Our laboratory has found AF to be an excellent inhibitor of trypanosomal AOX (TAO): its K_i against TAO is 2.38 n*M* (Minagawa *et al.*, 1997) compared with 10 μ*M* for the previously discovered TAO inhibitor SHAM (Njogu *et al.*, 1980). However, *in vitro* and *in vivo* experiments have revealed that AF-induced or SHAM-induced killing of trypanosomes is considerably enhanced when they are co-administered with 5 m*M* glycerol (Fairlamb *et al.*, 1977; Van der Meer & Versluijs-Broers, 1979; Minagawa *et al.*, 1997; Yabu *et al.*, 2006). This synergistic effect of glycerol is most likely to be mediated *via* an expected mass-action-induced inhibition of GK by the added glycerol, thereby blocking the anaerobic ATP generation of glycolysis in the parasites. Unfortunately, this nonphysiologically high concentration of glycerol required for co-administration with AF is toxic to the host. Although GK in conjunction with TAO is thus a promising target for chemotherapy, an effective and selective parasite GK inhibitor has not yet become available.

GK is ubiquitous in archaea, bacteria and eukaryotes, where it belongs to the sugar kinase/heat-shock protein 70/actin superfamily (Hurley, 1996). To date, prokaryotic GKs have been the most widely studied. Of the eukaryotes, structural information is only available on *Plasmodium falciparum* GK, but GK is not essential for growth of the asexual blood stages in this organism (Schnick *et al.*, 2009). Kinetic studies also revealed a striking difference between the GKs of trypanosomes and those of other organisms (Kralova *et al.*, 2000). In *T. b. brucei* GK is encoded by five identical tandemly arranged genes (Colasante *et al.*, 2006) and plays an essential role in the survival of the parasite, especially in the absence of oxygen or in the presence of TAO inhibitors (Minagawa *et al.*, 1997), owing to its ability to catalyze the reverse reaction leading to the production of ATP required by the parasites. One may wonder whether the ability of the trypanosomal GK to catalyze the reverse reaction, in contrast to the human enzyme, is purely a consequence of the compartmentalization in glycosomes of the former or whether structure-based catalytic differences also make a contribution. We therefore perceive the parasite GK to be an interesting subject for structural investigation in terms of fundamental enzymology as well as drug-target exploitation. Here, we report the preliminary X-ray diffraction analysis of GK from *T. b. gambiense*, which may lead us to the design of parasite-specific GK inhibitors that spare the host enzyme. Since *T. b. brucei* TAO has also been crystallized recently (Kido *et al.*, 2010), X-ray structure analysis of both enzymes will aid us in the search for a new generation of chemotherapeutic agents against BSFs.

2. Materials and methods

2.1. Cloning and expression of TbgGK

Complementary DNA (cDNA) libraries were prepared from stocks of the bloodstream forms of *T. b. gambiense* (IL2343) and *T. b. rhodesiense* (Tbr; IL1501J21) using Toyobo reverse transcriptase. The cDNAs served as templates for the amplification of their GK-encoding genes (*gk*) by PCR using 5′-CACCATGAAGTACGTCGGATCCATT-3′ and 5′-GTACAACTTTGCCCACTTCGTCCTC-3′ as forward and reverse primers, respectively, with *PfuUltra* II Fusion HS DNA polymerase (Stratagene). The amplicons were gel-purified using the Toyobo gel-purification method. Plasmid constructs were obtained by cloning the blunt-ended gene into the pET151/D-TOPO plasmid vector (Invitrogen) by a ligation-independent cloning procedure. Cloning in this vector leads to the addition of an N-terminal tag containing a His_6 sequence, a V5 epitope and a tobacco etch virus (TEV) protease cleavage site (for removal of the fused 4 kDa tag) to the expressed recombinant protein.

One Shot TOP10 *Escherichia coli* cells were transformed with the Tbg or Tbr *gk*-pET151/D-TOPO plasmid construct by heat shock. Colonies were grown on Luria–Bertani (LB) plates containing 100 μg ml^{-1} carbenicillin and positive clones carrying the inserted gene were confirmed by colony PCR and selected for liquid culturing in LB media for construct amplification. Plasmid extraction from the cultured TOP10 cells was achieved using a Toyobo MagExtractor kit and was subjected to further confirmation by a combination of nested PCR and digestion with *Nco*I. Gene sequencing using the construct and designed sequencing primers was conducted using the dye-terminator method with an ABI Prism310 genetic analyzer (Applied Biosystems). The nucleotide sequence of *gk* revealed that the Tbg and Tbr GKs were exactly identical at the protein level; hence, Tbg *gk* was picked and used in this study. The recombinant plasmid was transformed into the JM109 (DE3 + pRARE2) *E. coli* strain (Novagen) for protein expression. Colonies of the transformants grown on an LB plate containing 100 μg ml^{-1} carbenicillin and 50 μg ml^{-1} chloramphenicol were selected and grown aerobically in LB medium containing the same concentrations of antibiotics.

The expression conditions were optimized for the amount and the activity of GK in the cytosolic fractions using activity measurements and SDS–PAGE by varying the concentration of the expression inducer isopropyl β-D-1-thiogalactopyranoside (IPTG), the temperature and the post-induction time before transformant harvest. The best yield was achieved with 25 μ*M* IPTG, growth at 293 K and post-induction for 8 h.

2.2. Assay of GK activity

The TbgGK activity was assayed using the reverse reaction of TbgGK (glycerol 3-phosphate + ADP → glycerol + ATP). To 1.0 ml of the reaction mixture (1 m*M* EDTA, 5 m*M* $MgSO_4$, 0.5 m*M* $NADP^+$, 50 m*M* glucose, 2 m*M* ADP, 10 m*M* glycerol 3-phosphate and one unit of hexokinase and glucose-6-phosphate dehydrogenase), TbgGK was added at 300 K. Using the ATP produced by TbgGK, hexokinase converts glucose to glucose 6-phosphate and finally glucose-6-phosphate dehydrogenase produces NADPH from glucose 6-phosphate and $NADP^+$. The rate of NADPH accumulation was spectrophotometrically monitored at 340 nm using a Jasco V-660 spectrophotometer.

2.3. Purification of recombinant TbgGK

For large-scale preparation, the transformant was grown at 293 K in 10 l LB medium for 8 h after induction and was harvested by centrifugation at 10 000*g*. The *E. coli* pellet was washed twice in 50 m*M* Tris–HCl buffer pH 7.6 containing 0.1 m*M* phenylmethylsulfonyl fluoride (PMSF) and was resuspended in 300 ml lysis buffer [100 m*M* phosphate buffer pH 6.8, 300 m*M* NaCl, 10 m*M* $MgSO_4$, 0.1 m*M* PMSF, 1 mg ml^{-1} lysozyme and 10%(*v*/*v*) glycerol]. The cell suspension was kept on ice for 30 min, passed twice through a French pressure cell operated at 140 MPa to break the cells and then subjected to centrifugation at 26 000*g* to remove unbroken cells and inclusion bodies. The supernatant was further centrifuged at 146 000*g* to remove residual undissolved material and then applied onto an Ni–NTA Agarose column (Qiagen; 1.5 × 15 cm) pre-equilibrated with 100 m*M* phosphate buffer pH 6.8 containing 20 m*M* imidazole, 300 m*M* NaCl, 10 m*M* $MgSO_4$ and 1%(*v*/*v*) glycerol. After washing the column with 100 ml of the same buffer, rTbgGK was eluted with 500 ml of buffer containing a linear gradient of 20–500 m*M* imidazole. Fractions containing active rTbgGK of higher purity as assessed

by SDS–PAGE (Laemmli, 1970) were pooled, concentrated to approximately 40 mg ml^{-1} using a centrifugal ultrafiltration tube (Amicon Ultra-15, 30 kDa cutoff; Millipore) and stored at 253 K in the presence of 50%(*v*/*v*) glycerol until the next purification step. About 5 mg of the affinity-purified protein was further purified by gel-filtration chromatography using a Superdex 200 (1 × 30 cm) column (GE Healthcare Biosciences) equilibrated with 100 m*M* phosphate buffer pH 6.8 containing 0.3 *M* NaCl and 1%(*v*/*v*) glycerol. Elution was carried out at a flow rate of 0.5 ml min^{-1} on a high-performance liquid-chromatography (HPLC) instrument. Each fraction (0.5 ml) was analyzed by SDS–PAGE and fractions containing highly pure rTbgGK were pooled. After buffer exchange to 10 m*M* MOPS buffer pH 6.8, 10 m*M* $MgSO_4$ and 1%(*v*/*v*) glycerol, the purified rTbgGK was concentrated to about 10 mg ml^{-1} for crystallization experiments. The addition of $MgSO_4$ and glycerol was crucial for preservation of the rTbgGK activity. The concentration of rTbgGK was estimated using the calculated molar extinction coefficient at 280 nm (ε_{280} = 81 080), giving an A_{280} of 1.0 for the pure rTbgGK solution at 0.74 mg ml^{-1}.

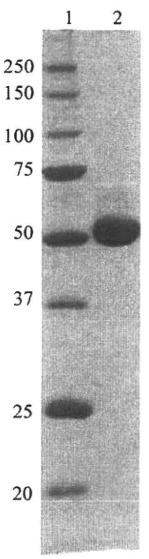

Figure 1
12.5% SDS–PAGE gel stained with Coomassie Brilliant Blue R-250 showing the apparent homogeneity of the purified rTbgGK. Lane 1, molecular-weight markers (kDa); lane 2, rTbgGK purified by Ni–NTA affinity chromatography and Superdex 200 gel filtration.

2.4. Crystallization and X-ray diffraction data collection

Crystallization conditions were initially screened at 277 and 293 K using the sitting-drop vapour-diffusion method in a 96-well Corning CrystalEX microplate with conical flat bottom (Hampton Research). A 0.5 µl droplet containing about 10 mg ml^{-1} rTbgGK dissolved in 10 m*M* MOPS buffer pH 6.8, 10 m*M* $MgSO_4$ and 1%(*v*/*v*) glycerol was mixed with an equal volume of reservoir solution and the droplet was allowed to equilibrate against 100 µl reservoir solution. In the initial screening experiment, commercially available screening kits from Hampton Research (Crystal Screen, Crystal Screen II, Grid Screen Ammonium Sulfate, Grid Screen PEG 6000, Grid Screen MPD and Quick Screen) and from Emerald BioStructures (Wizard Screen I and II) were used as the reservoir solutions. However, most of the conditions gave only heavy protein precipitates and the screening was unsuccessful. Screening was then carried out using a 5 mg ml^{-1} rTbgGK solution and twice-diluted reservoir solutions. Out of 290 conditions screened, tiny crystals and their aggregates appeared at 277 and 293 K from reservoir solutions containing 2.5–5%(*w*/*v*) PEG 6000 in the pH range 6.0–8.0. The conditions were further optimized by varying the buffer pH (5.6–8.4), the molecular weight of the PEG and its concentration [1–10%(*w*/*v*) for PEG 3350 and PEG 6000; 10–30%(*w*/*v*) for PEG 400]. Finally, single crystals suitable for X-ray diffraction experiments were obtained using a reservoir solution consisting of 30%(*w*/*v*) PEG 400 and 100 m*M* HEPES pH 7.0 within 24 h.

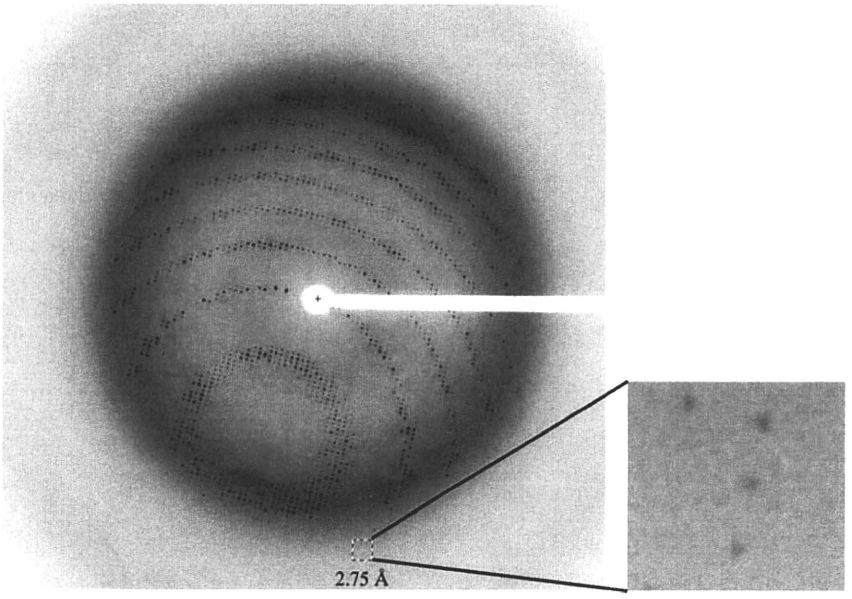

Figure 2
A typical X-ray diffraction pattern of an rTbgCK crystal. The detector edge corresponds to 2.4 Å resolution and an enlarged image of the indicated area around 2.75 Å resolution is shown. The exposure time was 1 s, with an oscillation angle of 1.0°.

X-ray diffraction experiments were performed under cryocooled conditions (100 K) on the BL41XU (λ = 1.000 Å; Rayonix MX225HE CCD detector) and BL44XU (λ = 0.900 Å; Bruker DIP-6040 detector system) beamlines at SPring-8 (Harima, Japan) and the BL17A (λ = 1.000 Å; ADSC Quantum 270 detector) beamline at the Photon Factory (Tsukuba, Japan). A crystal mounted in a nylon loop was transferred to and soaked briefly in reservoir solution containing 40%(*w*/*v*) PEG 400 and then flash-frozen at 100 K in a stream of nitrogen gas. A total of 180 images were recorded with an oscillation angle of 1.0°, an exposure time of 1 s per image and a crystal-to-detector distance of 200 mm. The diffraction data were processed and scaled with the *HKL*-2000 software package (Otwinowski & Minor, 1997).

3. Results and discussion

Gene-sequence analyses for the cDNAs of *T. b. gambiense* and *T. b. rhodesiense gk*s revealed a total of seven point differences when compared with the *gk* sequence from *T. b. brucei* (TREU927; accession No. XM_822408); only one of these differences (T212 in *T. b. brucei* to C in *T. b. gambiense* and *T. b. rhodesiense*) resulted in a change of a single amino acid (Phe71 in TbbGK to Ser71 in TbgGK and TbrGK) in the 512 amino-acid residues of TbgGK. The nucleotide-sequence data for the cDNAs of *T. b. gambiense* and *T. b. rhodesiense gk*s have been deposited in the DDBJ/EMBL/GenBank nucleotide-sequence databases with accession Nos. AB517984 and AB517985, respectively.

His_6-tagged rTbgGK with 545 amino-acid residues (60.4 kDa) was overexpressed and purified to homogeneity by a combination of Ni–NTA affinity chromatography and Superdex 200 gel-filtration chromatography (Fig. 1). About 80 mg purified enzyme with a specific activity of 31.7 μmol min^{-1} mg^{-1} was obtained from a 10 l culture. The rTbgGK protein eluted from the Superdex 200 column with a retention time corresponding to a molecular weight of about 119 kDa, indicating that the enzyme exists as a homodimer in solution.

