polarity of multiple gRNA-guided uridine insertion/deletion RNA editing. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 3542–3547.
- Koslowsky, D.J., Bhat, G.J., Read, L.K. and Stuart, K. (1991) Cycles of progressive realignment of gRNA with mRNA in RNA editing. *Cell*, **67**, 537–546.
- Blum, B., Bakalara, N. and Simpson, L. (1990) A model for RNA editing in kinetoplastid mitochondria: 'guide' RNA molecules transcribed from maxicircle DNA provide the edited information. *Cell*, **60**, 189–198.
- Blum, B. and Simpson, L. (1990) Guide RNAs in kinetoplastid mitochondria have a nonencoded 3' oligo(U) tail involved in recognition of the preedited region. *Cell*, **62**, 391–397.
- Koslowsky, D.J., Reifur, L., Yu, L.E. and Chen, W. (2004) Evidence for U-tail stabilization of gRNA/mRNA interactions in kinetoplastid RNA editing. *RNA Biol.*, **1**, 28–34.
- Leung, S.S. and Koslowsky, D.J. (2001) RNA editing in *Trypanosoma brucei*: characterization of gRNA U-tail interactions with partially edited mRNA substrates. *Nucleic Acids Res.*, **29**, 703–709.
- Ammerman, M.L., Presnyak, V., Fisk, J.C., Foda, B.M. and Read, L.K. (2010) TbRGG2 facilitates kinetoplastid RNA editing initiation and progression past intrinsic pause sites. *RNA (New York, N.Y.)*, **16**, 2239–2251.
- Simpson, R.M., Bruno, A.E., Bard, J.E., Buck, M.J. and Read, L.K. (2016) High-throughput sequencing of partially edited trypanosome mRNAs reveals barriers to editing progression and evidence for alternative editing. *RNA (New York, N.Y.)*, **22**, 677–695.
- Sturm, N.R., Maslov, D.A., Blum, B. and Simpson, L. (1992) Generation of unexpected editing patterns in *Leishmania tarentolae* mitochondrial mRNAs: misediting produced by misguiding. *Cell*, **70**, 469–476.
- Ochsenreiter, T. and Hajduk, S.L. (2006) Alternative editing of cytochrome c oxidase III mRNA in trypanosome mitochondria generates protein diversity. *EMBO Rep.*, **7**, 1128–1133.
- Ochsenreiter, T., Cipriano, M. and Hajduk, S.L. (2008) Alternative mRNA editing in trypanosomes is extensive and may contribute to mitochondrial protein diversity. *PLoS One*, **3**, e1566.
- Ochsenreiter, T., Anderson, S., Wood, Z.A. and Hajduk, S.L. (2008) Alternative RNA editing produces a novel protein involved in mitochondrial DNA maintenance in trypanosomes. *Mol. Cell Biol.*, **28**, 5595–5604.
- Rusche, L.N., Cruz-Reyes, J., Piller, K.J. and Sollner-Webb, B. (1997) Purification of a functional enzymatic editing complex from *Trypanosoma brucei* mitochondria. *EMBO J.*, **16**, 4069–4081.
- Rusche, L.N., Huang, C.E., Piller, K.J., Hemann, M., Wirtz, E. and Sollner-Webb, B. (2001) The two RNA ligases of the *Trypanosoma brucei* RNA editing complex: cloning the essential band IV gene and identifying the band V gene. *Mol. Cell Biol.*, **21**, 979–989.
- Cruz-Reyes, J., Zhelonkina, A.G., Huang, C.E. and Sollner-Webb, B. (2002) Distinct functions of two RNA ligases in active *Trypanosoma brucei* RNA editing complexes. *Mol. Cell Biol.*, **22**, 4652–4660.
- Aphasizhev, R., Aphasizheva, I. and Simpson, L. (2003) A tale of two TUTases. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 10617–10622.
- Ernst, N.L., Panicucci, B., Igo, R.P.J., Panigrahi, A.K., Salavati, R. and Stuart, K. (2003) TbMPS7 is a 3' terminal uridylyl transferase (TUTase) of the *Trypanosoma brucei* editosome. *Mol. Cell*, **11**, 1525–1536.
- Panigrahi, A.K., Allen, T.E., Stuart, K., Haynes, P.A. and Gygi, S.P. (2003) Mass spectrometric analysis of the editosome and other

- multiprotein complexes in *Trypanosoma brucei*. *J. Am. Soc. Mass Spectrom.*, **14**, 728–735.
34. Panigrahi, A.K., Schnauffer, A., Ernst, N.L., Wang, B., Carmean, N., Salavati, R. and Stuart, K. (2003) Identification of novel components of *Trypanosoma brucei* editosomes. *RNA (New York, N.Y.)*, **9**, 484–492.
