Evolution of RNA editing in kinetoplastid protozoa

Dmitri A. Maslov*, Herbert A. Avila*, James A. Lake*† & Larry Simpson*‡§

* Department of Biology, † Molecular Biology Institute, ‡ Howard Hughes Medical Institute, University of California, Los Angeles, 5–748 MacDonald Building, 10833 Le Conte Avenue, Los Angeles, California 90024, USA
§ To whom correspondence should be addressed

THE editing of RNA in trypanosomatid mitochondria involves the insertion and occasional deletion of uridine residues within coding regions of maxicircle messenger RNA transcripts. The extent to which the transcripts of homologous genes undergo editing differs in different species. In some, entire genes are edited (pan-editing), whereas in others, editing is limited to the 5′ termini of editing domains (5′ editing)[1,2]. Here we investigate which type of editing is ancestral and which is derived, by analysing RNA editing in the different lineages, using a kinetoplastid phylogeny reconstructed from nuclear small subunit ribosomal RNA sequences. We conclude that the ancestral cryptogenes were pan-edited, and we hypothesize that the 5′-edited homologues were generated by several independent events from partially edited RNAs, in which case editing may be a more primitive mechanism than previously thought.

The order Kinetoplastida contains two major subgroups that differ in morphological characters and life cycles: the trypanosomatids and the bodonids and related cryptobiids[3]. RNA editing occurs in all trypanosomatid species so far examined and also in the cryptobiid, *Trypanoplasma borreli* (data not shown). A rooted phylogenetic tree constructed using nuclear small subunit rRNA sequences from monogenetic trypanosomatids (which are parasitic in a single invertebrate host), digenetic trypanosomatids (which parasitize two hosts), and a cryptobiid, is shown in Fig. 1. Using *Euglena gracilis* as an outgroup, the tree is rooted in the *T. borreli* branch. The extreme divergence of the *E. gracilis* sequence from those of the trypanosomatids raises the possibility that this root is artefactually produced by unequal rate effects. However, additional support for this root is provided by the presence of two flagella in both cryptobiids[4] and euglenoids[5], and one flagellum in all other taxa. Within the trypanosomatid branch, the most deeply diverging lineage is the salivarian African trypanosome, *Trypanosoma brucei*, followed by the stercorarian trypanosomes, *T. cruzi* and *T. sp.* E1–CP (a fish parasite). The monogenetic species *Blastocrithidia culicis* and *Herpetomonas muscarum* branch off next, together with a monophyletic clade containing the monogenetic species *Crithidia fasciculata* and the digenetic species *Leishmania tarentolae*.

To investigate the evolution of editing, we studied three mitochondrial genes for which both pan-editing and 5′ editing has been reported. These genes encode NADH dehydrogenase subunit 7 (ND7), cytochrome *c* oxidase subunit III (COIII), and maxicircle unidentified reading frame 4 (MURF4), which may represent ATPase subunit 6 (ref. 6).

The MURF4 gene is completely pan-edited in *T. brucei*, and is 5′ pan-edited in *L. tarentolae*[6]. We have sequenced the MURF4 genes from the remaining taxa to determine their editing type. The evidence that MURF4 is pan-edited in *T. cruzi* is shown in Fig. 2. An example of a complementary DNA clone of a partially edited RNA amplified by polymerase chain reaction PCR is shown which has an open reading frame (ORF) at the 3′ end which translates into a sequence highly similar to the *T. brucei* MURF4 sequence. The upstream sequence represents a misedited sequence, such as those that frequently occur at 'junction regions' in editing intermediates[7,8]. As the entire DNA