In a screening of 290 crystallization conditions, crystals of rTbgGK were obtained using PEGs as precipitant. After optimization of the crystallization conditions, the best crystals, which diffracted X-rays to a resolution of 2.75 Å (Fig. 2), were grown at 293 K using a reservoir solution containing 30%(*w*/*v*) PEG 400 and 0.1 *M* HEPES buffer pH 7.0. The crystals attained typical dimensions of about 0.25 × 0.1 × 0.05 mm in 2 d (Fig. 3). Analysis of the symmetry and systematic absences in the recorded diffraction patterns revealed that the crystals of rTbgGK belonged to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters a = 63.84, b = 121.50, c = 154.59 Å. Assuming the presence of two rTbgGK molecules (2 × 60.4 kDa) in the asymmetric unit, the V_M value was calculated to be 2.5 $Å^3$ Da^{-1}, with an estimated solvent content of 50% (Matthews, 1968); these values are within the range commonly observed for protein crystals. A total of 135 987 observed reflections recorded on 180 images were merged to 31 848 unique reflections in the 50.0–2.75 Å resolution range with an R_{merge} of 5.5%. The data-collection and processing statistics are shown in Table 1.

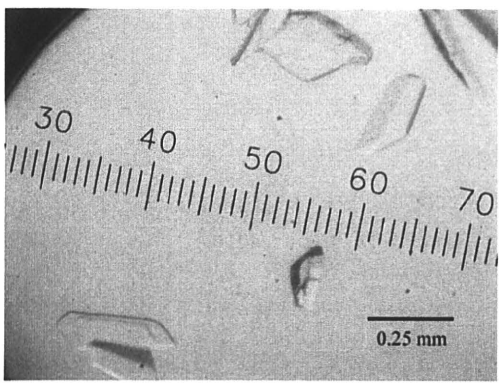

Figure 3
Crystals of rTbgGK obtained by the sitting-drop vapour-diffusion method using PEG 400 as a precipitant.

Table 1
Diffraction data statistics for the crystal of rTbgGK.

Values in parentheses are for the highest resolution shell.

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 63.84, b = 121.50, c = 154.59
V_M† ($Å^3$ Da^{-1})	2.5
Solvent content† (%)	50
X-ray source	BL41XU, SPring8
Wavelength (Å)	1.000
Temperature (K)	100
Resolution (Å)	50–2.75 (2.85–2.75)
Total reflections	135987
Unique reflections	31848
Completeness (%)	97.1 (95.7)
$R_{merge}(I)$‡ (%)	5.5 (46.1)
$\langle I/\sigma(I)\rangle$	18.4 (3.0)

† Assuming the presence of two molecules in the asymmetric unit. ‡ $R_{merge}(I) = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is their average.

An attempt to solve the structure using the molecular-replacement method with the *MOLREP* program (Vagin & Teplyakov, 1997) from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) was carried out using the refined coordinates of GK from *P. falciparum* (PDB code 2w41; 40% amino-acid sequence identity to rTbgGK; Schnick *et al.*, 2009). A promising solution with a homodimeric structure was obtained (correlation coefficient and R factor of 0.406 and 51.6%, respectively). Using the molecular-replacement solution, the structure is being subjected to refinement. In parallel with the refinement, we are now trying to obtain crystals of rTbgGK complexed with ligands, including substrates and substrate analogues. *In silico* screening of potential inhibitors from a compound library of the Chemical Biology Research Initiative at the University of Tokyo is also under way.

It should be noted that the amino-acid sequence of TbgGK was identical to that of TbrGK and showed only one difference from that of TbbGK. Therefore, inhibitors of TbgGK should also be effective against other trypanosome GKs. Since TbgGK provides a greater potential as the primary target of chemotherapy, detailed structures of TbgGK complexed with inhibitors will help structure-based drug design aimed at African trypanosomiasis.

We thank all the staff members of beamlines BL41XU and BL44XU at SPring-8 and BL17A at the Photon Factory for their help with the X-ray diffraction experiments. This work was supported by a grant from the Targeted Proteins Research Program (TPRP) and was supported in part by a Grant-in-Aid for Creative Scientific Research (18GS0314 to KK) from the Japan Society for the Promotion of Science and a Grant-in-Aid for Scientific Research on Priority Areas (18073004 to KK) from the Ministry of Education, Culture, Sports, Science and Technology, Japan. EOB is supported by a Japanese

Government Scholarship from the Ministry of Education, Science, Culture, Sports, Science and Technology.

References

Brun, R., Schumacher, R., Schmid, C., Kunz, C. & Burri, C. (2001). *Trop. Med. Int. Health*, **6**, 906–914.

Chaudhuri, M., Ott, R. D. & Hill, G. C. (2006). *Trends Parasitol.* **22**, 484–491.

Colasante, C., Ellis, M., Ruppert, T. & Voncken, F. (2006). *Proteomics*, **6**, 3275–3293.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Fairlamb, A. H., Opperdoes, F. R. & Borst, P. (1977). *Nature (London)*, **265**, 270–271.

Guerra, D. G., Decottignies, A., Bakker, B. M. & Michels, P. A. (2006). *Mol. Biochem. Parasitol.* **149**, 155–169.

Haanstra, J. R., van Tuijl, A., Kessler, P., Reijnders, W., Michels, P. A., Westerhoff, H. V., Parsons, M. & Bakker, B. M. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 17718–17723.

Hannaert, V., Bringaud, F., Opperdoes, F. R. & Michels, P. A. (2003). *Kinetoplastid Biol. Dis.* **2**, 1–30.

Hurley, J. H. (1996). *Annu. Rev. Biophys. Biomol. Struct.* **25**, 137–162.

Kido, Y., Shiba, T., Inaoka, D. K., Sakamoto, K., Nara, T., Aoki, T., Honma, T., Tanaka, A., Inoue, M., Matsuoka, S., Moore, A., Harada, S. & Kita, K. (2010). *Acta Cryst.* F**66**, 275–278.

Kralova, I., Rigden, D. J., Opperdoes, F. R. & Michels, P. A. (2000). *Eur. J. Biochem.* **267**, 2323–2333.

Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Michels, P. A., Hannaert, V. & Bringaud, F. (2000). *Parasitol. Today*, **16**, 482–489.

Minagawa, N., Yabu, Y., Kita, K., Nagai, K., Ohta, N., Meguro, K., Sakajo, S. & Yoshimoto, A. (1997). *Mol. Biochem. Parasitol.* **84**, 271–280.

Njiokou, F., Laveissière, C., Simo, G., Nkinin, S., Grébaut, P., Cuny, G. & Herder, S. (2006). *Infect. Genet. Evol.* **6**, 147–153.

Njogu, R. M., Whittaker, C. J. & Hill, G. C. (1980). *Mol. Biochem. Parasitol.* **1**, 13–29.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Schnick, C., Polley, S. D., Fivelman, Q. L., Ranford-Cartwright, L. C., Wilkinson, S. R., Brannigan, J. A., Wilkinson, A. J. & Baker, D. A. (2009). *Mol. Microbiol.* **71**, 533–545.

Singha, U. K., Peprah, E., Williams, S., Walker, R., Saha, L. & Chaudhuri, M. (2008). *Mol. Biochem. Parasitol.* **159**, 30–43.

Stevens, J. R. & Brisse, S. (2004). *The Trypanosomiases*, edited by I. Maudlin, P. Holmes & M. Miles, pp. 1–23. Wallingford: CAB International.

Van Der Meer, C. & Versluijs-Broers, J. A. (1979). *Exp. Parasitol.* **48**, 126–134.

Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.

World Health Organization (2006). *African trypanosomiasis*. http://www.who.int/mediacentre/factsheets/fs259/en/.

Yabu, Y., Suzuki, T., Nihei, C., Minagawa, N., Hosokawa, T., Nagai, K., Kita, K. & Ohta, N. (2006). *Parasitol. Int.* **55**, 39–43.

6186–6194 *Nucleic Acids Research, 2010, Vol. 38, No. 18*
doi:10.1093/nar/gkq449

Published online 27 May 2010

Extensive frameshift at all AGG and CCC codons in the mitochondrial cytochrome *c* oxidase subunit 1 gene of *Perkinsus marinus* (Alveolata; Dinoflagellata)

Isao Masuda, Motomichi Matsuzaki* and Kiyoshi Kita

Department of Biomedical Chemistry, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Received December 14, 2009; Revised and Accepted May 7, 2010

ABSTRACT

Diverse mitochondrial (mt) genetic systems have evolved independently of the more uniform nuclear system and often employ modified genetic codes. The organization and genetic system of dinoflagellate mt genomes are particularly unusual and remain an evolutionary enigma. We determined the sequence of full-length cytochrome *c* oxidase subunit 1 (*cox1*) mRNA of the earliest diverging dinoflagellate *Perkinsus* and show that this gene resides in the mt genome. Apparently, this mRNA is not translated in a single reading frame with standard codon usage. Our examination of the nucleotide sequence and three-frame translation of the mRNA suggest that the reading frame must be shifted 10 times, at every AGG and CCC codon, to yield a consensus COX1 protein. We suggest two possible mechanisms for these translational frameshifts: a ribosomal frameshift in which stalled ribosomes skip the first bases of these codons or specialized tRNAs recognizing non-triplet codons, AGGY and C CCCU. Regardless of the mechanism, active and efficient machinery would be required to tolerate the frameshifts predicted in *Perkinsus* mitochondria. To our knowledge, this is the first evidence of translational frameshifts in protist mitochondria and, by far, is the most extensive case in mitochondria.

INTRODUCTION

Mitochondria, the energy-producing organelles in eukaryotic cells, possess their own genomes. Mitochondrial (mt) genomes have been reduced relative to those of their bacterial ancestors by a series of evolutionary events, including massive gene transfers to the nuclear genome and gene loss (1). Most of the mt genomes sequenced to date are single, circular, double-stranded DNA molecules that typically encode dozens of genes for respiratory electron-transport chain proteins, ATP synthase proteins, ribosomal RNA (rRNA) and transfer RNA (tRNA). However, due to independent evolutionary events across eukaryotic taxa, mt genomes are very diverse with regard to physical structure, genome size and gene content. For example, mt genomes of land plants are highly expanded (up to 2.4 Mbp in muskmelon) (2), and the smallest mt genome reported is a 6-kb long linear molecule in apicomplexan parasites (3). An mt genome with unusual organization—several hundred linear DNA molecules coding one or a few genes—is found in the icthyosporean *Amoebidium* (4). In Euglenozoan flagellate *Diplonema*, one mt gene is separated into multiple fragments, each encoded on a different mini circlular molecule (5,6).

mt gene expression is distinct from that in the nucleus, and mitochondria are notable for having alternative genetic codes. One well-known code alteration is codon reassignment in which codons are not decoded as designated in the standard codon table. For example, the UGA codon in mitochondria of many eukaryotes (other than land plants) codes for tryptophan rather than a stop (7); AGR codons (R = A or G) in inchordata mitochondria code for glycine rather than arginine (8); and CUN codons (N = A, U, G or C) in yeast mitochondria code for threonine instead of leucine (9). Some codon reassignments, even those that result in the same coding change, are suggested to have evolved independently in separate taxa; one example is the reassignment of UAG codon to leucine in chlorophycean and in fungal mitochondria (7,10–12).

Dinoflagellate mt genomes are known for their remarkable organization and genetic systems. Although the overall mt genome structure is not yet determined, these are suggested to be composed of a number of heterogeneous DNA molecules that resulted from rampant homologous recombination (13,14). The entire mt genome size

*To whom correspondence should be addressed. Tel: +81 3 5841 8202; Fax: +81 3 5841 3444; Email: mzaki@m.u-tokyo.ac.jp

is estimated to be at least 30 kb but is probably much larger (15). The genome encodes a strictly limited set of genes: three protein-coding genes, *cox1*, *cox3* and *cob*, and several fragmented rRNA genes. Curiously, the three protein-coding genes lack canonical start (AUG) and stop (UAA, UAG and UGA) codons in the 5′ and 3′ terminal regions, respectively (13–17). Transfer RNA genes have not been detected in any of the dinoflagellate mt genomes, and most of these dinoflagellate mt genomes comprise non-coding and pseudogene sequences (13,17,18). Recent studies on two basally-branching dinoflagellates have further highlighted the complexity of these mt genomes; the mt genomes of both *Oxyrrhis marina* and *Amphidinium carterae* are comprised of a number of DNA molecules bearing multiple copies of the three protein-coding genes with different intergenic contexts to one another (15,19). Particularly in the latter species, long intergenic sequences containing extensive inverted repeats are predicted to form many stem-loop structures (15).

Some of the unusual characteristics of dinoflagellate mt genomes are shared with those of parasitic apicomplexans, albeit with significant differences in mt genome organization (14,20). Apicomplexa is the sister lineage to dinoflagellates and is composed of a variety of protozoan parasites, including the malaria parasites *Plasmodium* spp. Generally, the mt genomes of apicomplexans are linear and ~6 kb long, the smallest of the known mt genomes (3). The genomes are tightly packed and have the same three protein-coding genes as dinoflagellates, as well as fragmented rRNA genes; the protein-coding genes also lack canonical start and stop codons (21,22). Although the mt genomes of these two sister lineages, which share unusual features, have not been fully characterized for the mechanisms of gene expression, the shared gene content suggests that the drastic gene reduction in genome content occurred before the divergence of these lineages. In contrast, the significant difference in dinoflagellate and apicomplexa mt genome structures indicates that drastic mt genome reorganization events occurred after the two lineages split and independently diverged from their common ancestor (14).

To further understand the uniformity and diversity of mt genomes of dinoflagellates and apicomplexans, we are characterizing the mt genome of *Perkinsus* spp., which are well-known, aquatic unicellular parasites of various commercially important bivalve mollusks. In particular, the most studied species, *P. marinus*, parasitizes the eastern oyster, *Crassostrea virginica*, causing mass mortality in the host species (23). Genus *Perkinsus* is assumed to be the most basal of the dinoflagellate lineages discovered to date, branching just after the split between dinoflagellates and apicomplexans (24). Molecular studies on this organism are currently limited, but due to its industrial and phylogenetical significance, the genome project for *P. marinus* is being undertaken by scientists at the J. Craig Venter Institute (JCVI; formerly, The Institute for Genomic Research, TIGR) and scientists at the Department of Microbiology and Immunology, University of Maryland School of Medicine/Institute of Marine and Environmental Technology (IMET; formerly, at the Center of Marine Biotechnology, UMBI) (25). Although we have observed DNA in the mitochondria of *P. marinus* using DNA- and mitochondria-specific dyes (26), critical molecular data and annotated mt gene sequences are not been available from either the National Center for Biotechnology Information (NCBI) database or the previously available TIGR draft genome database (note that the *P. marinus* genomic data set is currently being curated at JCVI).

In this article, we report the first cloning and characterization of a *Perkinsus* mt gene. We used PCR with degenerate primers, ultracentrifugal isolation both of mitochondria and mt genome, and pulsed-field gel electrophoresis. Although these initial attempts detected neither the partial nor whole mt genome, we identified short fragments of mt gene remnants inserted into the nuclear genome of *P. marinus* in the previous TIGR database. We obtained the full-length mRNA sequence for *cox1*, which codes for mt cytochrome *c* oxidase subunit 1, by PCR and RACE. The primary sequence of this mRNA shared several features with orthologs from related species, and together with Southern hybridization data, the codon usage suggested that this gene resides in the *P. marinus* mt genome. Unexpectedly, multiple sequence alignments and a three-frame translation indicated that the translation of this mRNA employs a modified decoding system. We discussed the primary sequence features of this mRNA and further described the possibility of a unique, modified translational decoding system in *Perkinsus* mitochondria.

MATERIALS AND METHODS

Strains and culture conditions

The *P. marinus* strain CRTW-3HE was purchased from the American Type Culture Collection (ATCC, no. 50439) and maintained at 26°C in ATCC medium 1886. Discontinued products were substituted as follows: Lipid Mixture (1000×; L5146; Sigma) replaced Lipid Concentrate (100×; 21900-014; Gibco) and Instant Ocean Sea Salt (Aquarium Systems) replaced artificial seawater (S1649; Sigma). Strains of *P. honshuensis* and *P. olseni* were provided by Dr Tomoyoshi Yoshinaga (The University of Tokyo) and maintained in the same manner.