 35. Panigrahi, A.K., Ernst, N.L., Domingo, G.J., Fleck, M., Salavati, R. and Stuart, K.D. (2006) Compositionally and functionally distinct editosomes in *Trypanosoma brucei*. *RNA (New York, N.Y.)*, **12**, 1038–1049.
 36. Stuart, K.D., Schnauffer, A., Ernst, N.L. and Panigrahi, A.K. (2005) Complex management: RNA editing in trypanosomes. *Trends Biochem. Sci.*, **30**, 97–105.
 37. Trotter, J.R., Ernst, N.L., Carnes, J., Panicucci, B. and Stuart, K. (2005) A deletion site editing endonuclease in *Trypanosoma brucei*. *Mol. Cell*, **20**, 403–412.
 38. Carnes, J., Trotter, J.R., Peltan, A., Fleck, M. and Stuart, K. (2008) RNA editing in *Trypanosoma brucei* requires three different editosomes. *Mol. Cell Biol.*, **28**, 122–130.
 39. Carnes, J., Soares, C.Z., Wickham, C. and Stuart, K. (2011) Endonuclease associations with three distinct editosomes in *Trypanosoma brucei*. *J. Biol. Chem.*, **286**, 19320–19330.
 40. Goringer, H.U. (2012) 'Gestalt,' composition and function of the *Trypanosoma brucei* editosome. *Annu. Rev. Microbiol.*, **66**, 65–82.
 41. Hashimi, H., Cicova, Z., Novotna, L., Wen, Y.-Z. and Lukes, J. (2009) Kinetoplastid guide RNA biogenesis is dependent on subunits of the mitochondrial RNA binding complex 1 and mitochondrial RNA polymerase. *RNA (New York, N.Y.)*, **15**, 588–599.
 42. Weng, J., Aphasizheva, I., Etheridge, R.D., Huang, L., Wang, X., Falick, A.M. and Aphasizhev, R. (2008) Guide RNA-binding complex from mitochondria of trypanosomatids. *Mol. Cell*, **32**, 198–209.
 43. Ammerman, M.L., Hashimi, H., Novotna, L., Cicova, Z., McEvoy, S.M., Lukes, J. and Read, L.K. (2011) MRB3010 is a core component of the MRB1 complex that facilitates an early step of the kinetoplastid RNA editing process. *RNA (New York, N.Y.)*, **17**, 865–877.
 44. Ammerman, M.L., Tomasello, D.L., Faktorova, D., Kafkova, L., Hashimi, H., Lukes, J. and Read, L.K. (2013) A core MRB1 complex component is indispensable for RNA editing in insect and human infective stages of *Trypanosoma brucei*. *PLoS one*, **8**, e78015.
 45. Huang, Z., Faktorova, D., Krizova, A., Kafkova, L., Read, L.K., Lukes, J. and Hashimi, H. (2015) Integrity of the core mitochondrial RNA-binding complex 1 is vital for trypanosome RNA editing. *RNA (New York, N.Y.)*, **21**, 2088–2102.
 46. Ammerman, M.L., Downey, K.M., Hashimi, H., Fisk, J.C., Tomasello, D.L., Faktorova, D., Kafkova, L., King, T., Lukes, J. and Read, L.K. (2012) Architecture of the trypanosome RNA editing accessory complex, MRB1. *Nucleic Acids Res.*, **40**, 5637–5650.
 47. Gazestani, V.H., Nikpour, N., Mehta, V., Najafabadi, H.S., Moshiri, H., Jardim, A. and Salavati, R. (2016) A protein complex map of *Trypanosoma brucei*. *PLoS Negl. Trop. Dis.*, **10**, e0004533.
 48. McAdams, N.M., Ammerman, M.L., Nanduri, J., Lott, K., Fisk, J.C. and Read, L.K. (2015) An arginine-glycine-rich RNA binding protein impacts the abundance of specific mRNAs in the mitochondria of *Trypanosoma brucei*. *Eukaryotic Cell*, **14**, 149–157.
 49. Kumar, V., Madina, B.R., Gulati, S., Vashisht, A.A., Kanyumbu, C., Pieters, B., Shakir, A., Wohlschlegel, J.A., Read, L.K., Mooers, B.H.M. *et al.* (2016) REH2C helicase and GRBC subcomplexes may base pair through mRNA and small guide RNA in kinetoplastid editosomes. *J. Biol. Chem.*, **291**, 5753–5764.