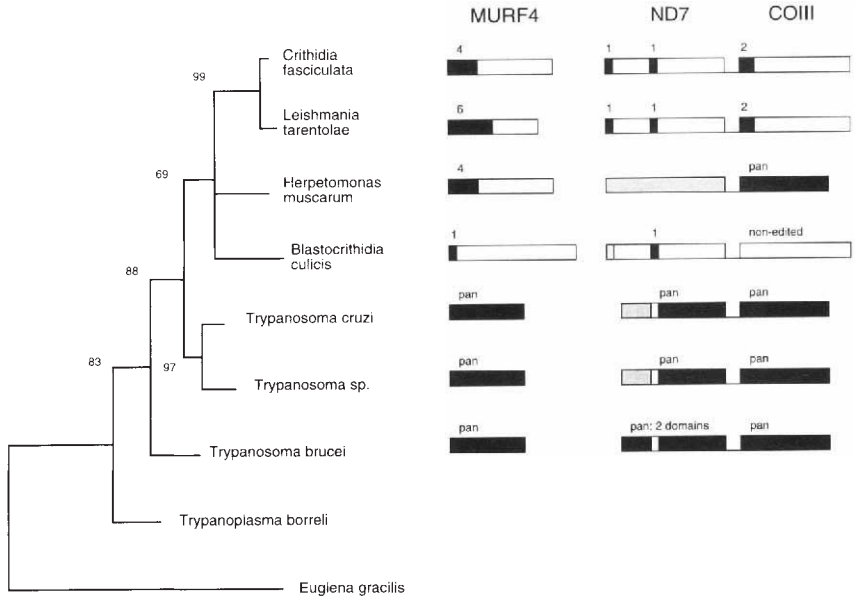

FIG. 1 Phylogeny of kinetoplastid RNA editing. Aligned sequences of 18S rRNAs from *T. brucei*[25], *T. cruzi*[26], *L. tarentolae*[27], *C. fasciculata*[28] and *E. gracilis*[25] were retrieved from the Ribosomal Database Project[29] and sequences of *T. borreli*, *Trypanosoma* sp. E1-CP, *B. culicis* and *H. muscarum* were obtained by sequencing several independently cloned PCR-amplified gene fragments by standard methods. GenBank accession numbers are L14840 (*T. borreli*), L14841 (*Trypanosoma* sp. E1-CP), L18872 (*H. muscarum*), and U05679 (*B. culicis*). Sequences were aligned using the interactive editor[30]. The culture of *T. borreli*[31] was kindly provided by J. Lom. DNA from E1-CP cells was isolated in the Lom laboratory by D.A.M. *H. muscarum* (30261) and *B. culicis* (30268) were obtained from the ATCC. The majority consensus parisomony tree[32], which was constructed using 200 bootstrap replicates, is shown. This tree is consistent with the three most parismonious trees, the first of which (1,174 steps) showed *B. culicis* branching off earliest, followed by *H. muscarum* and then by the *C. fasciculata* and *L. tarentolae* clade. The second tree (2 more steps) made *H. muscarum* and *B. culicis* a monophyletic group, and the third tree (4 more steps) had *H. muscarum* and *B. culicis* reversed from their order in the first tree. The maximum likehood method[33] (fast DNAml program provided by G. Olsen) produced a topology identical to that of the second most parsimonious tree. Paralinear distances[34], a new algorithm based on very general assumptions, gave the same topology as the tree shown here. Our conclusions on the evolution of editing remain valid with any one of these topologies, and are even stronger for the third most parsimonious tree. The rooted tree shown here indicates that a digenetic life cycle may have evolved independently several times. This differs from our previous speculation that the digenetic life cycle evolved late in the evolution of this group of organisms[35] and is consistent with the previous hypothesis of Hoare[36]. A diagrammatic representation of the extent of editing in MURF4, ND7 and COIII is shown on the right. Black boxes correspond to pre-edited regions, white boxes to non-edited regions, and grey areas to a lack of information. Numbers above the PERs indicate the actual number of guide RNAs involved (that is, the number of editing blocks) or the estimated miminal number of gRNAs required. GenBank accession numbers are U05816 (*C. fasciculata* MURF4), U05812 (*H. muscarum* MURF4), U05813 (*B. culicis* MURF4), U05879 (*T. cruzi* MURF4), U05818 (*Trypanosoma* sp. E1-CP MURF4), U05817 (*B. culicis* ND7), U05881 (*T. cruzi* ND7), U05815 (*Trypanosoma* sp. E1-CP ND7), U05814 (*B. culicis* COIII), U05878 (*T. cruzi* COIII), U05815 (*Trypanosoma* sp. E1-CP COIII).