Nucleic acid preparation

Perkinsus cells were collected by centrifugation at 800*g* for 5 min and re-suspended in extraction buffer [100 mM Tris, 100 mM boric acid and 50 mM ethylenediamine tetraacetic acid (EDTA), pH 8.0]. Cell suspensions were treated with sodium dodecyl sulfate at 60°C for 30 min. Total DNA was purified using standard phenol–chloroform extraction and ethanol precipitation methods. Total RNA was prepared using TRIzol Reagent (Invitrogen) according to the manufacturer's protocols, followed by the poly(A)$^+$-RNA enrichment with PolyATract mRNA Isolation System III (Promega). Complementary DNA (cDNA) was synthesized with SMART RACE cDNA amplification kit (Clontech) following manufacturer's instruction.

PCR, RACE, cloning and sequencing

PCR was performed using Takara Ex *Taq* (Takara Bio) or *PfuUltra* II HS DNA polymerase (Stratagene). We prepared reaction mixtures according to the manufacturers' instructions. Amplification was performed as follows: denaturation at 94°C for 4 min followed by 35 cycles of 94°C for 30 s, a primer annealing gradient from 40 to 50°C for 30 s, and extension at 72°C for 1 min, followed by a final extension at 72°C for 7 min. Primer set Pmcox1F1 and Pmcox1R3, based on the *cox1*-like sequences of nuclear DNA of mt origin (Numt) found in database ('Results' section), was used to amplify the partial sequence of the *Perkinsus* mt *cox1* (Supplementary Figure S1). Pmcox1R3 was then used in combination with a degerenate primer cox1-3f, which was designed based on *cox1* orthologs from related species (Supplementary Figure S1), to additionally sequence the upstream region of *Perkinsus* mt *cox1*. After determining the full *P. marinus cox1* (*Pmcox1*) mRNA sequence, we designed the primer set Pmcox1fullF and Pmcox1fullR for use in PCR of the nearly full-length *Pmcox1* both from genomic DNA (gDNA) and cDNA. Primer sequences are listed in Supplementary Table S1.

We performed RACE experiments using Takara Ex *Taq* with *P. marinus* cDNA as the template. Reaction mixtures were prepared according to the instructions of the cDNA synthesis kit manufacturer. Reaction conditions were 35 cycles of 94°C for 30 s, 48°C for 30 s, and 72°C for 3 min, and a final extension at 72°C for 7 min. Primers for 5′ and 3′ RACE were Pmcox1-5RACE and Pmcox1-3RACE, respectively (Supplementary Table S1).

PCR and RACE products were separated by electrophoresis on 1.2% agarose gel containing 1× Tris–Borate EDTA (TBE) buffer and target products were extracted with the MagExtractor PCR & Gel Clean up kit (Toyobo). The gel-purified products were then cloned using the TOPO TA cloning kit for Sequencing (Invitrogen). The recombinant plasmids containing PCR or RACE products were extracted from transformed *E. coli* (strain DH5α) using MagExtractor Plasmid (Toyobo). Both strands of cloned products were sequenced with the DYEnamic ET Terminator Cycle Sequencing kit (GE Healthcare) on an ABI310 automatic sequencer (Applied Bioscience). Sequences were determined from more than three clones, unless otherwise stated. For nearly full-length gene fragments, direct sequencing was performed on four independently obtained PCR products. Consensus sequences were determined from the alignments of multiple sequences. The assembled full-length mRNA sequence was deposited to the DNA Data Bank of Japan (accession no. AB513789).

Sequence analysis

Sequences were aligned with Clustal X 1.83 (27) and amino acid sequences were predicted using the ExPASy translate tool (http://www.expasy.org/tools/dna.html). Codon usage in several *P. marinus* genes was calculated using the Countcodon program (Kazusa DNA Res. Inst., http://www.kazusa.or.jp/codon/countcodon.html).

Accession numbers for *P. marinus* nuclear genes are as follows: *ispC* (AB284362), *sod1* (AY095212), *sod2* (AY095213) and *act1* (AY436364).

Southern hybridization

DNA fragments for use as probes were amplified by PCR using the following primer sets (for primer sequences, see Supplementary Table S1): Pmcox1pF and Pmcox1pR for *Pmcox1*, nucLSU-7f and nucLSU-7r for large subunit ribosomal DNA (LSU rDNA), PmNumt1F and PmNumt1R for *cox1*-like Numt1 and its flanking regions, and PmNumt2F and PmNumt2R for Numt2 and its flanking regions. The amplified fragments were cloned as described earlier. The extracted plasmids were digested with EcoRI overnight except for two Numt-plasmids, which were digested with both NotI and PstI, and the fragments were purified, labeled and hybridized to *P. marinus* genomic DNA with or without restriction enzyme digestion. The probes were used for detection with the AlkPhos Direct Labelling and Detection System with CDP-Star (GE Healthcare) as follows. First, 1 μg of *P. marinus* genomic DNA was digested with each restriction enzyme overnight at 37°C. Digested and uncut DNA was subjected to electrophoresis on a 0.3% agarose gel and transferred onto Hybond N^+ nylon membrane (GE Healthcare) overnight. Purified probe (100 ng) was labeled with alkaline phosphatase and hybridized to the membrane-linked genomic DNA overnight at 42°C. The membrane was washed and incubated with the substrate CDP-Star, and the chemiluminescence signal was detected using LAS-4000 (Fujifilm).

RESULTS

Primary sequence of *P. marinus cox1*

Preliminary searches for mt genome fragments of *P. marinus* in the NCBI databases of May 2008 and the *P. marinus* draft genome database at TIGR using mt gene sequences of dinoflagellates and apicomplexans as queries did not produce any sequences that were supported with statistical significance ($E < 0.01$). The identified sequences were checked carefully by eye while referring to the amino acid alignment of COX1 from related species to identify highly conserved amino acid residues in the partial sequences, and two contigs were found to harbor *cox1*-like fragments, albeit these were only partial and tiny fragments (Supplementary Figure S1). Contig no. 22713 (available as part of AAXJ01000589 in Genbank/EMBL/DDBJ) contained a fragment with 75.0% AT, that showed 68% predicted-amino acid identity (17/25 residues) with *O. marina* COX1 (ABK57983) and was found to include functionally essential amino acid residues His276 and Glu278 [amino acid numbers according to Iwata *et al.* (28)]. Another fragment in contig no. 22822 (available as part of AAXJ01000147) had 70.7% AT and showed 52% predicted-amino acid identity (20/38 residues) with *O. marina* COX1 and conserved His325 and His326 (Supplementary Figure S1A). We realized that the base composition of these *cox1*-like fragments differed from those of the flanking regions

(<55% AT). The flanking regions did not show sequence similarity to *cox1* and were discovered to harbor nuclear genes like RNA helicase gene and clathrin-associated protein gene, the former of which contained the *cox1*-like fragment in one of its intronic regions (Supplementary Figure S1B). These observations imply that these *cox1*-like, AT-rich fragments are nuclear DNA of mt origin (Numts), which are DNA fragments that had been transferred from mt genomes into the nucleus and, in many cases, have become transcriptionally inactive.

Because the *cox1*-like Numts and the true mt *cox1* are likely to have similar sequences, we used two primer sets for PCR: (i) Pmcox1F1 and Pmcox1R3, both of which were derived from the Numt sequences and (ii) Pmcox1R3 and a degenerate primer cox1-3f, which was designed based on *cox1* sequences of closely related species. In each case, there was a distinct single DNA amplification from total *P. marinus* DNA template. Sequencing of these PCR products confirmed the lengths at 167 and 434 bp, respectively, with the former being completely included in the latter. To obtain the full-length sequence of this gene, we performed 5′ and 3′ RACE using internal primers Pmcox1-5RACE and Pmcox1-3RACE, respectively, with *P. marinus* cDNA as the template. After cloning and sequencing five clones for each of the RACE products (~700 bp each), we amplified the nearly full-length sequence (~1400 bp) of both gDNA and cDNA using specific primer sets (Pmcox1fullF and Pmcox1fullR) followed by direct sequencing of multiple independent PCR products. The sequences of the RACE products and the nearly full-length sequence were manually assembled to determine the full-length mRNA sequence (1434 bp) of this gene, which was confirmed to contain sequences identical to the PCR and RACE fragments obtained above. Conversely, this mRNA contained regions which are similar to, but not identical to Numt sequences, and their flanking regions were completely different from each other (Supplementary Figure S1C). There were no substitutions, insertions and deletions between sequences from gDNA and cDNA, suggesting that RNA editing does not occur in this gene. The overall AT content of this gene was 80.9%. As a whole, this gene was similar to *cox1* of dinoflagellates and apicomplexans with an $E < 10^{-70}$; hereafter, we refer to this sequence as *Pmcox1* mRNA.

Genomic localization of *Pmcox1*

To determine the localization of *Pmcox1* in *P. marinus* genomes (nucleus or organelles), we conducted Southern hybridization using total DNA because it was difficult to isolate pure mt DNA or intact mitochondria from *P. marinus*. *Pmcox1* signals constituted a smear in the low molecular-weight region (<10 kb) of uncut genomic DNA, which is far lower than the expected position for chromosomal DNA (Figure 1). Similarly, *Pmcox1* signals formed a smear for the digestion of total DNA with BamHI, EcoRI or HindIII. A distinct signal was only observed (1–2 kb region) for the digestion of total DNA with AccI. Given the high AT content of *Pmcox1*, it is

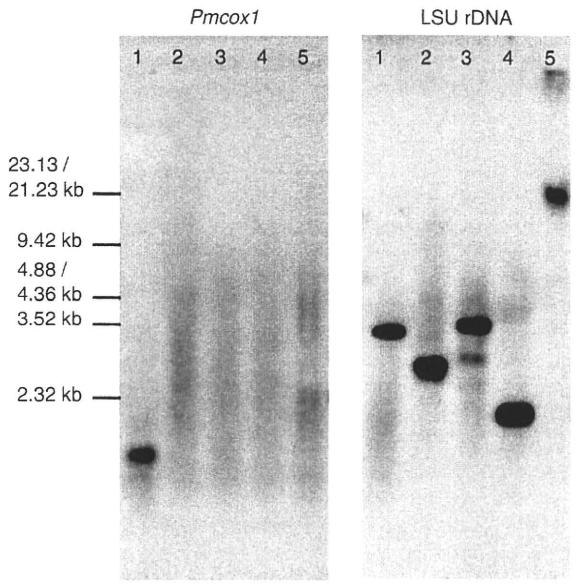

Figure 1. Southern hybridization with *Pmcox1* and a nuclear control probe. Southern hybridization using *Pmcox1* (left) and nuclear LSU rDNA (right) probes. Lanes 1–4, *P. marinus* genomic DNA digested with AccI (1), BamHI (2), EcoRI (3) and HindIII (4); lane 5, uncut genomic DNA.

natural that AccI was the only restriction enzyme tried here which cut *P. marinus* mt genome sequences around *Pmcox1*.

In sharp contrast to the *Pmcox1* probe, the probe for the nuclear LSU rDNA hybridized to the stacked, high molecular-weight, chromosomal DNA in the uncut DNA sample (Figure 1). Moreover, one or two distinct LSU rDNA band(s) were detected in genomic DNA digested with AccI, BamHI, EcoRI or HindIII. The LSU rDNA signals indicate the high quality of genomic DNA and that the restriction digests were complete. The smear signals from the *Pmcox1* probe suggest that *Pmcox1* resides on small (<10 kb) heterogeneous non-chromosomal DNA. Like the LSU rDNA probe, probes for Numts and its flanking regions hybridized to the undigested chromosomal DNA without a smear signal, indicating that they reside on chromosomal DNA (Supplementary Figure S2).

Prediction of amino acid sequence

The amino acid sequence predicted to be encoded by the primary *Pmcox1* mRNA sequence unexpectedly could not be translated in its entirety using the standard codon table in a single reading frame; several stop codons appeared in all three frames (Figure 2A). Performing Blastx-based search using the entire *Pmcox1* mRNA sequence as a query identified several partial COX1-like amino acid sequences that appeared separately in all three reading frames (gray boxes in Figure 2A). In total, we found eleven COX1-like 'coding-blocks' (gray boxes numbered I–XI in Figure 2B) that cover almost the entire sequence of *Pmcox1*, though discontinuously.

To understand the discontinuity in the COX1-like amino acid sequences, we aligned the *Pmcox1* mRNA

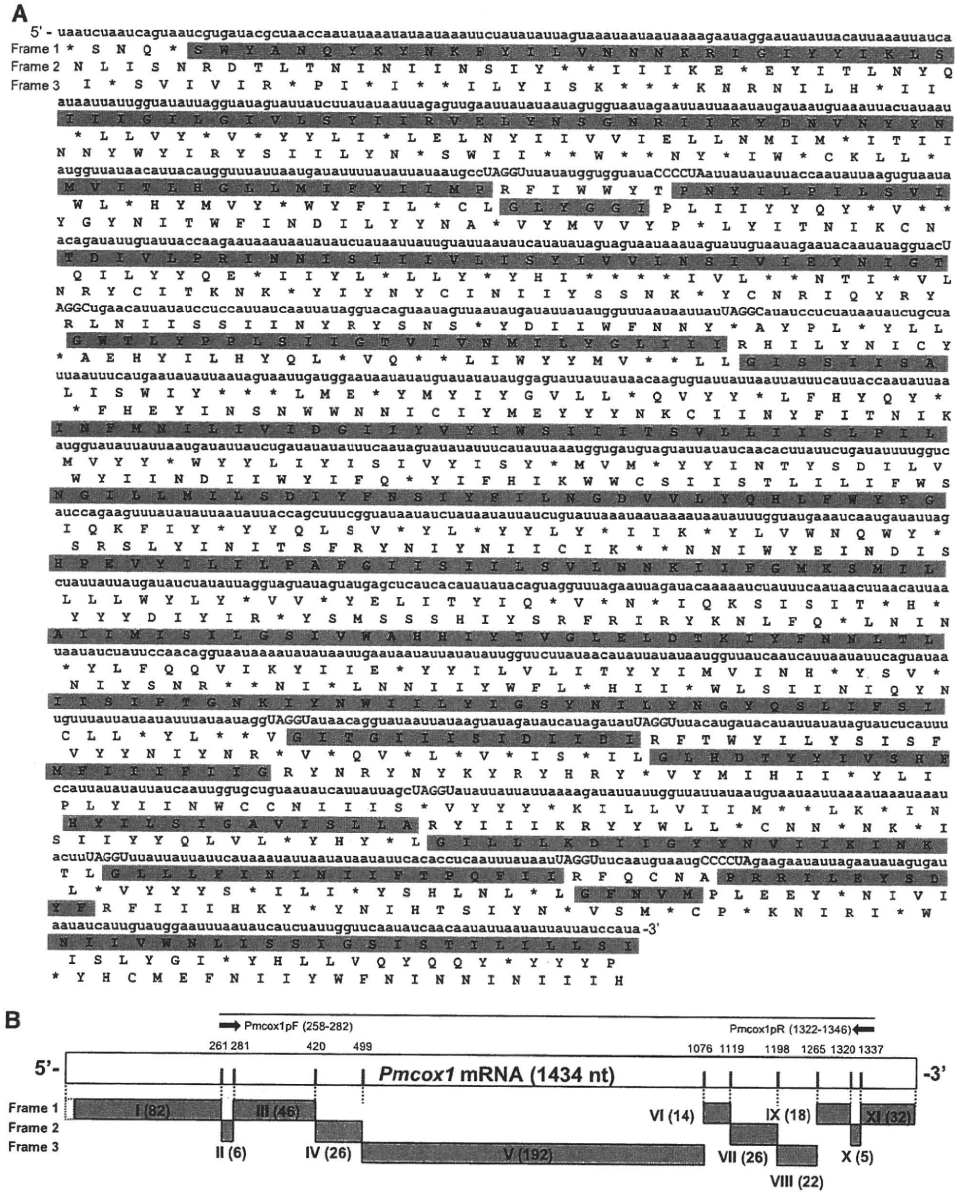

Figure 2. (**A**) Nucleotide sequence of *Pmcox1* and its three frame translation. Gray boxes denote COX1-like amino acid sequences. Nucleotides in uppercase represent the sequence motifs found in transition areas of the reading frame where COX1-like amino acid sequences are deduced. AGG and CCC codons in the motifs are highlighted in red and blue, respectively. (**B**) Schematic model of *Pmcox1* mRNA and the COX1-like blocks. COX1-like amino acid sequences distributed across three reading frames are represented by gray boxes labeled with Roman numerals. Numbers in parenthesis adjacent to the Roman numerals indicate the number of the amino acid residues. Red bars at the end of COX1-like blocks indicate UAGGY motifs containing in-frame AGG, and blue bars indicate CCCCUA motifs including in-frame CCC. Numbers above the open box indicate the positions of the first nucleotide in each motif. The uppermost bar indicates the probe region used for Southern hybridization with arrows indicating the primer locations.

sequence with *cox1* sequences of related species (Supplementary Figure S3). Among the four *cox1* sequences, *Pmcox1* was the most divergent and contained the largest number of insertions and deletions (indels). Curiously, there were 10 one- or two-base indels specifically in the *Pmcox1* mRNA that occurred in the context of UAGGY (8 of 10) or CCCCUA (2 of 10) motifs (shown on a black background in Supplementary Figure S3). Most intriguingly, these 10 regions appear to coincide with the transitions between coding-blocks, and AGG and CCC appear in-frame preceded by the predicted COX1-coding blocks (Figure 2).

