論文の内容の要旨

論文題目　Mitochondrial gene expression system involving extensive frameshift in a shellfish pathogen *Perkinsus*

和訳　貝類寄生虫 *Perkinsus* における高頻度なフレームシフトを伴うミトコンドリア遺伝子発現系

指導教員　北 潔　教授

東京大学 大学院医学系研究科 国際保健学専攻
博士後期課程　平成 20 年 4 月入学

増田　功

I. Introduction

I-i. Apicomplexan parasites and their mitochondria as drug targets

The spread of infectious diseases has long been a big worldwide problem. Despite the generations of efforts in preventing the spread, millions of people die in a year. Malaria is one of the most devastating infectious diseases, which is majorly prevalent in tropical and subtropical areas. The malaria parasites *Plasmodium* spp. are members of the phylum Apicomplexa, a large eukaryotic assemblage comprised of unicellular parasites. This phylum encompasses so many causative agents of human infectious diseases such as *Toxoplasma* and *Cryptosporidium*, which occasionally cause fatal illnesses. It is an urgent task to develop effective chemotherapy for these parasitic diseases, so that finding potential drug targets is highly needed.

The mitochondrion is a eukaryotic organelle functioning as the center of energy metabolism in cells. Especially, the mitochondrial electron-transport chain (mtETC) directly contributes to the energy production by forming the proton gradient which drives the ATP synthase. Intensive studies have revealed that the mtETC of apicomplexans are divergent and therefore can be promising drug targets. For instance, the cyanide-resistant alternative oxidase (AOX) in *Cryptosporidium* is sensitive to ascofuranone, and atovaquone specifically binds to the complex III of *Plasmodium* mtETC, interfering with the electron-flow. Another notable feature of apicomplexan mitochondria is observed in their genomes. Generally, apicomplexan mt genomes are 6-8 kb long linear molecules, the known smallest mt genomes. This mt genome element contains only three protein-coding genes and unusually fragmented rRNA genes, and the protein-coding genes seem to lack canonical start and stop codons. Given the importance of the gene products and the probable difference in the expression mechanism between parasites and

mammalian hosts, they can also be drug targets. As there remain several problems against the establishment of chemotherapy, such as the emergence of drug-resistant strains, it is necessary to further characterize the parasite's mitochondria and examine their potential as drug targets.

I-ii. *Perkinsus*, a shellfish pathogen closely related to apicomplexans

Some of the characteristic features of the apicomplexan mitochondria are shared in the related lineages. In particular, the sister lineage dinoflagellates share some features seen in mt genomes with apicomplexans. Researchers have recently paid attentions to the ancestral species of apicomplexans and dinoflagellates, as it can be anticipated that a comprehensive understanding on physiology of many apicomplexan and dinoflaegllate species would be obtained. One closely related organism to apicomplexans is a shellfish pathogen *Perkinsus*, which is branched between the two groups. This close relationship, along with the similarities in the ultrastructure and life cycle, implies that *Perkinsus* shares several biological features with apicomplexans and is expected as a model organism that would characterize mitochondria of apicomplexans and dinoflagellates. It is also notable that *Perkinsus* can be maintained in host-free *in vitro* culture and sufficient amount of proteins for biochemical assays are easily obtained. Accordingly, it is of great significance to characterize the mitochondrial energy conversion system including the mtETC functions and mt gene expression system of *Perkinsus*, from an evolutionary viewpoint.

In such backgrounds, I started to investigate mitochondria of *Perkinsus*. As almost all current studies on *Perkinsus* do not focus on mitochondria, the main purpose of this study is to create a platform to characterize *Perkinsus* mitochondria. I attempted to i) establish a preliminary preparation protocol for *Perkinsus* mitochondria, ii) outline the *Perkinsus* mtETC by *in silico* analyses and biochemical experiments, and iii) determine and analyze the mt gene sequence.

II. Attempts for preparation of enriched *Perkinsus* mitochondria

Initially, I attempted to establish an enrichment protocol for *Perkinsus* mitochondria. To mildly disrupt the cells, I tried chemical disruption methods using digitonin in the presence or absence of cellulase, as the *Perkinsus* cell is surrounded by the cell wall. But these methods did not lyse the cell membrane and cell wall, unfortunately. Among several physical disruption methods tested, the N_2 cavitation method was utilized, which is also employed in cell homogenization of *Plasmodium* to obtain mitochondria. In preliminary experiments, *Perkinsus* cells were disrupted by incubation with N_2 gas dissolved at 300 psi at 4°C, and the crude organellar fraction was subsequently ultracentrifuged at 23,000 rpm for 1 h. The resultant sample was separated into 24 fractions by peristaltic pump. Several fractions clearly showed the activity of the mtETC complex II, succinate dehydrogenase (SDH). This is the first biochemical evidence of mitochondria with an active mtETC in *Perkinsus*. Among them, fractions Nos. 12-14 showed higher total and specific SDH activities, indicating that these fractions contained more mitochondria. For a purification method, I need to modify and improve the protocol. The