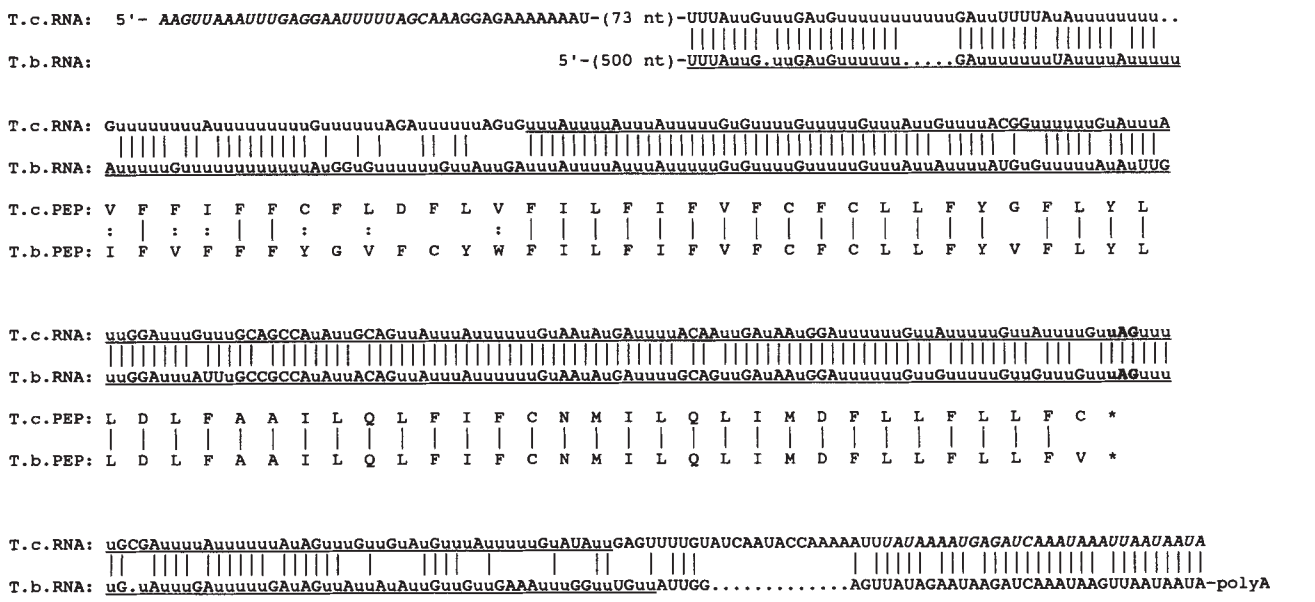

FIG. 2 MURF4 is a pan-edited cryptogene in *T. cruzi*. The sequence of one particular cDNA clone of a partially edited mRNA molecule is shown. The portion that corresponds to the fully edited consensus sequence, which was confirmed by sequencing additional clones, underlined. This sequence was aligned with the *T. brucei* edited sequence according to the amino-acid alignment shown below (indicated by PEP). Matches are indicated by vertical bars and conservative changes by colons. Sequences used to PCR-amplify RNAs are italicized. Uridines added by editing are shown in lower case, termination codons in bold, and gaps by dots.