Based on these observations, we postulated that modified decoding, which could shift the reading frame, occurs in the translation of *Pmcox1* mRNA. In our hypothesis, specifically, when an in-frame AGG or CCC

appears, the reading frame should be shifted forward by one or two bases, respectively. Accordingly, we prepared a putative PmCOX1 amino acid sequence in the following manner. We eliminated the A residues of the UAGGY motifs and made a +1 frameshift, making GGY instead of AGG in-frame. We also deleted the first two C residues of CCCCUA motifs and made a +2 frameshift, making CCU instead of CCC in-frame. This model accounts for all the *Perkinsus*-specific one- and two-base indels and connects the 11 'blocks' into one consecutive coding sequence. The alignment of our putative PmCOX1 sequence with counterparts from related organisms shows the conservation of functionally important amino acid residues (Figure 3, black boxes). This sequence also conserves the glycine and proline residues, which are most common in the proximity of the UAGGY and CCCCUA motifs. The potential mechanisms for these frameshifts will be further discussed later.

Based on this amino acid sequence, we identified the following characteristics of codon usage in *Pmcox1*. Around the 5′ terminal regions, no AUG codon that is likely to act as start codon was identified. Canonical stop (UAA, UAG and UGA) codons were not observed in 3′ terminal regions, as is often the case with mt genes of dinoflagellates and apicomplexans. Comparison of the COX1 amino acid alignment and nucleotide sequence also showed well-conserved tryptophan residues among related species that appeared to be coded by UGA codons in *Pmcox1* (Figure 2A and open boxes in Figure 3). On the whole, *Pmcox1* utilizes only 35 different codons whereas nuclear genes use 53–60 (Supplementary Table S2).

DISCUSSION

Pmcox1 is located in the mt genome

Using the newly determined sequence of *Pmcox1* mRNA and nearly the full-length of its genomic counterpart, we find evidence to suggest that this gene is located in the mt genome. First, Southern hybridization of total DNA from *P. marinus* shows the localization of the *Pmcox1* gene that is distinct from that of the nuclear LSU rDNA. Signal from a *Pmcox1* probe formed a smear in the relatively low molecular-weight regions of uncut total DNA, while LSU rDNA probe hybridized to stacked, uncut DNA with high molecular weight, i.e. chromosomal DNA (Figure 1). These results indicate that *Pmcox1* resides on the relatively small DNA molecules distinct from chromosomal, nuclear DNA. The present hybridization data (Figure 1) is congruent with previously reported results on other dinoflagellates (16,29,30), suggesting that *Pmcox1* is encoded on multiple heterogeneous DNA molecules, which is similar to the structure found for other dinoflagellate mt genomes.

Second, canonical start and stop codons are not found in the terminal regions of *Pmcox1* (Figure 2A). As the mt

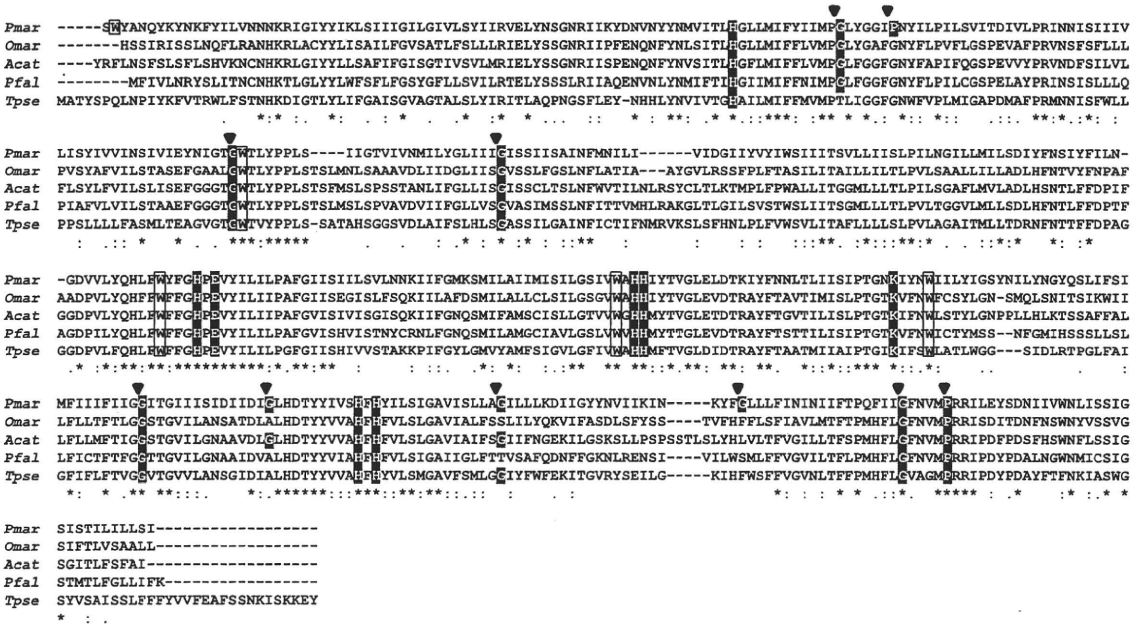

Figure 3. Alignment of multiple COX1 amino acid sequences of *P. marinus* and related protists. The sequence for *Perkinsus* was predicted based on the frameshift model. Asterisks, colons and dots indicate identical residues, conserved and semi-conserved substitutions, respectively. Residues highlighted in black are conserved amino acid essential for cytochrome *c* oxidase function: His94 (ligand for heme a), His276 (ligand for CuB), Glu278 (D-channel), His 325, His326 (ligand for CuB), Lys354 (for K-channel), His411 (ligand for heme a3) and His413 [ligand for heme a; numbers are according to *Paracoccus denitrificans* homolog. (28)]. The red and blue arrowheads indicate the motifs (UAGGY and CCCCUA, respectively) where the reading frame is hypothetically shifted. Note that the glycine and proline residues at the frameshift sites are often highly conserved (highlighted in red and blue, respectively). The open boxes indicate tryptophans coded by UGA codons in *Pmcox1* and those conserved at the homologous positions in related species. *cox1* sequences were obtained from the NCBI database for *Oxyrrhis marina* (*Omar*, EF680822), *Alexandrium catenella* (*Acat*, AB374235), *Plasmodium falciparum* (*Pfal*, AY282930) and *Thalassiosira pseudonana* (*Tpse*, DQ186202). Note that the ciliate genes are not included because they are highly divergent and the gene length also differs greatly from those of related alveolates.

genes of dinoflagellates and apicomplexans do not possess AUG start codon and stop codons, these are assumed to utilize alternative start and stop mechanisms (16,17,22,29,30). All of the *Perkinsus* nuclear genes examined here had AUG start and stop codons in the expected positions based on comparisons to orthologs from related species. These observations support that *Pmcox1* resides in the mt genome.

Lastly, overall codon usage showed significant differences between *Pmcox1* and *Perkinsus* nuclear genes (Supplementary Table S2). Moreover, several UGA codons, which typically function as stop codons in nuclear genes but often code for tryptophan in mt genes, were present in the *Pmcox1* mRNA and appeared to code for tryptophan (Figures 2A and 3). While we have no direct evidence that *Pmcox1* is located in the mt genome, these multiple lines of evidence strongly support the conclusion that this gene is not located in the nuclear but in the mt genome.

mt gene translation of *Perkinsus* involves multiple frameshifts

Surprisingly, the *Pmcox1* mRNA is apparently not translated in a single reading frame. Because we detected the cyanide-sensitive enzyme activity of cytochrome *c* oxidase according to the method described previously (31), functional COX1 protein most likely exists in *Perkinsus* mitochondria. Furthermore, we obtained partial sequences of *Pmcox1* orthologs from two other *Perkinsus* species, *P. olseni* and *P. honshuensis* (Supplementary Figure S4). Their nucleotide sequence identity to *Pmcox1* was >96%, and the UAGGY motif was conserved. There were no gaps in the alignment and all substitutions were synonymous, indicating the selective pressure to conserve the amino acid sequence in *Pmcox1* and these orthologs. Taken together with there being no *cox1*-like sequence other than *Pmcox1*, these results further emphasize that *Pmcox1* is functional and is translated with the aid of an unusual mechanism that requires multiple frameshifts (Figures 2 and 3).

At present, we are unable to show direct evidence that translation of *Pmcox1* mRNA requires frameshifts because we have not directly sequenced the PmCOX1 protein. However, the predicted PmCOX1 amino acid sequence reinforces the validity of our frameshift model. As a major functional component of mt cytochrome *c* oxidase, COX1 reduces molecular oxygen to water using electrons from cytochrome *c* and transports protons from the mt matrix to the intermembrane space. The amino acid sequence of PmCOX1 predicted by the frameshift model retains the conserved residues that are essential for these reactions (see Figure 3 and its legend) (28). The reading frame is possibly shifted back by one base (−1 frameshift) at the CCCCUA motif, but this is less likely because it would require the insertion of one extra amino acid residue into the alignment.

Moreover, this frameshift motif may be conserved in another mt gene. We identified a *cob*-like fragment from *P. marinus* whole-genome shotgun assemblies (AAXJ01022806) in a Blast-based search using dinoflagellate mt gene sequences. This fragment included four conserved UAGGY motifs and one GAGGY motif where the reading frame appeared to be shifted forward by one base to connect discontinuous COB-like amino acid sequences to form a plausible COB protein (Supplementary Figure S5). In contrast, the deduced amino acid sequences for *Perkinsus* nuclear genes shown in Supplementary Table S1 did not include such translational frameshifts. These observations strongly indicate that an unconventional event occurred during translation, specifically in mitochondria of *P. marinus*, and also of other *Perkinsus* species. Our data are the first evidence of a frameshift-dependent translation system in protist mitochondria.

Possible mechanisms suggested for frameshift in *Perkinsus* mt genes

If *Pmcox1* mRNA is read in all three frames to generate PmCOX1, an unconventional mechanism must exist in the *Perkinsus* mt translation system to shift the reading frame systematically. One possible mechanism is a ribosomal frameshift, a phenomenon observed in a wide range of organisms which results in a shift forward or backward in the reading frame during translation (32). In the case of +1 ribosomal frameshift, a rarely used codon or a stop codon in the ribosome A site is suggested to induce the ribosome to stall and allow the reading frame to be subsequently shifted forward by skipping one base (33,34). Ribosomal frameshifts have also been found in mt genes from various animals, and a +1 frameshift is suggested at specific codons (35–40). Based on previous studies, we hypothesized that ribosomes in *Perkinsus* mitochondria skip the A residue in the first position of the in-frame AGG in the shared UAGGY motif and the first two C residues in the CCCCUA motif by shifting forward by one base at in-frame CCC (Figure 4A). These two types of frameshifts at the rarely used AGG and CCC codons change the reading frame and allow the discontinuous COX1-like amino acid sequences to be joined, which produces the preferred amino acid residues at the frameshift sites (Figure 3).

Alternatively, specialized tRNAs that recognize non-triplet codons may be utilized at frameshift sites during translation. Naturally occurring deviant tRNAs recognize four-base codons and act as suppressors of non-sense mutations, and artificial tRNAs bearing modified loops can recognize quadruplet and even quintuplet codons (41–44). In the case of *Pmcox1*, specialized tRNAs may recognize AGGY (for glycine) and CCCCU (for proline) to enable the proposed frameshifts (Figure 4B). With these tRNAs, the reading frame would be shifted by one and two base(s), respectively, and one contiguous COX1 protein would be translated. Specialized tRNAs with altered decoding capacity may be used in *Perkinsus* mitochondria, although such mt tRNAs have not yet been identified from any organism.

Regardless of the mechanism, it should be noted that the efficiency of translational frameshift depends on the nucleotide sequence and the abundance of tRNAs, but 100% efficiency has never been observed (45). Lower

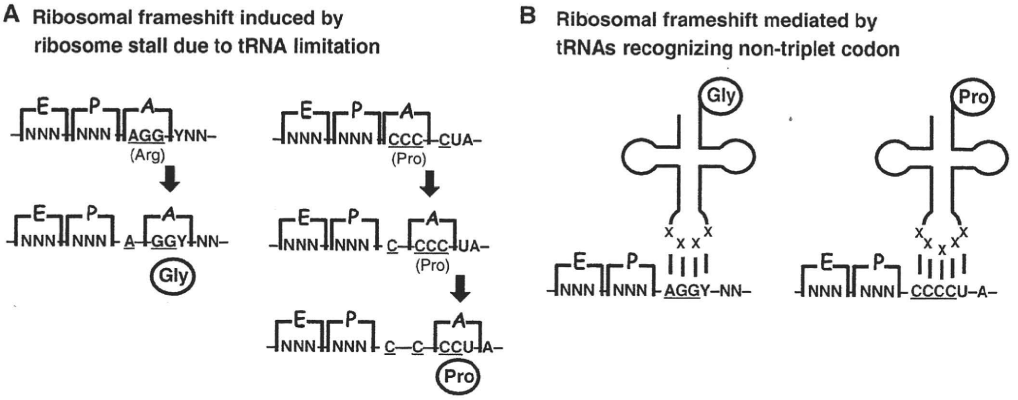

Figure 4. Possible mechanisms of frameshift-dependent translation at AGG and CCC codons of *Perkinsus* mt genes. (**A**) A ribosome stalled by tRNA limitation induces a ribosomal frameshift. When an in-frame AGG or CCC moves into the ribosome A site during translation, the ribosome stalls due to the limitation of the corresponding tRNA molecules, and the reading frame is subsequently shifted to the +1 frame. Translation then restarts in the new frame. In the case of CCCCUA, two consecutive frameshifts occur. (**B**) Specialized tRNAs recognize non-triplet codons, AGGY and CCCCU, and the reading frame shifts forward by one (AGGY) or two (CCCCU) bases.

frameshift efficiencies are not lethal to organisms known to have frameshift-dependent genes because there is only one (most cases) or at most two [for nuclear genes of some ciliates like *Euplotes* (46)] ribosomal frameshifts per gene. In contrast, frameshift must occur at as many as 10 sites to produce a complete COX1 protein in *Perkinsus*, which is a surprisingly high number. If one frameshift failure occurs at any of the 10 sites due to low efficiency, only a truncated COX1 protein, and not the full-length protein, will be synthesized to deleterious effect on respiratory function of *Perkinsus*. It is known that 'stimulatory' elements such as upstream Shine-Dalgarno-like sequences or downstream pseudoknot structures promote efficient frameshifts (47,48). There are, however, no such sequences associated with the frameshift in *Pmcox1*.