specific distribution of mitochondria will be endorsed by co-distribution of other mtETC enzyme activities and by immunoblotting assay. It should be ensured that other organelles like plastids and endoplasmic reticulum are not co-distributed with mitochondria. Also, the intactness of mitochondria has to be ensured for density-based fractionation. For an easier method to see if the cells are broken, I am now establishing transformed cell lines which have green fluorescent protein signal in mitochondria. After these refinements, it will surely contribute to the specific studies of *Perkinsus* mitochondria.

III. Genomic and biochemical characterization of *Perkinsus* mtETC proteins

To outline the *Perkinsus* mtETC, genomic and biochemical analyses were conducted. Biochemical assays detected activities of the mtETC complexes II, III, IV and AOX, and their combined activities as well. Then BLASTp-based sequence searches were performed in public database using mtETC protein sequences from related species as queries. As a result, I revealed that *Perkinsus* possessed homologs for main subunits of the mtETC complexes II-IV and the F_OF_1-ATP synthase with high sequence similarity to those of related organisms. Curiously, I did not find homologs for CybL, CybS (complex II), cytochrome *b* (complex III), cytochrome *c* oxidase subunit 1 (COX1), and COX3 (complex IV). In spite of their absence, the activities of SDH and succinate-ubiquinone reductase (SQR) which were attributed to the complex II were clearly observed by biochemical assays. This indicated the presence of the CybL and CybS, which are possibly divergent as suggested for *Plasmodium*. Similarly, the activities of the complexes III and IV were demonstrated although the three important subunits were missing. Importantly, these three genes are encoded exclusively in mt genome in apicomplexans and dinoflagellates. It was proposed that these genes are highly divergent, or that *Perkinsus* mt genome sequences were just not added to the public database. Additionally, I found homologs for unconventional enzymes: the type II nicotiamide dinucleotide dehydrogenase (NDH2) observed in many apicomplexans, and the AOX. The apparent ratio of AOX contribution to the quinol oxidation was around 40%. As a whole, I confirmed here that the *Perkinsus* mtETC contained conventional complexes II-IV, the alternative type of NDH, and the additional terminal oxidase AOX, with detectable activities. The presence of alternative enzymes supports the similarity in the mtETC organization between apicomplexans and *Perkinsus*.

IV. Analysis of *Perkinsus* mitochondrial gene and its specific expression system

IV-i. Determinatoin of the full-length mRNA sequence of *P. marinus cox1* and its genomic localization

It is intriguing that the size and the organization of the mt genome vary greatly among the eukaryotic taxa, and the mt gene expression system is more subjected to code change like codon reassignment and recoding, than nuclear system. As mentioned before, apicomplexans possess

highly unusual mt genomes and the gene expression system remains to be characterized. Interestingly, dinoflagellate mt genomes encode the same set of genes as apicomplexans, despite that they are composed of heterogeneous DNA molecules. Since *Perkinsus* mt gene sequences were not found at all by BLAST-based sequence searches in public database, I tried to obtain the *P. marinus* cryptic mt genes by molecular biological approaches. I discovered two tiny AT-rich fragments in *P. marinus* contigs in public database showing high predicted-amino-acid identity with the partial COX1 sequence of a dinoflagellate. Based on the features of flanking sequences, these *cox1*-like fragments were found to be the remnant DNA that had been transferred from mt genomes into the nuclear genome. As it was anticipated that these fragments were similar to the mitochondrial counterparts they had originated from, I prepared PCR primers based on these sequences and amplified the inner part of the authentic mitochondrial *cox1* from *P. marinus*. After sequencing, I amplified and determined the full-length *cox1* mRNA sequence by performing RACE. This mRNA was 1434 nt long, and the overall AT content was 80.9%. As a whole, this gene was similar to *cox1* of apicomplexans and dinoflagellates with an E-value less than 10^{-70}; hereafter, I refer to this sequence as *Pmcox1*.

To determine the genomic localization of *Pmcox1* and to infer the *Perkinsus* mt genome structure, I conducted Southern hybridization using total *P. marinus* DNA. Probes for the nuclear large subunit ribosomal DNA hybridized to the stacked, high molecular-weight, chromosomal DNA in the uncut DNA sample. In sharp contrast, *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. The smear signals suggested that *Pmcox1* resides on relatively small, heterogeneous non-chromosomal DNA. These hybridization data are congruent with previous reports on other dinoflagellates, suggesting that *Pmcox1* is encoded on multiple heterogeneous DNA molecules, which is similar to the structure inferred for dinoflagellate mt genomes.