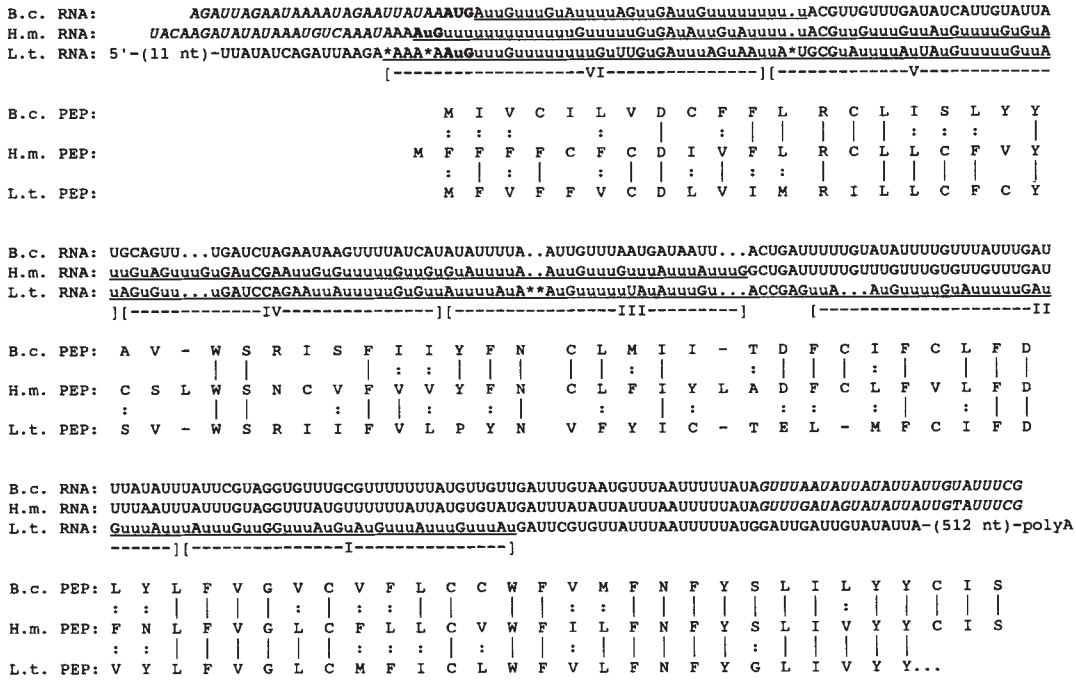

FIG. 3 Alignment of edited MURF4 sequence from *B. culicis*, *H. muscarum* and *L. tarentolae*. Nucleotide alignments were made according to amino-acid sequence alignments. *L. tarentolae* editing blocks are indicated below the RNA sequences. Uridines added by editing are shown in lower case, uridines deleted by asterisks. Other designations are as for Fig. 2.

sequence of the PCR-amplified MURF4 fragment from *T. cruzi* is extremely (G+A)-rich (data not shown), as in the case of the pan-edited gene in *T. brucei*, and as editing proceeds 3′ to 5′, we conclude that MURF4 is completely pan-edited in *T. cruzi*. Sequence analysis of the amplified MURF4 maxicircle fragment from another stercorarian trypanosome, *Trypanosoma* sp. E1-CP, showed that this gene is also pan-edited (data not shown).

The phylogenetic positions of *B. culicis*, *H muscarum* and *C. fasciculata* made it important to investigate the editing of the MURF4 gene in these species also. Several clones of edited *B. culicis* MURF4 RNAs were obtained which contained 23 uridines inserted in 7 sites. The ORF derived from the edited sequence together with the non-edited downstream genomic sequence is similar to that obtained from edited *L. tarentolae* MURF4 sequence, thereby precisely localizing the pre-edited region (PER) in *B. culicis* (Fig. 3). By comparison with *L. tarentolae*, which has a total of six contiguous editing blocks mediated by six overlapping genomic (g)RNAs[9], only a single gRNA corresponding to the last editing block would be required in *B. culicis* for the observed editing events. A fully edited sequence for MURF4 mRNA from *H. muscarum* was deduced from a collection of overlapping partially edited RNAs. Translation of

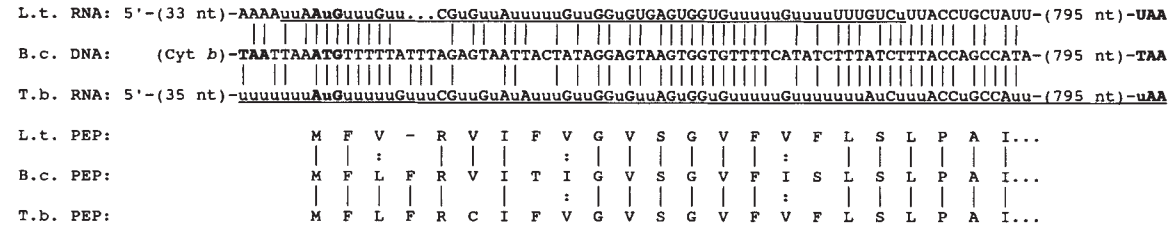

FIG. 4 Comparison of the COIII DNA sequence of *B. culicis* with fully edited mRNA sequences from *L. tarentolae* and *T. brucei*. The edited portions of the sequences are underlined. The initiation and termination codons of COIII in the three species and the termination codon of the upstream cytochrome *b* gene in *B. culicis* are shown in bold. The translated amino-acid sequence is shown below.