Based on these observations, we suggest that the complete translation of *Pmcox1*, a *Perkinsus* mt gene, requires a mechanism that is quite accurate for high frequency and high efficiency frameshifts. 'Ten times per gene' is by far the highest frequency among the reported ribosomal frameshifts. We suggest that the function of the frameshift mechanism in *Perkinsus* mitochondria is far more efficient and active than that of the frameshifts in other organisms. Elucidation of the amino acid sequence of *Pmcox1* is still ongoing and is required to confirm the frameshift model and also to identify the start and stop codons within *Pmcox1* mRNA. We will also investigate the translational machinery in *Perkinsus* mitochondria to understand the mechanisms that promote these 'extensive' frameshifts.

ACCESSION NUMBER

AB513789.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Y. Watanabe (The University of Tokyo) for helpful comments and discussions about codon recognition, Dr T. Mogi (The University of Tokyo) for valuable information on COX1, and Dr R. Kamikawa (University of Tsukuba) for providing critical comments on mt genomes of dinoflagellates. We are also grateful to Dr T. Yoshinaga (The University of Tokyo) for providing the isolates of *P. honshuensis* and *P. olseni*.

FUNDING

Grants-in-Aid for Creative Scientific Research (18GS0314 to K.K.) and for JSPS Fellows (2105920 to I.M.) from the Japan Society for the Promotion of Science (JSPS). I. M. is a JSPS research fellow. Funding for open access charge: Grants-in-Aid for JSPS Fellows (2105920 to I.M.) from the Japan Society for the Promotion of Science (JSPS).

Conflict of interest statement. None declared.

REFERENCES

1. Burger,G., Gray,M.W. and Lang,B.F. (2003) Mitochondrial genomes: anything goes. *Trends Genet.*, **19**, 709–716.
2. Ward,B.L., Anderson,R.S. and Bendich,A.J. (1981) The mitochondrial genome is large and variable in a family of plants (cucurbitaceae). *Cell*, **25**, 793–803.
3. Feagin,J.E. (1992) The 6-kb element of *Plasmodium falciparum* encodes mitochondrial cytochrome genes. *Mol. Biochem. Parasitol.*, **52**, 145–148.
4. Burger,G., Forget,L., Zhu,Y., Gray,M.W. and Lang,B.F. (2003) Unique mitochondrial genome architecture in unicellular relatives of animals. *Proc. Natl Acad. Sci. USA*, **100**, 892–897.
5. Marande,W., Lukes,J. and Burger,G. (2005) Unique mitochondrial genome structure in diplonemids, the sister group of kinetoplastids. *Eukaryot. Cell*, **4**, 1137–1146.
6. Marande,W. and Burger,G. (2007) Mitochondrial DNA as a genomic jigsaw puzzle. *Science*, **318**, 415.
7. Sengupta,S., Yang,X. and Higgs,P.G. (2007) The mechanisms of codon reassignments in mt genetic codes. *J. Mol. Evol.*, **64**, 662–688.

8. Kondow,A., Suzuki,T., Yokobori,S., Ueda,T. and Watanabe,K. (1999) An extra $tRNA^{Gly}$(U*CU) found in ascidian mitochondria responsible for decoding non-universal codons AGA/AGG as glycine. *Nucleic Acids Res.*, **27**, 2554–2559.
9. Osawa,S., Collins,D., Ohama,T., Jukes,T.H. and Watanabe,K. (1990) Evolution of the mitochondrial genetic code. III. Reassignment of CUN codons from leucine to threonine during evolution of yeast mitochondria. *J. Mol. Evol.*, **30**, 322–328.
10. Ohama,T., Inagaki,Y., Bessho,Y. and Osawa,S. (2008) Evolving genetic code. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.*, **84**, 58–74.
11. Kück,U., Jekosch,K. and Holzamer,P. (2000) DNA sequence analysis of the complete mitochondrial genome of the green alga *Scenedesmus obliquus*: evidence for UAG being a leucine and UCA being a non-sense codon. *Gene*, **253**, 13–18.
12. Laforest,M.J., Roewer,I. and Lang,B.F. (1997) Mitochondrial tRNAs in the lower fungus *Spizellomyces punctatus*: tRNA editing and UAG 'stop' codons recognized as leucine. *Nucleic Acids Res.*, **25**, 626–632.
13. Nash,E.A., Nisbet,R.E., Barbrook,A.C. and Howe,C.J. (2008) Dinoflagellates: a mitochondrial genome all at sea. *Trends Genet.*, **24**, 328–335.
14. Waller,R.F. and Jackson,C.J. (2009) Dinoflagellate mt genomes: stretching the rules of molecular biology. *Bioessays*, **31**, 237–245.
15. Nash,E.A., Barbrook,A.C., Edwards-Stuart,R.K., Bernhardt,K., Howe,C.J. and Nisbet,R.E. (2007) Organization of the mitochondrial genome in the dinoflagellate *Amphidinium carterae*. *Mol. Biol. Evol.*, **24**, 1528–1536.
16. Jackson,C.J., Norman,J.E., Schnare,M.N., Gray,M.W., Keeling,P.J. and Waller,R.F. (2007) Broad genomic and transcriptional analysis reveals a highly derived genome in dinoflagellate mitochondria. *BMC Biol.*, **5**, 41.
17. Kamikawa,R., Nishimura,H. and Sako,Y. (2009) Analysis of the mitochondrial genome, transcripts, and electron transport activity in the dinoflagellate *Alexandrium catenella* (Gonyaulacales, Dinophyceae). *Phycol. Res.*, **57**, 1–11.
18. Norman,J.E. and Gray,M.W. (2001) A complex organization of the gene encoding cytochrome oxidase subunit 1 in the mitochondrial genome of the dinoflagellate, *Crypthecodinium cohnii*: homologous recombination generates two different *cox1* open reading frames. *J. Mol. Evol.*, **53**, 351–363.
19. Slamovits,C.H., Saldarriaga,J.F., Larocque,A. and Keeling,P.J. (2007) The highly reduced and fragmented mitochondrial genome of the early-branching dinoflagellate *Oxyrrhis marina* shares characteristics with both apicomplexan and dinoflagellate mt genomes. *J. Mol. Biol.*, **372**, 356–368.
20. Gray,M.W., Lang,B.F. and Burger,G. (2004) Mitochondria of protists. *Annu. Rev. Genet.*, **38**, 477–524.
21. Conway,D.J., Fanello,C., Lloyd,J.M., Al-Joubori,B.M., Baloch,A.H., Somanath,S.D., Roper,C., Oduola,A.M., Mulder,B., Povoa,M.M. *et al.* (2000) Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. *Mol. Biochem. Parasitol.*, **111**, 163–171.
22. Rehkopf,D.H., Gillespie,D.E., Harrell,M.I. and Feagin,J.E. (2000) Transcriptional mapping and RNA processing of the *Plasmodium falciparum* mitochondrial mRNAs. *Mol. Biochem. Parasitol.*, **105**, 91–103.
23. Villalba,A., Reece,K.S., Ordás,M.C., Casas,S.M. and Figueras,A. (2004) Perkinsosis in molluscs: a review. *Aquat. Living. Resour.*, **17**, 411–432.
24. Saldarriaga,J.F., McEwan,M.L., Fast,N.M., Taylor,F.J. and Keeling,P.J. (2003) Multiple protein phylogenies show that *Oxyrrhis marina* and *Perkinsus marinus* are early branches of the dinoflagellate lineage. *Int. J. Syst. Evol. Microbiol.*, **53**, 355–365.
25. Joseph,S.J., Fernández-Robledo,J.A., Gardner,M.J., El-Sayed,N.M., Kuo,C.-H., Schott,E.J., Wang,H., Kissinger,J.G. and Vasta,G.R. (2010) The alveolate *Perkinsus marinus*: biological insights from EST gene discovery. *BMC Genom.*, **11**, 228.
26. Matsuzaki,M., Kuroiwa,H., Kuroiwa,T., Kita,K. and Nozaki,H. (2008) A cryptic algal group unveiled: a plastid biosynthesis pathway in the oyster parasite *Perkinsus marinus*. *Mol. Biol. Evol.*, **25**, 1167–1179.
27. Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**, 4876–4882.
28. Iwata,S., Ostermeier,C., Ludwig,B. and Michel,H. (1995) Structure at 2.8 A resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Nature*, **376**, 660–669.
29. Chaput,H., Wang,Y. and Morse,D. (2002) Polyadenylated transcripts containing random gene fragments are expressed in dinoflagellate mitochondria. *Protist*, **153**, 111–122.
30. Norman,J.E. and Gray,M.W. (1997) The cytochrome oxidase subunit 1 gene (*cox1*) from the dinoflagellate, *Crypthecodinium cohnii*. *FEBS Lett.*, **413**, 333–338.
31. Matsumoto,J., Sakamoto,K., Shinjyo,N., Kido,Y., Yamamoto,N., Yagi,K., Miyoshi,H., Nonaka,N., Katakura,K., Kita,K. *et al.* (2008) Anaerobic NADH-fumarate reductase system is predominant in the respiratory chain of *Echinococcus multilocularis*, providing a novel target for the chemotherapy of alveolar echinococcosis. *Antimicrob. Agents Chemother.*, **52**, 164–170.
32. Farabaugh,P.J. (2000) Translational frameshifting: implications for the mechanism of translational frame maintenance. *Prog. Nucleic Acid Res. Mol. Biol.*, **64**, 131–170.
33. Belcourt,M.F. and Farabaugh,P.J. (1990) Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. *Cell*, **62**, 339–352.
34. Vimaladithan,A. and Farabaugh,P.J. (1994) Special peptidyl-tRNA molecules can promote translational frameshifting without slippage. *Mol. Cell. Biol.*, **14**, 8107–8116.
35. Härlid,A., Janke,A. and Arnason,U. (1997) The mtDNA sequence of the ostrich and the divergence between paleognathous and neognathous birds. *Mol. Biol. Evol.*, **14**, 754–761.
36. Mindell,D.P., Sorenson,M.D. and Dimcheff,D.E. (1998) An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. *Mol. Biol. Evol.*, **15**, 1568–1571.
37. Beckenbach,A.T., Robson,S.K. and Crozier,R.H. (2005) Single nucleotide +1 frameshifts in an apparently functional mitochondrial cytochrome *b* gene in ants of the genus *Polyrhachis*. *J. Mol. Evol.*, **60**, 141–152.
38. Russell,R.D. and Beckenbach,A.T. (2008) Recoding of translation in turtle mitochondrial genomes: programmed frameshift mutations and evidence of a modified genetic code. *J. Mol. Evol.*, **67**, 682–695.
39. Milbury,C.A. and Gaffney,P.M. (2008) Complete mitochondrial sequence of the eastern oyster *Crassostrea virginica*. *Mar. Biotechnol.*, **7**, 697–712.
40. Rosengarten,R.D., Sperling,E.A., Moreno,M.A., Leys,S.P. and Dellaporta,S.L. (2008) The mitochondrial genome of the hexactinellid sponge *Aphrocallistes vastus*: evidence for programmed translational frameshifting. *BMC Genomics*, **9**, 33.
41. Magliery,T.J., Anderson,J.C. and Schultz,P.G. (2001) Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in *Escherichia coli*. *J. Mol. Biol.*, **307**, 755–769.
42. Atkins,J.F. and Björk,G.R. (2009) A gripping tale of ribosomal frameshifting: extragenic suppressors of frameshift mutations spotlight P-site realignment. *Microbiol. Mol. Biol. Rev.*, **73**, 178–210.
43. Anderson,J.C., Wu,N., Santoro,S.W., Lakshman,V., King,D.S. and Schultz,P.G. (2004) An expanded genetic code with a functional quadruplet codon. *Proc. Natl Acad. Sci. USA*, **101**, 7566–7571.
44. Wang,L., Xie,J. and Schultz,P.G. (2006) Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.*, **35**, 225–249.
45. Sundararajan,A., Michaud,W.A., Qian,Q., Stahl,G. and Farabaugh,P.J. (1999) Near-cognate peptidyl-tRNAs promote +1 programmed translational frameshift in yeast. *Mol. Cell.*, **4**, 1005–1015.
46. Klobutcher,L.A. (2005) Sequencing of random *Euplotes crassus* macronuclear genes supports a high frequency of +1 translational frameshifting. *Eukaryot. Cell*, **4**, 2098–2105.
47. Gesteland,R.F. and Atkins,J.F. (1996) Recoding: dynamic reprogramming of translation. *Annu. Rev. Biochem.*, **65**, 741–768.
48. Giedroc,D.P. and Cornish,P.V. (2009) Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res.*, **139**, 193–208.

Parasitology International 59 (2010) 560–564

Contents lists available at ScienceDirect

Parasitology International

journal homepage: www.elsevier.com/locate/parint

Trypanosome alternative oxidase, a potential therapeutic target for sleeping sickness, is conserved among *Trypanosoma brucei* subspecies ☆

Kosuke Nakamura [a,b,*,1], Sunao Fujioka [a], Shinya Fukumoto [c,d], Noboru Inoue [d], Kimitoshi Sakamoto [a], Haruyuki Hirata [c], Yasutoshi Kido [a], Yoshisada Yabu [e], Takashi Suzuki [e], Yoh-ichi Watanabe [a], Hiroyuki Saimoto [f], Hiroshi Akiyama [b], Kiyoshi Kita [a,*]

[a] Department of Biomedical Chemistry, Graduate School of Medicine, The University of Tokyo, 7-3-1, Hongo, Tokyo 113-0033, Japan
[b] National Institute of Health Sciences, Tokyo 158-8501, Japan
[c] Center for Disease Biology and Integrative Medicine, Faculty of Medicine, The University of Tokyo, Tokyo 113-0033, Japan
[d] National Research Center for Protozoan Diseases, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido 080-8555, Japan
[e] Department of Molecular Parasitology, Nagoya City University, Graduate School of Medical Sciences, Nagoya 467-8601, Japan
[f] Department of Chemistry and Biotechnology, Graduate School of Engineering, Tottori University, Tottori 680-8552, Japan

ARTICLE INFO

Article history:
Received 26 June 2010
Received in revised form 17 July 2010
Accepted 23 July 2010
Available online 3 August 2010

Keywords:
Human African Trypanosomiasis
Sleeping sickness
Trypanosome alternative oxidase
Chemotherapy

ABSTRACT

Trypanosoma brucei rhodesiense and *T. b. gambiense* are known causes of human African trypanosomiasis (HAT), or "sleeping sickness," which is deadly if untreated. We previously reported that a specific inhibitor of trypanosome alternative oxidase (TAO), ascofuranone, quickly kills African trypanosomes *in vitro* and cures mice infected with another subspecies, non-human infective *T. b. brucei*, in *in vivo* trials. As an essential factor for trypanosome survival, TAO is a promising drug target due to the absence of alternative oxidases in the mammalian host. This study found TAO expression in HAT-causing trypanosomes; its amino acid sequence was identical to that in non-human infective *T. b. brucei*. The biochemical understanding of the TAO including its 3 dimensional structure and inhibitory compounds against TAO could therefore be applied to all three *T. brucei* subspecies in search of a cure for HAT. Our *in vitro* study using *T. b. rhodesiense* confirmed the effectiveness of ascofuranone (IC_{50} value: 1 nM) to eliminate trypanosomes in human infective strain cultures.