IV-ii. Frameshifts at all AGG and CCC codons suggested for translation of *Pmcox1*

Pmcox1 lacked canonical start and stop codons in terminal regions, similar to apicomplexan and dinoflagellate mt genes. This indicates that *Perkinsus* also utilizes alternative mechanisms for initiation and termination of translation in mitochondria, as inferred also in the apicomplexans and dinoflagellates. Unexpectedly, this mRNA was not translated in a single reading frame with standard codon usage; several stop codons appeared in all three frames. By BLASTx-based searches, I identified eleven partial COX1-like amino acid sequences that appeared separately in all three reading frames. Most intriguingly, the conserved AGGY (8 sites) or CCCCU (2 sites) motifs appeared to coincide with the transitions between the coding-blocks, and AGG and CCC always appeared in-frame preceded by the predicted COX1-coding blocks (Figure 1). Based on these observations, I postulated a modified decoding which could shift the reading frame during translation: specific frameshifts at every in-frame AGG and CCC. Accordingly, I prepared a putative PmCOX1 amino acid sequence in the following manner. I eliminated the A residues of the AGGY motifs and made a +1 frameshift, making GGY instead of AGG in-frame. I also deleted the first two C residues of CCCCU motifs and made a +2

frameshift, making CCU instead of CCC in-frame. Along this manner, all the 11 “blocks” were connected into one consecutive coding sequence.

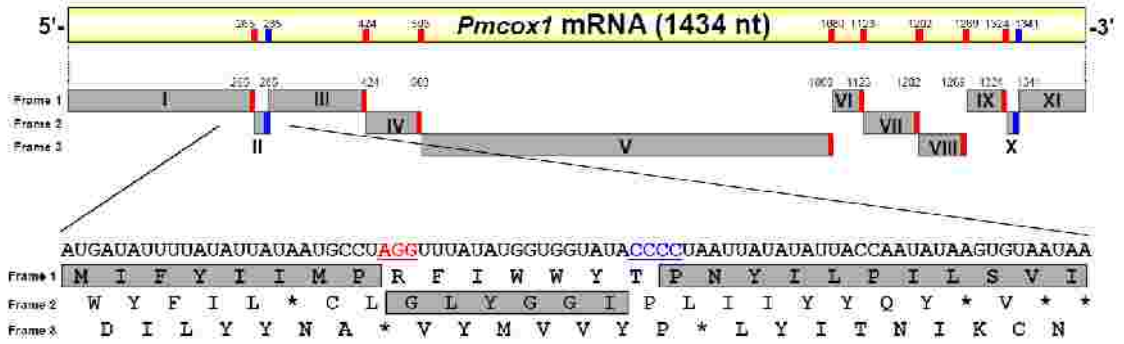

Figure 1. Schematic model of *Pmcox1* mRNA and the COX1-like blocks. COX1-like amino acid sequences distributed across three reading frames are represented by gray boxes labeled with Roman numerals. Red and blue bars at the end of COX1-like blocks indicate AGGY motifs containing in-frame AGG, and CCCCU motifs including in-frame CCC, respectively. The nucleotide sequence and its three frame translation around the junction of the coding blocks I-III are described, with AGG and CCC codons colored in red and blue, respectively.

This predicted PmCOX1 sequence based on my frameshift model retains the conserved residues essential for COX functions, thereby it reinforces the validity of the frameshift model. The PmCOX1 sequence also conserves the glycine and proline residues, which are most common in the proximity of the AGGY and CCCCU motifs, respectively. Moreover, the highly conserved tryptophans coded by UGA were deduced only with the frameshift model for PmCOX1, which is often observed in mt genes of other organisms. The frameshift model is further supported by the conservation of all the ten frameshift motifs in the *cox1* ortholog from another *Perkinsus* species, *P. olseni*. Besides, the *cob*-like fragment found in the *P. marinus* whole-genome shotgun assembly contained five in-frame AGG which would induce +1 frameshift to connect discontinuous COB-like fragments. Considering the absence of such frameshift motifs in *Perkinsus* nuclear genes, it was assumed that an unconventional event occurred during translation exclusively in mitochondria of *Perkinsus* species. Notably, this is the first evidence of a frameshift-involving translation system in protist mitochondria.