the edited and non-edited sequences allowed a precise localization of the PER (Fig. 3). This editing would require minimally four overlapping gRNAs which would correspond to the gRNAs for the four last editing blocks in *L. tarentolae*. A similar situation was found in the case of the MURF4 gene in *C. fasciculata*, which exhibited a PER four editing blocks in length (data not shown).

The genes for ND7 and COIII are adjacent on the maxicircle. DNA and edited RNA sequences have been reported for *C. fasciculata*[10], *L. tarentolae*[11], *T. brucei*[12,13] and *H. muscarum*[14] (COIII only). The ND7 mRNA in *T. brucei* is pan-edited in two large domains, whereas in *L. tarentolae* and in *C. fasciculata*, the ND7 mRNA is 5′ edited in two separate editing blocks which correspond exactly to the 5′ termini of the two editing domains in *T. brucei*. Similarly, the two contiguous editing blocks in the *L. tarentolae* COIII mRNA correspond exactly to the 5′ terminus of the single large pan-edited domain in *T. brucei*. The COIII gene in *H. muscarum* is pan-edited[14]. We have extended this comparative analysis to the phylogenetically significant stercorarian trypanosomes and the *B. culicis* lineages (Fig. 1). Sequence analysis of PCR-amplified maxicircle fragments from *T. cruzi* and *Trypanosoma* sp. E1-CP (data not shown) showed that COIII and ND7 (at least in domain I) are also pan-edited as in *T. brucei*. But in the case of *B. culicis*, the ND7 domain I is represented by a single 5′ editing block, as in *L. tarentolae* and *C. fasciculata* (data not shown). On the other hand, the COIII gene in *B. culicis* (Fig. 4) is completly non-edited, as shown by the fact that the translated ORF has a high similarity with COIII protein sequences from other trypanosomatids (69% identity and 84% similarity with the *T. brucei* COIII sequence for example). This is an important finding because it represents the first example of a gene that is present in a completely non-edited version in one species and in a pan-edited version in another species.

The multiple forms of the edited genes for which comparative sequence analysis is available are shown in Fig. 1 on the kinetoplastid phylogenetic tree. Pan-edited MURF4, ND7 and COIII genes are mainly found in the early diverging branches, multiple forms of less edited genes are found in the later separated lineages. The most parsimonious interpretation is that pan-editing of the MURF4 and COIII genes is the primitive state. For COIII, only two independent changes of pan-editing to 5′ editing are required: one for the *Blastocrithidia* branch and another for the *Crithidia–Leishmania* clade if pan-editing is ancestral, versus three independent changes from 5′ to pan-editing if 5′ editing is primitive. For the MURF4 gene, only one change of pan-editing to 5′ editing is required. We therefore conclude that pan-editing is a primitive character state that was present in an ancestral trypanosomatid.

Our findings are consistent with the hypothesis that at several times during the course of evolution of the trypanosomatid flagellates, pan-edited genes were replaced with partially edited genes. One possible mechanism is that partially edited RNAs were reverse-transcribed into cDNAs, which were incorporated into the maxicircle genome and replaced the original pan-edited genes[1,15–17]. The existence of a non-edited COIII gene in *B. culicis* indicates that such a mechanism could also completely eliminate editing. This raises the possibility that the ND1, ND4, ND5 and COI genes, which are non-edited in all kinetoplastids examined, may have originally been pan-edited and, in the course of evolution, were replaced by fully edited cDNAs.