1. Introduction

Trypanosoma brucei, comprising three subspecies, *brucei*, *rhodesiense* and *gambiense*, causes African trypanosomiasis in mammalian hosts. *T. b. brucei* is known to cause *nagana* disease in wild and domestic animals but is non-infective to humans due to the lytic action in normal human serum [1]. On the other hand, *T. b. rhodesiense* and *T. b. gambiense* are known to cause human African trypanosomiasis (HAT), or "sleeping sickness," which currently afflicts 50,000–70,000 people and threatens over 50 million people in Africa according to a recent estimation from the World Health Organization (http://www.who.int/en/) [2]. Because African trypanosomes can escape the host's immune system by switching their cell surface antigen, variant surface glycoprotein (VSG), development of a vaccine against HAT is difficult [3]. Current chemotherapeutic treatment differs depending on the *T. brucei* subspecies and the developmental stage of the disease [4], thereby complicating diagnosis for proper treatment. Adverse events from long-term treatment, serious toxicity due to side effects, relapse and increasing prevalence of drug resistance all hamper efficient treatment of HAT [1]. Hence, there is an urgent need for a new chemotherapeutic drug therapy that can treat all disease developmental stages of different *T. brucei* subspecies.

Parasites have exploited unique energy metabolic pathways by adapting to the natural host habitat. Mitochondria in the bloodstream form of African trypanosomes are considered to be a promising chemotherapeutic target because of their unique properties not found in mammals but which are critical for parasite survival [3]. That is, these trypanosomes lack cytochromes and utilize glucose as the sole source of energy through the glycolytic pathway [3]. Trypanosome alternative oxidase (TAO) is the essential terminal enzyme for reoxidization of NADH produced during glycolysis by these parasites, and in addition to glycerol-3-phosphate dehydrogenase, it is a primary mitochondrial electron transport protein [3]. Because there

Abbreviations: HAT, Human African Trypanosomiasis; TAO, trypanosome alternative oxidase.

☆ New nucleotide sequences for alternative oxidase derived from *Trypanosoma brucei rhodesiense* and *Trypanosoma brucei gambiense* reported in this paper are available in the EMBL, GenBank and DDBJ databases under the accession nos. AB261994 and AB261993, respectively.

* Corresponding authors. Tel.: +81 3 5841 3526; fax: +81 3 5841 3444.
E-mail addresses: kosnakamura@nihs.go.jp (K. Nakamura), kitak@m.u-tokyo.ac.jp (K. Kita).
[1] Present address: National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan. Tel./fax: +81 3 3700 9397.

1383-5769/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.parint.2010.07.006

is no alternative oxidase in the mammalian host, specific inhibitor against TAO would cause no side effects, unlike the currently available anti-trypanosomiasis drugs. In this context, TAO from *T. b. brucei* has thus been extensively studied as a possible drug target against African trypanosomiasis [3,5–9].

Ascofuranone, an antibiotic isolated from a phytopathogenic fungus (*Ascochyta visiae*), was found to specifically target TAO based on biochemical analyses and structure–activity relationship studies using derivatives of ascofuranone, indicating the site of inhibition to be the same or close to its quinol binding pocket [6,10]. It should be noted that this antibiotic was also found to cure *T. b. brucei*-infected mice *in vivo* by oral administration as well as intraperitoneal administration [11]. For further biochemical study including 3 dimensional structure analysis, *E. coli* expression system for the recombinant TAO was established [7,12]. During this study, we have recognized a strain-specific difference in amino acid sequences, particularly around the C-terminus, between our *T. b. brucei* strain (TC221) [13] and a strain analyzed by another group (EATRO110) [14]. Furthermore, we examined TAO from animal infectious strains using TC221 and two other *T. b. brucei* strains, ILTat1.4 and GUTat3.1, and concluded that these strains have completely identical amino acid sequences [13]. Subsequently, the genome sequencing project using *T. b. brucei* strain (TREU927) revealed a single copy number of the TAO gene coding for identical sequences with our strains analyzed [15].

In this study, amino acid sequences of TAO from verified human infective *T. b. rhodesiense* and *T. b. gambiense* strains were deduced from their cDNA sequences. The therapeutic efficacy of the TAO inhibitor was also analyzed *in vitro* using cultures of a human infective strain, *T. b. rhodesiense*.

2. Materials and methods

2.1. Trypanosome strains

T. b. rhodesiense (IL1501) and *T. b. gambiense* (IL3253) strains were originally provided by the International Livestock Research Institute (ILRI), Nairobi, Kenya.

(i) IL1501: This parasite was originally isolated from a human patient with HAT in Kenya in 1980 [16], subsequently subcloned *in vitro* for experimental use by the limiting dilution method at the Department of Biomedical Chemistry, the University of Tokyo. This strain was confirmed as *T. b. rhodesiense* in this study.
(ii) IL3253: This parasite was isolated from a human patient with HAT in South Sudan [16]. This strain was confirmed as *T. b. gambiense* in this study.

2.2. Trypanolytic assay for human serum resistance

Human serum resistance was investigated based on a modification of the original *in vitro* trypanolytic assay established by Rifkin [17,18] as reported previously [19]. Briefly, *in vitro* grown trypanosomes in HMI-9 medium were harvested during the exponential phase, 1×10^6/ml trypanosomes were suspended in HMI-9 medium containing 10% (v/v) heat-inactivated human serum, and the suspension was cultured in a CO_2 incubator at 37 °C. Human serum was obtained from a volunteer with no history of African trypanosome infection after overnight fasting. The medium was partly changed (>50%) every 2 days. End-point survival of the parasites was determined after at least 8 days by microscopy (200× magnification) and scored as "human serum resistance" [20].

2.3. Total genome extraction

Total genomic DNA was extracted from trypanosomes by the eukaryotic genomic DNA extraction method [21]. Briefly, phosphate-buffered saline (PBS) (pH 7.4)-washed cell pellet was lysed by lysis buffer (100 mM Tris–HCl (pH 8.0), 10 mM EDTA, 0.2% (w/v) SDS, 200 mM NaCl) and then incubated with the addition of 100 μg/mL proteinase K at 56 °C overnight. After cooling to room temperature, salt solution (3 M NaCl, 0.5 M KCl, 10 mM Tris–HCl, pH 8.0) and phenol–chloroform–isoamyl alcohol (25:24:1) were added to the sample and mixed. Genomic DNA was precipitated from the aqueous supernatant by adding isopropanol and washed with 70% (v/v) ethanol. The purified DNA was dissolved in sterilized distilled water and utilized as a template for PCR identification.

2.4. Identification of HAT-causing T. brucei subspecies by PCR

PCR for the identification of *T. brucei* subspecies was carried out using *Taq* DNA polymerase (Invitrogen, Tokyo, Japan) according to the manufacturer's instructions. Specific primers used are indicated in Table 1. PCR products were separated by 1% (w/v) agarose gel electrophoresis containing ethidium bromide (0.2 μg/ml) and visualized by an ultraviolet transilluminator.

2.5. Sequence analysis of alternative oxidase from the identified T. brucei subspecies

Total RNA of trypanosomes was extracted using TRIzol (Invitrogen) and subsequently reverse-transcribed with ReverTra Ace (Toyobo, Osaka, Japan) using the following oligo(dT) primer with an adaptor sequence, [5′-GACTCGAGTCGACATCGAT$_{17}$-3′]. The resulting cDNA was used as a template for PCR. PCR was performed using *PfuTurbo* DNA polymerase (Agilent Technologies, Inc., Santa Clara, CA). Highly conserved sequences between the TAO genes of *T. b. brucei* (GenBank accession no. AB070614) and *T. vivax* (GenBank accession no. AB070521) were used to design primers to amplify the TAO gene from different subspecies. The forward and reverse primers used were [5′-CTGGAGCCTTCC-3′] and [5′-TCGATGTCGACTCGAGTC-3′], respectively. The terminal open reading frame sequence was analyzed by 5′- and 3′-RACE using a spliced leader sequence and the oligo(dT) primer adapter sequence as the primers, respectively. Table 1 summarizes the primers used for the primary PCR for RACE. Amplified PCR products were sequenced using BigDye Terminator v3.0 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions. Sequences were determined by ABI PRISM 310 Genetic Analyzer (Applied Biosystems).

2.6. In vitro ascofuranone sensitivity assay

Ascofuranone was isolated directly from fermentation of the filamentous fungus, *A. visiae*, and handled as previously described [5]. Stock

Table 1
Oligonucleotides used in this study.

Target[a]	Primer sequence	Remarks
SRA	5′-ATAGTGACAAGATGCGTACTCAACGC-3′ 5′-AATGTGTTCGAGTACTTCGGTCACGCT-3′	Ref. [22]
R-ES	5′-GTGGCGGCAGCAAAAGTATCATC-3′ 5′-ACACTCCAACACTCTCCTATC-3′	Ref. [22]
TgsGP	5′-GCTGCTGTGTTCGGAGAGC-3′ 5′-GCCATCGTGCTTGCCGCTC-3′	Ref. [23]
TAO	5′-AACGCTATTATTAGAACAGTTTCTG-3′ 5′-AAGGTGGCGCAACATKCC-3′	For 5′ RACE
TAO	5′-GACTCGAGTCGACATCGAT$_{17}$-3′ 5′-GACCTYATCAACGTGATYC-3′	For 3′ RACE

[a] Abbreviations used are: SRA, human serum resistance-associated protein; R-ES, expression site of SRA; TgsGP, *T. b. gambiense*-specific glycoprotein; and TAO, trypanosome alternative oxidase.

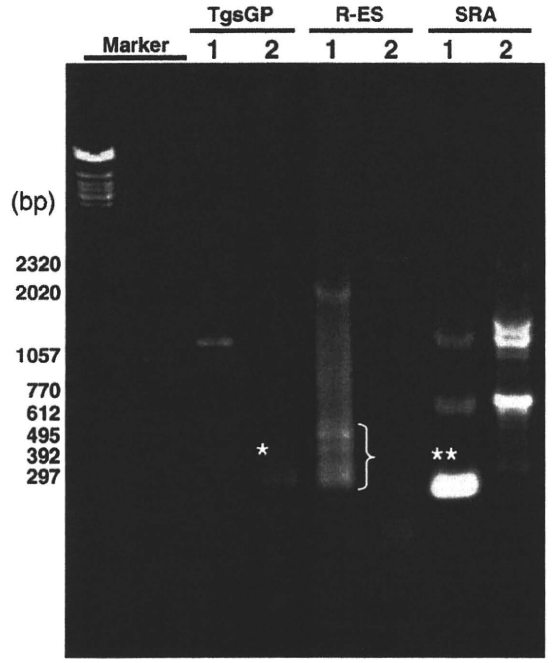

Fig. 1. Genotypical identification of *T. brucei* subspp. *rhodesiense*, *gambiense* and *brucei*. PCR was performed using genomic DNA as the template and the primers indicated in Table 1. Single asterisk indicates the 308-bp band positive for the *T. b. gambiense*-specific glycoprotein (TgsGP) gene; bracket indicates the bands positive for the human serum resistance-associated gene (SRA) expression site (R-ES) at 290, 370, 450, 530 bp; and double asterisks indicate the 284-bp SRA positive band. Lane 1, IL1501; lane 2, IL3253.

solution of ascofuranone was prepared in 100% dimethylsulfoxide (DMSO; Nacalai Tesque, Kyoto, Japan) at 10 mM and stored at −20 °C until use. Human infective *T. b. rhodesiense* at the exponential phase were centrifuged at 700×g for 5 min at room temperature and gently suspended in fresh medium to 500 cells/μl. Assays were performed in sterile 96-well microtiter plates with non-coated surface, with each well containing 100 μl parasite culture (5×10^4 bloodstream form). The first row of wells was filled with 150 μl cell suspension of the test sample at the initial concentration, and subsequent one-third serial dilutions were prepared by transferring 50 μl of the first row samples to the next row containing 100 μl cell suspension. After 18 h incubation, 10 μl Alamar Blue (Invitrogen) was added to each well and incubated for another 6 h. Absorbances of Alamar Blue in reduced (570 nm) and oxidized (600 nm) states were measured by Benchmark Plus Microplate Spectrophotometer (Bio-Rad Laboratories, Inc., Tokyo, Japan). Plates were inspected under an inverted microscope to confirm no growth changes in the controls with the same concentration of DMSO as the solvent, for determination of the MIC values, the lowest drug concentration showing no trypanosomes with normal morphology and motility compared to controls. Each sample was tested in triplicate for three independent assays.

3. Results

3.1. Identification of human infective trypanosome subspecies

A human infective form of *T. brucei* subspecies was genotypically identified by PCR using primers shown in Table 1. Genotypical identification of the trypanosome parasite was necessary since the spread of trypanosomes to new endemic areas in Africa through a movement of infected livestock and refugee as carriers has been reported [24]. One of the distinctive characteristics between *T. b. rhodesiense* and *T. b. gambiense* is human serum resistance [25]. *T. b. rhodesiense* can be differentiated by the presence of the human serum resistance-associated gene, SRA, whereas no mechanism of resistance against human serum lysis has been reported for *T. b. gambiense* [26,27]. Thus, genomic PCR targeting gene for SRA and related gene were conducted to specifically identify the *T. b. rhodesiense* strain. As shown in Fig. 1, SRA- and its expression site (R-ES)-positive PCR bands at 284 bp and at 290, 370, 450, 530 bp, respectively, were observed. This result clearly confirmed the trypanosome strain IL1501 to be *T. b. rhodesiense*. The trypanolytic assay further confirmed that a cultured IL1501 clone is resistant to lysis in the presence of human serum allowing propagation (data not shown). Moreover, as the PCR of the *T. b. gambiense*-specific glycoprotein (TgsGP) gene is reportedly unique for *T. b. gambiense* strains [23], this method was also performed for the IL3253 strain. The results were positive for TgsGP, producing a product at 308 bp, but negative for SRA or R-ES (Fig. 1), verifying IL3253 to be *T. b. gambiense*. These results are summarized in Table 2.

3.2. TAO in three T. brucei subspecies

Two positions of nucleotide polymorphism in the coding sequence of TAO are present among *T. b. brucei* strains, TC221 (GenBank accession no. AB070617.1), ILTat1.4 (GenBank accession no. AB070614.1) and EATRO 110 (GenBank accession no. U52964.2), at the 246th position (thymine in EATRO110 and cytosine in TC221 and ILTat1.4) and at the 984th position (thymine in TC221 and cytosine in ILTat1.4 and EATRO110). Importantly, the deduced amino acid sequences of TAO among all *T. b. brucei* strains analyzed are identical. Using identified *T. b. rhodesiense* and *T. b. gambiense* strains, open reading frame sequences for their TAO genes were verified from cDNA. As shown in Fig. 2, the open reading frame of TAO was 987 nucleotides in length. In comparison with the non-human infective *T. b. brucei* strain (TC221), nucleotide polymorphism was observed at the 246th position (cytosine in TC221 and IL3253 and thymine in IL1501) and 984th position (thymine in TC221 and cytosine in IL3253 and IL1501). The coding sequences of TAO in IL1501 and IL3253 were deposited in GenBank with accession numbers AB261994 and AB261993, respectively. The deduced amino acid sequence was found to be completely conserved having a length of 329 amino acids for all *T. brucei* subspecies.

3.3. In vitro ascofuranone sensitivity assay using human serum-resistant trypanosomes

Effects of ascofuranone were analyzed using the human serum-adapted IL1501 strain (*T. b. rhodesiense*). The 24-h drug sensitivity assay using the growth indicator, Alamar Blue, at 10% (v/v) showed

Table 2
Characterization of trypanosome strains.