IV-iii. Possible mechanisms for the unusually frequent frameshift in translation of *Perkinsus* mt genes

I propose two possible mechanisms for this unusual frameshift in *Perkinsus* mitochondrial translation system. The first one is a ribosomal frameshift observed in a wide range of organisms during translation. In the case of +1 ribosomal frameshift, observed also in some mt genes, 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 1 base. Based on previous studies, I hypothesized that ribosomes in *Perkinsus* mitochondria skip

the A residue in the first position of the in-frame AGG in the AGGY motif and the first two C residues in the CCCCU motif by shifting forward by one base at in-frame CCC (Figure 2A). Alternatively, specialized tRNAs that recognize non-triplet codons may be utilized at frameshift sites. Naturally occurring deviant tRNAs recognize four-base codons and act as suppressors of nonsense mutations, and artificial tRNAs bearing modified loops can recognize quadruplet and even quintuplet codons. In the case of *Pmcox1*, specialized tRNAs may recognize AGGY (for glycine) and CCCCU (for proline) to enable the proposed frameshifts (Figure 2B).

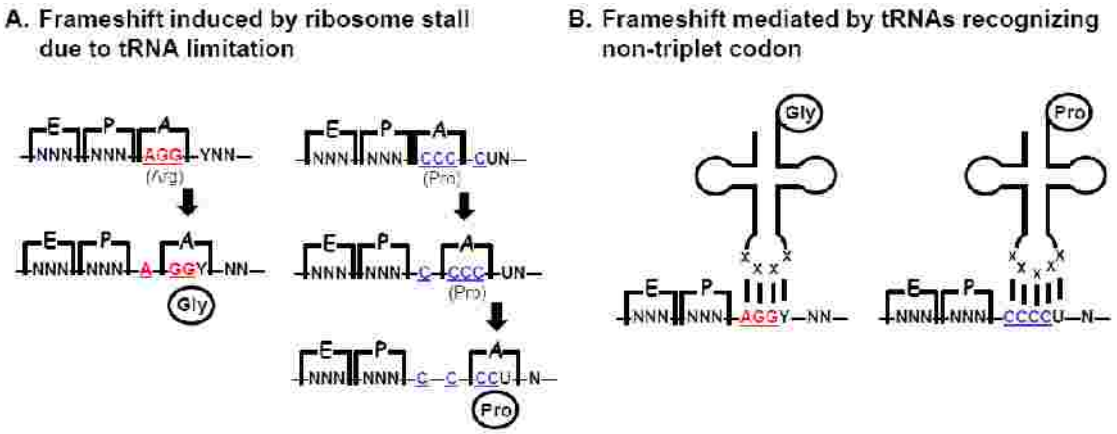

Figure 2. Possible mechanisms of frameshift-dependent translation at AGG and CCC codons of *Perkinsus* mt genes.

Regardless of the mechanism, it should be noted that the 100% frameshift efficiency has never been observed so far. Lower frameshift efficiencies are not lethal to organisms with frameshift-dependent genes, because there is only one or at most two frameshifts per one gene. In contrast, frameshift must occur at as many as ten sites to produce a complete COX1 protein in *Perkinsus*, which is by far the highest frequency among the relevant studies. With only one failure in any of ten frameshifts due to low efficiency, a truncated PmCOX1 protein will be synthesized which will affect the respiration. The frameshift-promoting elements like upstream Shine-Dalgarno-like sequences or downstream pseudoknot structures were not found around the frameshift sites in *Pmcox1*. Based on these observations, I suggest that the complete translation of *Pmcox1* requires a quite accurate mechanism for highly frequent and efficient frameshifts. I will analyze the actual PmCOX1 sequence, and also investigate the translation machinery in *Perkinsus* mitochondria to understand the mechanisms allowing these “extensive” frameshifts.

V. Conclusions and perspectives

This study is a pioneering work on *Perkinsus* mitochondria, and contains two major important points with great scientific significance. The first point is that I created the foundation for studying the *Perkinsus* mitochondria. I outlined the *Perkinsus* mtETC by genomic and biochemical approaches, and determined and characterized the first *Perkinsus* mt gene sequence. I demonstrated the similarities in the mtETC organization and mt gene expression machinery

between *Perkinsus* and apicomplexan parasites. These observations, along with the ease in handling, will prompt us to consider that *Perkinsus* may be a highly suitable material to investigate the properties of mtETC enzymes and the mt gene expression machinery of apicomplexans, both of which are the candidates of antiparasitic drug targets.

The second is that I discovered an unusual phenomenon: extremely frequent frameshifts in translation. Identifying the components employed in translation like tRNAs and ribosomes in *Perkinsus* mitochondria would delineate a specific molecular mechanism which extends our general view on decoding and might give rise to invaluable insights into the essence of translation. Furthermore, considering its phylogenetic position and the threat to fisheries, this work and future related studies will be informative also to the evolutionary protistology and fisheries sciences. As a conclusion, this study encompasses basic and essential information in studying mitochondria of *Perkinsus* as a relative of apicomplexans, and presents a discovery of an unusual and tantalizing phenomenon, making contributions to a wide range of scientific fields.