A possible driving force for the appearance of lineages in which pan-edited genes have been substituted by partially or completely edited reverse transcripts could be the loss of minicircle-encoded gRNA genes. In *L. tarentolae* each minicircle encodes a single gRNA which only overlaps other gRNAs in the anchor sequence region[18], unlike in *T. brucei*, in which gRNAs overlap even in the guiding region[19]. We have shown previously that the relative abundance of different minicircle sequence classes in *L. tarentolae* varies from as few as 30 copies per network of 10,000 minicircles to as many as 2,000 copies[9]. As minicircles apparently segregate randomly between daughter networks, there is a finite chance that a daughter cell will not receive at least one copy of a particular minicircle. A more dramatic loss of one or more minicircle classes could also occur by the phenomenon of 'transkinetoplastidy'[20]. The subsequent loss of the corresponding gRNA would result in the disruption of an editing cascade. If we make the assumption that a low level of homologous retroposition of partially edited RNAs substituting for pan-edited genes occurs continuously among the 50 or so maxicircle molecules per network, then the cells carrying a retroposed partially edited gene would not require the missing gRNA and would survive. Additional support for this hypothesis is provided by the observation that a loss of minicircle classes is observed in nature (*Trypanosoma equiperdum*)[21,22] and in the laboratory (*L. tarentolae*-UC strain) (O. Thiemann, D.A.M. and L.S., unpublished results) when cyclical transmission has been prevented. The preservation of pan-editing in cyclically transmitted *T. brucei* is possibly due to the utilization of editing as a translational control mechanism during the complex life cycle of this species[23]. The maintenance of an extremely large redundant gRNA repertoire could be an adaptation to prevent the loss of gRNA genes or could indicate that this also is a primitive state.

When and how was RNA editing acquired by the kinetoplastids? Editing could have evolved in the ancestral kinetoplastids by a gene duplication mechanism[1,24], or alternatively could have been inherited from the original proto-mitochondrion. To address these questions, an investigation of the possibility of RNA editing in euglenoids and in other mitochondrial-containing protists from the earliest branching lineages, or in eubacteria related to the original endosymbionts, may prove informative. □

Received 3 November 1993; accepted 12 January 1994.

1. Simpson, L., Maslov, D. A. & Blum, B. in *RNA Editing—The Alteration of Protein Coding Sequences of RNA* (ed. Benne, R.) 53–85 (Horwood, New York, 1993).
2. Stuart, K. in *RNA Editing—The Alteration of Protein Coding Sequences of RNA* (ed. Benne, R.) 25–52 (Horwood, New York, 1993).
3. Vickerman, K. in *Biology of the Kinetoplastida* (eds Lumsden, W. H. R & Evans, D. A.) 1–34 (Academic, London, New York and San Francisco, 1976).
4. Lom, J. in *Biology of the Kinetoplastida* (eds Lumsden, W. H. R. & Evans, D. A.) 267–337 (Academic, London, New York and San Francisco, 1976).
5. Kivic, P. A. & Walne, P. L. *Origins of Life* 13, 269–288 (1984).
6. Bhat, G. J., Koslowsky, D. J., Feagin, J. E., Smiley B. L. & Stuart, K. *Cell* **61,** 885–894 (1990).