 50. Aphasizheva, I., Zhang, L., Wang, X., Kaake, R.M., Huang, L., Monti, S. and Aphasizhev, R. (2014) RNA binding and core complexes constitute the U-insertion/deletion editosome. *Mol. Cell Biol.*, **34**, 4329–4342.
 51. Foda, B.M., Downey, K.M., Fisk, J.C. and Read, L.K. (2012) Multifunctional G-rich and RRM-containing domains of TbrGG2 perform separate yet essential functions in trypanosome RNA editing. *Eukaryotic Cell*, **11**, 1119–1131.
 52. Kafkova, L., Ammerman, M.L., Faktorova, D., Fisk, J.C., Zimmer, S.L., Sobotka, R., Read, L.K., Lukes, J. and Hashimi, H. (2012) Functional characterization of two paralogs that are novel RNA binding proteins influencing mitochondrial transcripts of *Trypanosoma brucei*. *RNA (New York, N.Y.)*, **18**, 1846–1861.
 53. Dixit, S., Muller-McNicoll, M., David, V., Zarnack, K., Ule, J., Hashimi, H. and Lukes, J. (2017) Differential binding of mitochondrial transcripts by MRB8170 and MRB4160 regulates distinct editing fates of mitochondrial mRNA in *Trypanosoma*. *mBio*, **8**, e02288–e02316.
 54. Fisk, J.C., Ammerman, M.L., Presnyak, V. and Read, L.K. (2008) TbrGG2, an essential RNA editing accessory factor in two *Trypanosoma brucei* life cycle stages. *J. Biol. Chem.*, **283**, 23016–23025.
 55. Wickstead, B., Ersfeld, K. and Gull, K. (2002) Targeting of a tetracycline-inducible expression system to the transcriptionally silent minichromosomes of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **125**, 211–216.
 56. Ammerman, M.L., Fisk, J.C. and Read, L.K. (2008) gRNA/pre-mRNA annealing and RNA chaperone activities of RBP16. *RNA (New York, N.Y.)*, **14**, 1069–1080.
 57. Lott, K., Mukhopadhyay, S., Li, J., Wang, J., Yao, J., Sun, Y., Qu, J. and Read, L.K. (2015) Arginine methylation of DRBD18 differentially impacts its opposing effects on the trypanosome transcriptome. *Nucleic Acids Res.*, **43**, 5501–5523.
 58. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodological)*, **57**, 289–300.
 59. Hashimi, H., Zikova, A., Panigrahi, A.K., Stuart, K.D. and Lukes, J. (2008) TbrGG1, an essential protein involved in kinetoplastid RNA metabolism that is associated with a novel multiprotein complex. *RNA (New York, N.Y.)*, **14**, 970–980.
 60. Vondruskova, E., van den Burg, J., Zikova, A., Ernst, N.L., Stuart, K., Benne, R. and Lukes, 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**, 2429–2438.
 61. Kafkova, L., Debler, E.W., Fisk, J.C., Jain, K., Clarke, S.G. and Read, L.K. (2016) The major protein arginine methyltransferase in *trypanosoma brucei* functions as an enzyme-prozyme complex. *J. Biol. Chem.*, **292**, 2089–2100.
 62. Carnes, J., McDermott, S., Anupama, A., Oliver, B.G., Sather, D.N. and Stuart, K. (2017) In vivo cleavage specificity of *Trypanosoma brucei* editosome endonucleases. *Nucleic Acids Res.*, **45**, 4667–4686.
 63. Alatortsev, V.S., Cruz-Reyes, J., Zhelonkina, A.G. and Sollner-Webb, B. (2008) *Trypanosoma brucei* RNA editing: coupled cycles of U deletion reveal processive activity of the editing complex. *Mol. Cell Biol.*, **28**, 2437–2445.
 64. David, V., Flegontov, P., Gerasimov, E., Tanifuji, G., Hashimi, H., Logacheva, M.D., Maruyama, S., Onodera, N.T., Gray, M.W., Archibald, J.M. *et al.* (2015) Gene loss and error-prone RNA editing in the mitochondrion of *Perkinsella*, an endosymbiotic kinetoplastid. *MBio*, **6**, e01498–01415.
 65. Sturm, N.R. and Simpson, L. (1990) Partially edited mRNAs for cytochrome b and subunit III of cytochrome oxidase from *Leishmania tarentolae* mitochondria: RNA editing intermediates. *Cell*, **61**, 871–878.
 66. Kala, S., Moshiri, H., Mehta, V., Yip, C.W. and Salavati, R. (2012) The oligonucleotide binding (OB)-fold domain of KREPA4 is essential for stable incorporation into editosomes. *PLoS One*, **7**, e46864.