Strain	Host origin	Subspecies	Human serum trypanolytic assay (bloodstream form)	PCR genotype analysis			GenBank accession no.
				SRA	R-ES	TgsGP	
IL1501[a]	Human (Kenya)	*T. b. rhodesiense*	R	+	+	−	AB261994
IL3253	Human (South Sudan)	*T. b. gambiense*	R	−	−	+	AB261993
TC221	Wild animal	*T. b. brucei*	S	−	−	−	AB070617

[a]A sub-culture of the original strain underwent limited dilution and was used for analysis.
R, resistant; S, sensitive.

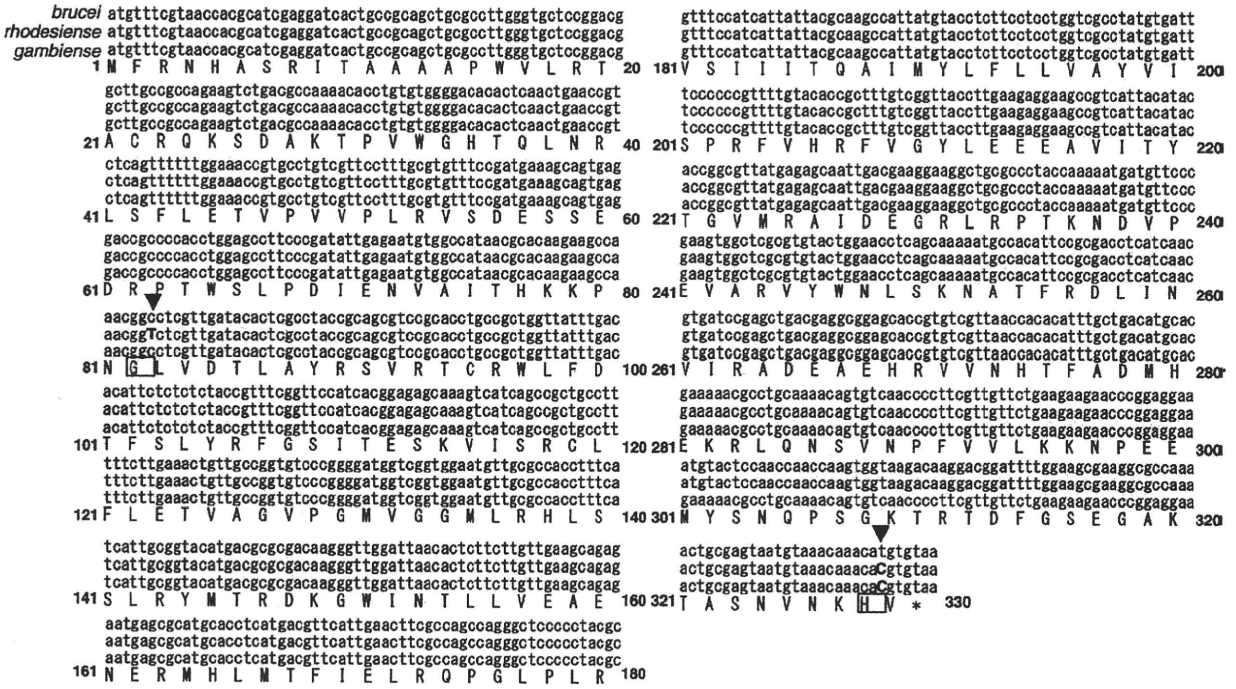

Fig. 2. Coding sequence alignment of TAO in *T. brucei* subspecies. The 987-nucleotide open reading frame sequence was translated to 329 amino acids by the universal genetic code using GENETYX-WIN program. Two positions of the amino acid sequence, which differs in the nucleotide coding sequence, are indicated in the box. Nucleotide sequence at the 246th position was cytosine in TC221 and IL3253 or thymine in IL1501 and at the 984th position was thymine in TC221 or cytosine in IL3253 and IL1501.

ascofuranone to have a dose-dependent trypanocidal effect (Fig. 3). The IC_{50} value was about 1 nM, and the MIC value was about 0.1 μM for this strain of *T. b. rhodesiense*.

4. Discussion

Sleeping sickness has a devastating impact on human health and prosperity in Africa. African trypanosome infection is always fatal unless treated. However, the currently available drugs for African trypanosomiasis have a limited effectiveness due to subspecies specificity, severe side effects and drug resistance [1,28–30]. In the past 30 years, the failure of control measures and treatment regimes has allowed this parasite to drastically increase, making it the biggest killer in certain areas of Africa, even surpassing HIV/AIDS in the provinces of Angola, Congo and Southern Sudan [31]. Currently, the choice of available drugs is limited due to the prohibitively high cost of developing new compounds showing a broad spectrum of activity [32]. Indeed, the appearance of a drug-resistant phenotype has been reported to lead to a growing number of resistant-type isolates [29,33]. Hence, a new chemotherapeutic drug that is effective at low doses, has no side effects, and is a potential treatment against all disease developmental stages of different *T. brucei* subspecies is needed.

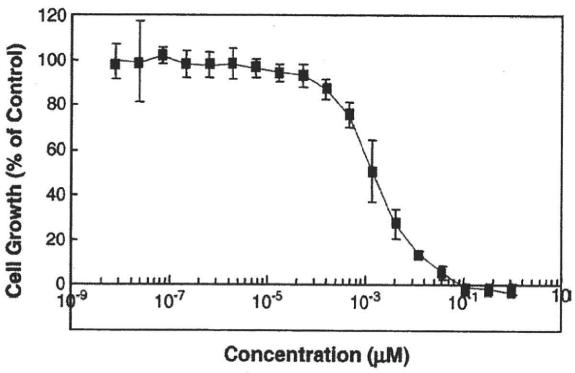

Fig. 3. *In vitro* ascofuranone sensitivity assay using human serum resistant strain, *T. b. rhodesiense*. A 24-h drug sensitivity assay was performed in sterile 96-well microtiter plates at a concentration of 500 cells/μl in the HMI-9 medium. The first row of wells was filled with 150 μl cell suspension of the test sample at the initial concentration and subsequent rows were filled with one-third serial dilutions by transferring 50 μl cell suspension from the first row to the next row of wells containing 100 μl cell suspension. After 18 h incubation, 10 μl Alamar Blue was added to each well and incubated for another 6 h. The relative difference in the reduction of Alamar Blue between the test and control samples was calculated according to the manufacturer's instructions by measuring absorbance at 570 nm and 600 nm. Data are means ± standard error of three independent experiments performed in triplicate.

Unique aspects of the parasitic mitochondria present promising chemotherapeutic targets [3], as seen by the action of the antimalarial drug atovaquone on the mitochondrial respiratory chain [34]. Atovaquone is effective against chloroquine-resistant strains, and is already being used for treatment [35]. The specific target is thought to be complex III (ubiquinol-cytochrome *c* reductase), and analysis using resistant mutants has shown that the drug acts on the ubiquinone oxidation site in the cytochrome *b* of complex III [36–38]. Along these lines, our previous screening for such inhibitors discovered two anthelmintic compounds in adult *Ascaris suum* mitochondria: nafuredin, which inhibits at the quinone-binding domain in complex I [39], and atpenin, which inhibits the quinol-fumarate reductase activity of complex II [40]. In the case of trypanosomes found in the bloodstream of mammalian hosts, glucose is their sole energy source used in the glycolytic pathway [3]. Therefore, a respiratory enzyme, such as TAO in African trypanosome mitochondrion, could be a potential chemotherapeutic target, as it is essential for reoxidizing NADH produced during glycolysis. Ascofuranone is a potent and specific TAO inhibitor [5,6,10]. Our previous studies have characterized ascofuranone as a novel trypanocidal compound that cures mice infected with *T. b. brucei* (ILTat1.4 strain) [11] and *T. vivax* (ILDat 1.2 strain) [41].

This study deduced for the first time the amino acid sequences of TAO from human infective strains of *T. b. rhodesiense* and *T. b. gambiense* using their cDNA. The sequences were found to be

completely identical to that of TAO found in non-human infective *T. b. brucei*. Another *T. brucei* strains registered as *gambiense*, IL2343 and Wellcome, were also deduced for the amino acid sequences of TAO. It was found that they all had identical amino acid sequences. Therefore, the amino acid sequences of the TAO are highly conserved among *T. brucei* subspecies. Quite recently, we have established a protocol to prepare a highly purified TAO and crystallize the enzyme [6,7]. Recent mutational analyses on the alternative oxidase from other species provided more information on the catalytic site for the target of inhibition [42]. These data allows us to understand an inhibitory mechanism of ascofuranone and to design more effective and safe derivatives of ascofuranone from in silico analysis. We are able to apply all of the result and information from the study on *T. b. brucei* TAO to combat against HAT. In fact, the drug sensitivity assay confirmed ascofuranone as a potent inhibitor for human infective trypanosome strain, *T. b. rhodesiense*, at the IC_{50} value of about 1 nM while 0.1 μM ascofuranone completely eliminates trypanosomes. We are further investigating the inhibitory effect of ascofuranone on different strains of *T. brucei*.

In summary, the therapeutic efficacy of TAO inhibitors like ascofuranone could extend to all three subspecies of *T. brucei*, including those that cause HAT. Further studies are underway to develop a derivative of ascofuranone that can treat all developmental stages of the African trypanosomiasis.

Acknowledgements

This work was supported in part by Grant-in-aid for Creative Scientific Research Grant 18GS0314, Grant-in-aid for Scientific Research on Priority Areas 18073004 from the Japanese Society for the Promotion of Science, Targeted Proteins Research Program from the Japanese Ministry of Education, Science, Culture, Sports and Technology (MEXT) and for research on emerging and re-emerging infectious diseases from the Japanese Ministry of Health and Welfare.

References

[1] Baral TN, Magez S, Stijlemans B, Conrath K, Vanhollebeke B, Pays E, et al. Experimental therapy of African trypanosomiasis with a nanobody-conjugated human trypanolytic factor. Nat Med 2006;12:580–4.
[2] Bacchi CJ. Chemotherapy of human African trypanosomiasis. Interdiscip Perspect Infect Dis 2009;2009:195040.
[3] Kita K, Nihei C, Tomitsuka E. Parasite mitochondria as drug target: diversity and dynamic changes during the life cycle. Curr Med Chem 2003;10:2535–48.
[4] Barrett MP. The rise and fall of sleeping sickness. Lancet 2006;367:1377–8.
[5] Minagawa N, Yabu Y, Kita K, Nagai K, Ohta N, Meguro K, et al. An antibiotic, ascofuranone, specifically inhibits respiration and *in vitro* growth of long slender bloodstream forms of *Trypanosoma brucei brucei*. Mol Biochem Parasitol 1997;84: 271–80.
[6] Kido Y, Sakamoto K, Nakamura K, Harada M, Suzuki T, Yabu Y, et al. Purification and kinetic characterization of recombinant alternative oxidase from *Trypanosoma brucei brucei*. Biochim Biophys Acta 2010;1797:443–50.
[7] Kido Y, Shiba T, Inaoka DK, Sakamoto K, Nara T, Aoki T, et al. Crystallization and preliminary crystallographic analysis of cyanide-insensitive alternative oxidase from *Trypanosoma brucei brucei*. Acta Crystallogr F Struct Biol Cryst Commun 2010;66:275–8.
[8] Ott R, Chibale K, Anderson S, Chipeleme A, Chaudhuri M, Guerrah A, et al. Novel inhibitors of the trypanosome alternative oxidase inhibit *Trypanosoma brucei brucei* growth and respiration. Acta Trop 2006;100:172–84.
[9] Clarkson Jr AB, Bienen EJ, Pollakis G, Grady RW. Respiration of bloodstream forms of the parasite *Trypanosoma brucei brucei* is dependent on a plant-like alternative oxidase. J Biol Chem 1989;264:17770–6.
[10] Nihei C, Fukai Y, Kawai K, Osanai A, Yabu Y, Suzuki T, et al. Purification of active recombinant trypanosome alternative oxidase. FEBS Lett 2003;538:35–40.
[11] Yabu Y, Yoshida A, Suzuki T, Nihei C, Kawai K, Minagawa N, et al. The efficacy of ascofuranone in a consecutive treatment on *Trypanosoma brucei brucei* in mice. Parasitol Int 2003;52:155–64.
[12] Fukai Y, Nihei C, Kawai K, Yabu Y, Suzuki T, Ohta N, et al. Overproduction of highly active trypanosome alternative oxidase in *Escherichia coli* heme-deficient mutant. Parasitol Int 2003;52:237–41.
[13] Fukai Y, Nihei C, Yabu Y, Suzuki T, Ohta N, Minagawa N, et al. Strain-specific difference in amino acid sequences of trypanosome alternative oxidase. Parasitol Int 2002;51:195–9.
[14] Chaudhuri M, Ajayi W, Hill GC. Biochemical and molecular properties of the *Trypanosoma brucei* alternative oxidase. Mol Biochem Parasitol 1998;95:53–68.
[15] Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et al. The genome of the African trypanosome *Trypanosoma brucei*. Science 2005;309:416–22.
[16] Inoue N, Narumi D, Mbati PA, Hirumi K, Situakibanza NT, Hirumi H. Susceptibility of severe combined immuno-deficient (SCID) mice to *Trypanosoma brucei gambiense* and *T. b. rhodesiense*. Trop Med Int Health 1998;3:408–12.
[17] Rifkin MR. Identification of the trypanocidal factor in normal human serum: high density lipoprotein. Proc Natl Acad Sci USA 1978;75:3450–4.
[18] Rifkin MR. *Trypanosoma brucei*: some properties of the cytotoxic reaction induced by normal human serum. Exp Parasitol 1978;46:189–206.
[19] Ortiz-Ordonez JC, Sechelski JB, Seed JR. Mechanism of lysis of *Trypanosoma brucei gambiense* by human serum. J Parasitol 1994;80:924–30.
[20] Brun R, Jenni L. Human serum resistance of metacyclic forms of *Trypanosoma brucei brucei, T. brucei rhodesiense* and *T. brucei gambiense*. Parasitol Res 1987;73: 218–23.
[21] Sambrook J, Russell DW. Molecular cloning: a laboratory manual 3rd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2001.
[22] Radwanska M, Chamekh M, Vanhamme L, Claes F, Magez S, Magnus E, et al. The serum resistance-associated gene as a diagnostic tool for the detection of *Trypanosoma brucei rhodesiense*. Am J Trop Med Hyg 2002;67:684–90.
[23] Radwanska M, Claes F, Magez S, Magnus E, Perez-Morga D, Pays E, et al. Novel primer sequences for polymerase chain reaction-based detection of *Trypanosoma brucei gambiense*. Am J Trop Med Hyg 2002;67:289–95.
[24] Picozzi K, Fevre EM, Odiit M, Carrington M, Eisler MC, Maudlin I, et al. Sleeping sickness in Uganda: a thin line between two fatal diseases. BMJ 2005;331: 1238–41.
[25] Ortiz JC, Sechelski JB, Seed JR. Characterization of human serum-resistant and serum-sensitive clones from a single *Trypanosoma brucei gambiense* parental clone. J Parasitol 1994;80:550–7.
[26] Raper J, Portela MP, Lugli E, Frevert U, Tomlinson S. Trypanosome lytic factors: novel mediators of human innate immunity. Curr Opin Microbiol 2001;4:402–8.
[27] Jackson AP, Sanders M, Berry A, McQuillan J, Aslett MA, Quail MA, et al. The genome sequence of *Trypanosoma brucei gambiense*, causative agent of chronic human African trypanosomiasis. PLoS Negl Trop Dis 2010;4:e658.
[28] Peregrine AS. Chemotherapy and delivery systems: haemoparasites. Vet Parasitol 1994;54:223–48.
[29] Boda C, Enanga B, Courtioux B, Breton JC, Bouteille B. Trypanocidal activity of methylene blue. Evidence for *in vitro* efficacy and *in vivo* failure. Chemotherapy 2006;52:16–9.
[30] Fairlamb AH. Chemotherapy of human African trypanosomiasis: current and future prospects. Trends Parasitol 2003;19:488–94.
[31] Matthews KR. The developmental cell biology of *Trypanosoma brucei*. J Cell Sci 2005;118:283–90.
[32] Simonite T. Protists push animals aside in rule revamp. Nature 2005;438:8–9.
[33] Dolan RB, Stevenson PG, Alushula H, Okech G. Failure of chemoprophylaxis against bovine trypanosomiasis on Galana Ranch in Kenya. Acta Trop 1992;51:113–21.
[34] Srivastava IK, Rottenberg H, Vaidya AB. Atovaquone, a broad spectrum antiparasitic drug, collapses mitochondrial membrane potential in a malarial parasite. J Biol Chem 1997;272:3961–6.
[35] Looareesuwan S, Viravan C, Webster HK, Kyle DE, Hutchinson DB, Canfield CJ. Clinical studies of atovaquone, alone or in combination with other antimalarial drugs, for treatment of acute uncomplicated malaria in Thailand. Am J Trop Med Hyg 1996;54:62–6.
[36] Kessl JJ, Lange BB, Merbitz-Zahradnik T, Zwicker K, Hill P, Meunier B, et al. Molecular basis for atovaquone binding to the cytochrome bc_1 complex. J Biol Chem 2003;278:31312–8.
[37] Syafruddin D, Siregar JE, Marzuki S. Mutations in the cytochrome *b* gene of *Plasmodium berghei* conferring resistance to atovaquone. Mol Biochem Parasitol 1999;104:185–94.
[38] Siregar JE, Syafruddin D, Matsuoka H, Kita K, Marzuki S. Mutation underlying resistance of *Plasmodium berghei* to atovaquone in the quinone binding domain 2 (Qo_2) of the cytochrome *b* gene. Parasitol Int 2008;57:229–32.
[39] Omura S, Miyadera H, Ui H, Shiomi K, Yamaguchi Y, Masuma R, et al. An anthelmintic compound, nafuredin, shows selective inhibition of complex I in helminth mitochondria. Proc Natl Acad Sci USA 2001;98:60–2.
[40] Miyadera H, Shiomi K, Ui H, Yamaguchi Y, Masuma R, Tomoda H, et al. Atpenins, potent and specific inhibitors of mitochondrial complex II (succinate-ubiquinone oxidoreductase). Proc Natl Acad Sci USA 2003;100:473–7.
[41] Yabu Y, Suzuki T, Nihei C, Minagawa N, Hosokawa T, Nagai K, et al. Chemotherapeutic efficacy of ascofuranone in *Trypanosoma vivax*-infected mice without glycerol. Parasitol Int 2006;55:39–43.
[42] Crichton PG, Albury MS, Affourtit C, Moore AL. Mutagenesis of the *Sauromatum guttatum* alternative oxidase reveals features important for oxygen binding and catalysis. Biochim Biophys Acta 2010;1797:732–7.