7. Sturm, N. R., Maslov, D. A., Blum, B. & Simpson, L. *Cell* **70,** 469–476 (1992).
8. Decker, C. J. & Sollner-Webb, B. *Cell* **61,** 1001–1011 (1990).
9. Maslov, D. A. & Simpson, L. *Cell* **70,** 459–467 (1992).
10. Van der Spek, H. *et al. EMBO J.* **10,** 1217–1224 (1991).
11. Blum, B., Bakalara, N. & Simpson, L. *Cell* **60,** 189–198 (1990).
12. Feagin, J. E., Abraham, J. M. & Stuart, K. *Cell* **53,** 413–422 (1988).
13. Koslowsky, D. J., Bhat, G. J., Perrollaz, A. L., Featin, J. E. & Stuart, K. *Cell* **62,** 901–911 (1990).
14. Landweber, L. F. & Gilbert, W. *Nature* **363,** 179–182 (1993).
15. Landweber, L. F. *BioSystems* **28,** 41–45 (1992).
16. Fink, G. R. *Cell* **49,** 5–6 (1987).
17. Derr, L. D. & Strathern, J. *Nature* **361,** 170–173 (1993).
18. Sturm, N. R. & Simpson, L. *Nucleic Acids Res.* **19,** 6277–6281 (1991).
19. Corell, R. A. *et al. Nucleic Acids Res.* **21,** 4313–4320 (1993).
20. Lee, S. T., Tarn, C. & Chang, K. P. *Molec. Biochem. Parasit.* **58,** 187–203 (1993).
21. Barrois, M., Riou, G. & Galibert, F. *Proc. natn. Acad. Sci. U.S.A.* **78,** 3323–3327 (1982).
22. Frasch, A. *et al. Biochim. biophys. Acta* **607,** 397–410 (1980).
23. Koslowsky, D. J., Riley G. R., Feagin, J. E. & Stuart, K. *Molec. cell. Biol.* **12,** 2043–2049 (1992).
24. Covello, P. S. & Gray, M. W. *Trends Genet.* **9,** 265–268 (1993).
25. Sogin, M., Elwood, H. & Gunderson, J. *Proc. natn. Acad. Sci. U.S.A.* **83,** 1383–1387 (1986).
26. Hernandez, R., Rios, P., Valdes, a. & Pinero, D. *Molec. Biochem. Parasit.* **41,** 207–212 (1990).
27. Briones, M. R. S. *et al. Molec. Biochem. Parasit.* **53,** 121–128 (1992).
28. Schnare, M., Collings, J. & Gray, M. *Curr. Genet.* **10,** 405–410 (1986).
29. Olsen, G. J. *et al. Nucleic Acids Res.* **20** suppl., 2199–2200 (1992).
30. Olsen, G. J. *Sequence Editor and Analysis Program* (University of Illinois, Urbana, 1990).
31. Peckova, H. & Lom, J. *Parasit. Res.* **76,** 553–558 (1990).
32. Swofford, D. L. *PAUP: Phylogenetic Analysis Using Parsimony, Version* 3.1 (Illinois Natural History Survey, Champaign 1993).
33. Felsenstein, J. *PHYLIP (Phylogeny Inference Pacakage) Version* 3.4 (University of Washington, Seattle 1991).
34. Lake, J. A. *Proc. natn. Acad. Sci. U.S.A.* **91,** 1455–1459 (1994).
35. Lake, J., de la Cruz, V., Ferreira, P., Morel, C. & Simpson, L. *Proc. natn. Acad. Sci. U.S.A.* **85,** 4779–4783 (1988).
36. Hoare, C. A. *The Trypanosomes of Mammals* (Blackwell Scientific, Oxford, 1972).

ACKNOWLEDGEMENTS. We thank G. Olsen and M. Sogin for computer programs and M. Peris for assistance with computer analysis and for discussion. This work was suported in part by grants from the NIH (L.S. and H.A.A.), UNDP/World Bank/WHO/TDR and the Rockefeller Foundation (L.S.), and the NSF (J.A.L.). D.A.M. is on leave from Moscow State University.

Opposite voltage gating polarities of two closely related connexins

Vytas K. Verselis, Christopher S. Ginter & Thaddeus A. Bargiello

Department of Neuroscience, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461-1602, USA

THE molecular mechanisms underlying the voltage dependence of intercellular channels formed by the family of vertebrate gap junction proteins (connexins) are unknown. All vertebrate gap junctions are sensitive to the voltage difference between the cells, defined as the transjunctional voltage, V_j (refs 1, 2), and most appear to gate by the separate actions of their component hemichannels[3–8]. The heterotypic Cx32/Cx26 junction displays an unpredicted rectification that was reported to represent a novel V_j dependence created by hemichannel interactions, mediated in part by the first extracellular loop E1 (ref. 9). Here we show that aspects of the rectification of Cx32/Cx26 junctions are explained by opposite gating polarities of the component hemichannels, and that the opposite gating polarity of Cx32 and Cx26 results from a charge difference in a single amino-acid residue located at the second position in the N terminus. We also show that charge substitutions at the border of the first transmembrane (M1) and E1 domains can reverse gating polarity and suppress the effects of a charge substitution at the N terminus. We conclude that the combined actions of residues at the N terminus and M1/E1 border form a charge complex that is probably an integral part of the connexin voltage sensor. A consistent correlation between charge substitution and gating polarity indicates that Cx26 and Cx32 voltage sensors are oppositely charged and that both move towards the cytoplasm upon hemichannel closure.

Comparison of E1 sequences of cloned connexins[10,11] indicates that the first two residues at the proposed M1/E1 border are different in Cx26 (KE) from those of most other members of the connexin gene family (ES). We exchanged these residues to create Cx32*KE and Cx26*ES and expressed them in *Xenopus* oocytes. The steady-state changes in normalized junctional conductance, G_j, in response to V_j for homotypic wild-type and mutant connexins are illustrated in Fig. 1. Cx32 homotypic junctions gate symmetrically (Fig. 1*a*), with identical kinetic and steady-state properties about $V_j = 0$. Cx26 homotypic junctions gate asymmetrically (Fig. 1*d*) owing to the presence of an additional dependence on the membrane potential termed V_m, or $V_{j\cdot 0}$ (refs 9, 12–14). Best fits of the steady-state data to a Boltzmann relation indicate a calculated gating charge of ~2 for Cx32 and ~4 for Cx26. Both Cx32*KE (Fig. 1*b*) and Cx26*ES (Fig. 1*e*) homotypic junctions retain symmetry about $V_j = 0$, but are more sensitive to V_j and have substantially faster kinetics than their wild-type counterparts. Although the $G_j - V_j$ relations of the mutant homotypic junctions, particularly Cx32*KE, have complex shapes that cannot be fitted well by a two-state Boltzmann relation, the maximum slopes in each case are steeper than those of the wild-type junctions, consistent with a substantial increase in the calculated gating charge (see Fig. 1 legend). All the currents decay in a multiexponential fashion (fits to data not shown), indicating, as do the steady-state $G_j - V_j$ relations, that multiple voltage-gated transitions occur in these channels.

In the heterotypic junction Cx26*ES/Cx26 (Fig. 1*f*), both the kinetic and the steady-state properties are asymmetric about $V_j = 0$, resembling Cx26 when the cell expressing Cx26 is made relatively positive, and resembling Cx26*ES when the cell expressing Cx26*ES is made relatively positive. This shows that the component hemichannels retain their characteristics as inferred from the corresponding homotypic junctions, and that Cx26 and Cx26*ES hemichannels close on relative positivity on their cytoplasmic side. The same polarity of closure has been shown for Cx38, Cx37 and Cx40 (refs 4–8). In the corresponding heterotypic junction Cx32*KE/Cx32 (Fig. 1*c*), the respective kinetic and steady-state hemichannel properties of Cx32 and Cx32*KE are also retained. But the surprising result is that the polarity of V_j sensitivity is opposite to that of Cx26, with Cx32 and Cx32*KE hemichannels closing on relative negativity on their cytoplasmic side (see also refs 15, 16).

To confirm the assignment of positive gating polarity to Cx26 and negative to Cx32, we examined the behaviour of heterotypic Cx32/Cx26, Cx32*KE/Cx26 and Cx32/Cx26*ES junctions. The expected properties of homotypic and heterotypic junctions in which the component hemichannels gate in response to opposite polarities of V_j are shown in Fig. 2*a*. These expectations were met in each case (Fig. 2*b*) by the display of a marked asymmetry of the G–V_j relation, with conductance decreasing only for the predicted polarity of V_j. The decrease in G is consistent with the closure of either or both hemichannels, as exemplified by the differences in the resultant G–V_j relations according to properties attributable to the component hemichannels. These results also confirm that gating is predominantly an intrinsic hemichannel property.

Localization of the protein domain responsible for the difference in gating polarity of Cx26 and Cx32 was accomplished by

TABLE 1 Summary of gating polarities for point mutations

Mutation	Polarity of closure	Mutation	Polarity of closure
Cx32N2D	+	Cx26D2R	−
Cx32N2E	+	Cx26D2K	−
Cx32N2Q	−	Cx26D2E	+
Cx32N2A	−	Cx32S11D	−
Cx32N2R	−	Cx32Y7D	NE
Cx32N2K	−	Cx32*KE	−
Cx26D2N	−	Cx26*ES	+
Cx26D2Q	−	Cx32*EE	+

Polarity of closure was determined by pairing the mutant connexin with both Cx26 and Cx32 (Fig. 2 legend). NE, no expression.