Mitochondrion 11 (2011) 273–278

Contents lists available at ScienceDirect

Mitochondrion

journal homepage: www.elsevier.com/locate/mito

Concatenated mitochondrial DNA of the coccidian parasite *Eimeria tenella*

Kenji Hikosaka [a], Yutaka Nakai [b], Yoh-ichi Watanabe [c], Shin-Ichiro Tachibana [a], Nobuko Arisue [d], Nirianne Marie Q. Palacpac [d], Tomoko Toyama [a,d], Hajime Honma [a,e], Toshihiro Horii [d], Kiyoshi Kita [c], Kazuyuki Tanabe [a,*]

[a] *Laboratory of Malariology, International Research Center of Infectious Diseases, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka 565-0871, Japan*
[b] *Laboratory of Sustainable Environmental Biology, Graduate School of Agricultural Science, Tohoku University, Osaki, Miyagi 989-6711, Japan*
[c] *Department of Biomedical Chemistry, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan*
[d] *Department of Molecular Protozoology, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka 565-0871, Japan*
[e] *Japan Society for the Promotion of Science, Japan*

ARTICLE INFO

Article history:
Received 28 June 2010
Received in revised form 12 October 2010
Accepted 25 October 2010
Available online 31 October 2010

Keywords:
Mitochondrion
Mitochondrial genome
Eimeria
Plasmodium
Apicomplexa
Nuclear mitochondrial DNA

ABSTRACT

Apicomplexan parasites of the genus *Plasmodium*, pathogens causing malaria, and the genera *Babesia* and *Theileria*, aetiological agents of piroplasmosis, are closely related. However, their mitochondrial (mt) genome structures are highly divergent: *Plasmodium* has a concatemer of 6-kb unit and *Babesia/Theileria* a monomer of 6.6- to 8.2-kb with terminal inverted repeats. Fragmentation of ribosomal RNA (rRNA) genes and gene arrangements are remarkably distinctive. To elucidate the evolutionary origin of this structural divergence, we determined the mt genome of *Eimeria tenella*, pathogens of coccidiosis in domestic fowls. Analysis revealed that *E. tenella* mt genome was concatemeric with similar protein-coding genes and rRNA gene fragments to *Plasmodium*. Copy number was 50-fold of the nuclear genome. Evolution of structural divergence in the apicomplexan mt genomes is discussed.

1. Introduction

Mitochondria, organelles essential for a range of cellular processes and cellular signaling, are ubiquitous in all eukaryotes. Mitochondrial (mt) genomes exhibit remarkable variation in structure and size (Gray et al., 2004), from the 6-kb genome in the malaria parasite *Plasmodium* (Feagin, 1992) to the large (180 to 2400 kb) mt genome in land plants (Ward et al., 1981; Palmer et al., 1992). *Plasmodium* belongs to the phylum Apicomplexa, with more than 5000 species, all clinically and/or economically important pathogens (Levine, 1988): *Eimeria*, responsible for the diseases of intestinal coccidiosis in intensively reared livestock; *Toxoplasma*, etiological agent of toxoplasmosis in immune-compromised patients and congenitally infected fetuses; *Cryptosporidium*, pathogens for cryptosporidiosis in humans and animals; *Babesia*, causing babesiosis in ruminants and humans; and *Theileria*, causal agents of tropical theileriosis and East Coast fever in cattle.

Mt genomes of a few apicomplexan genera have been studied, and available data suggest that they are remarkably diverse in structure and genome organization. The minuscule 6-kb tandemly repeated linear or concatenated mtDNA of *Plasmodium* encodes only three protein-coding genes (cytochrome *c* oxidase subunits I [*cox1*] and III [*cox3*] and cytochrome *b* [*cob*]) in addition to large subunit (LSU) and small subunit (SSU) ribosomal RNA (rRNA) genes (Preiser et al., 1996). The two rRNA genes are highly fragmented with 19 identified rRNA pieces (Feagin et al., 1997). The arrangement of these mt genes is completely conserved in the genus (Perkins, 2008). In the genera of *Babesia* and *Theileria*, known as piroplasms, closely related to *Plasmodium* (Lau, 2009), the mt genomes are monomeric linear, from 6.6 kb to 8.2 kb, with terminal inverted repeats on both ends (Kairo et al., 1994; Hikosaka et al., 2010). Although the *Babesia/Theileria* mt genomes encode the same three protein-coding genes, gene array and transcriptional direction are different. Furthermore, only six fragmented LSU have been identified in the *Babesia/Theileria* mt genomes, with fragmentation different from that of *Plasmodium*. Thus, the mt genomes of *Plasmodium* and *Babesia/Theileria* are structurally highly divergent regardless of their close relatedness. Although the mt genome of *Toxoplasma gondii* has yet to be sequenced, multiple copies of partial mt genes (*cox1* and *cob*) were found to be scattered throughout the nuclear genome (Ossorio et al., 1991). In *Cryptosporidium*, the mitochondrion is reduced to mitosome and has no DNA (Mogi and Kita, 2010). The phylum Apicomplexa, therefore, encompasses a large number of interesting genera to further understand the evolution of mt genomes.

* Corresponding author. Laboratory of Malariology, International Research Center of Infectious Diseases, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan. Tel.: +81 6 6879 4260; fax: +81 6 6879 4262.
E-mail address: kztanabe@biken.osaka-u.ac.jp (K. Tanabe).

1567-7249/$ – see front matter © 2010 Elsevier B.V. and Mitochondria Research Society. All rights reserved.
doi:10.1016/j.mito.2010.10.003

In Apicomplexa, *Plasmodium* and *Babesia/Theileria* belong to the class Haematozoea, and *Eimeria* and other intestinal coccidian parasites including *Toxoplasma* belong to the class Coccidea (Hausmann and Hülsmann, 1996). The genus *Eimeria* undergoes all of its developmental stages in one host, whereas parasites belonging to the class Haematozoea require two hosts to complete their life cycles, namely, the sexual development in invertebrate vectors and the asexual development in vertebrate host. Coccidian parasites have, nevertheless, complex developmental cycles: first, oocysts excreted from the hosts undergo differentiation (sporulation) in the environment and become infective. When ingested by a host animal, oocysts undergo rounds of discrete, expansive asexual reproduction (merogony and schizogony) in the intestine, followed by sexual differentiation, fertilization and shedding of unsporulated oocysts (Jeurissen et al., 1996). *Eimeria tenella* is one of the most important *Eimeria* species, as it causes intestinal coccidiosis in domestic fowls (*Gallus gallus*), imposing enormous economic losses (Shirley et al., 2004). In *E. tenella*, two extrachromosomal DNAs have been demonstrated by pulsed-field gel electrophoresis (Dunn et al., 1998): one is the 35 kb apicoplast genome (Cai et al., 2003) and the other is a smaller size mt genome. The primary structure and the gene organization of *E. tenella* mt genome, however, remain undetermined. In this study, we report the mt genome sequence of *E. tenella*, and show that the *E. tenella* mt genome has the form of a tandemly repeated linear element or concatemer structure, and contains 19 rRNA fragments as well as three protein-coding genes. This finding indicates that the mt genomes of both *Eimeria* and *Plasmodium* retain common structural features. We discuss evolution of structural divergence in the apicomplexan mt genomes. Additionally, we identified nuclear genome DNA fragments that are shared by the mt genome of *E. tenella*.

2. Materials and methods

2.1. Blast search for mt genome sequence

A contig of *Eimeria tenella* (Houghton strain), containing mtDNA, was retrieved from *Eimeria tenella* GeneDB (http://www.genedb.org/Homepage/Etenella) using the following gene names: cytochrome *c* oxidase subunit 1, cytochrome *c* oxidase subunit 3 and cytochrome *b*. Two unfinished genomic sequences (EIMER_contig_00018071 and EIMER_contig_00018452), encoding putative COX1 and putative COB, respectively, were obtained from the Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/Projects/E_tenella/).

2.2. DNA sequencing

E. tenella NIAH strain was maintained at Tohoku University by routine passage through chickens. Oocyst stage parasites were collected from feces of infected chickens and purified by the centrifugation method (Nakai et al., 1993). Purified oocysts were subjected to 5 times repeated freeze–thawing. Parasite genomic DNA was isolated using QIAamp DNA Blood Mini Kit (QIAGEN, Hilden, Germany). Genomic DNA of *Plasmodium gallinaceum* (8A strain) was kindly provided by the late M. Shahabuddin (NIAID/NIH, USA). Nucleotide sequences of the *E. tenella* mt genome; small subunit (SSU) and large subunit (LSU) rRNA genes of the *E. tenella* apicoplast genome; *cox1* and *cob* of the *P. gallinaceum* mt genome were determined by direct sequencing of polymerase chain reaction (PCR) products using specific primers (Supplementary Table 1A) designed from retrieved sequences. Amplification conditions, PCR product purification (QIAquick PCR purification kit, QIAGEN) and DNA sequencing of two independent PCR products were carried out as previously described (Hikosaka et al., 2010). Sequencing primers were designed to cover target regions in both directions. The sequences obtained in this study have been deposited in DDBJ/EMBL/GenBank with the following accession numbers: AB564272 to AB564276.

2.3. Gene annotation

Nucleotide sequences of *E. tenella* were aligned with reported sequences from *P. falciparum* (GenBank accession # M76611), *P. mexicanum* (EF079653), *B. bovis* (AB499088), *T. parva* (AB499089) and *T. annulata* (NW_001091933) by CLUSTAL W (Thompson et al., 1994) with manual corrections. Protein-coding genes were predicted using previously annotated sequences from the five parasite species. Putative rRNA genes were identified essentially as described (Hikosaka et al., 2010). MtDNA sequence or annotated rRNA gene fragments from *P. falciparum* (M76611) were used as query under suggested algorithm parameters (Freyhult et al., 2007) in NCBI BLAST 2.2 (Altschul et al., 1990). The termini of the candidate genes were assigned using aligned sequences and putative base-pairings between fragments proposed for *P. falciparum* mt rRNA fragments, and secondary structure predicted by CentroidHomfold (Hamada et al., 2009).

2.4. Southern blot hybridization

Genomic DNA of *E. tenella*, either undigested or digested with *Hind*III or *Pvu*II, were electrophoresed on 1.0% agarose gels in TAE (40 mM Tris–acetate, 1 mM EDTA) and transferred to a positively charged nylon membrane (Amersham Hybond-N+, GE Healthcare, Little Chalfont, England). PCR products specifically amplified for target regions of the *E. tenella* mt genome were labeled with digoxigenin-dUTP using the DIG High Prime DNA Labeling and Detection Starter Kit II (Roche Diagnostics, Rotkreuz, Switzerland). After overnight hybridization with the DIG-labeled DNA probes, blots were washed twice with 2× SSC, 0.1% SDS and twice with 0.5× SSC, 0.1% SDS, at 65 °C for 15 min. Hybridization signals were detected using Detection Starter Kit II. Southern blot hybridization was also done in parallel with *P. falciparum*, whose mt genome structure is known to be circular or linear concatenated with numerous branching off (Preiser et al., 1996).

2.5. Phylogenetic analysis

The concatenated amino acid sequences of COX1 and COB (755 sites) from 15 apicomplexan parasites (Supplementary Table 2) were used for phylogenetic analysis. A free-living ciliate, *Tetrahymena thermophila* (Brunk et al., 2003), was included as an outgroup. COX3 was not used, since *cox3* of *T. thermophila* is present in the nuclear (but not mt) genome. We constructed the maximum likelihood (ML) phylogenetic trees by the PROML program in PHYLIP version 3.68 (Felsenstein and Churchill, 1996). CODEML program in PAML version 4.2 (Yang, 2007) was used to estimate the Γ shape parameter value α. Bootstrap analysis was done by applying PROML to 100 re-sampled datasets produced by SEQBOOT program in PHYLIP.

A phylogenetic tree of LSU sequences of the mt genomes (LSUE, LSUF, LSUG and RNA10, 379 sites in total) was constructed using the ML method implemented in PAUP* 4.0 b10 (Swofford, 2002). SSU/LSU sequences of the apicoplast genomes (3045 sites) (Supplementary Table 2) were also used for phylogenetic analysis. The non-photosynthetic flagellate, *Astasia longa*, was included as an outgroup instead of *T. thermophila* which lacks a plastid genome. Bootstrap probability was estimated from 1000 heuristic replicates. For statistical comparisons among the best-tree and its alternatives, p-values of the KH test (Kishino and Hasegawa, 1989), the SH test (Shimodaira and Hasegawa, 1999) and the AU test (Shimodaira, 2002) were obtained.

2.6. Search for nuclear mitochondrial DNA in E. tenella

In some eukaryotes, nuclear genomes contain DNA segments which have a high sequence similarity to mtDNA (Caro et al., 2010; Hazkani-Covo et al., 2010). These sequences are considered to have been derived from mtDNA and thus designated as nuclear mtDNAs