DIVISION OF GENOME DYNAMICS

Professor:	*HORIUCHI, Takashi*
Associate Professor:	*KOBAYASHI, Takehiko*
Research Associate:	*JOHZUKA, Katsuki*
Technical Staff:	*MOROOKA, Naoki*
JSPS Research Fellow:	*CUI, Tailing*
Postdoctoral Fellows:	*KODAMA, Ken-ichi*
	AUSTEN, R. D. Ganley
	OHSUMI, Katsufumi
	SERIZAWA, Naomi
	WATANABE, Taka-aki
	IDE, Satoru
Graduate Student:	*ITAZU, Masako*
Secretary	*MIKAMI, Yuriko*

The genomes of higher organisms contain significant amounts of repetitive sequences which are, in general, unstable. At present, neither the physiological function(s) of these repeated sequences, nor the mechanisms responsible for their production or for controlling instability are fully understood. To clarify these aspects, we are pursuing several lines of investigation using *Escherichia coli*, *Saccharomyces cerevisiae* and Chinese Hamster Ovary (CHO) cells. In 2006 we discovered a new role of condensin in maintaining a long repeated structure of rDNA. We also addressed the question of why almost all genomes in prokaryotes are linear. To approach this, we attempted to linearize the circular *E. coli* genome by utilizing the linearization system of a lysogenic *E. coli* phage N15. Following the success of this endeavor, we are now comparing the properties of the linear and the circular genome strains.

I. Analysis of mechanisms maintaining the repeated structure of ribosomal RNA genes

In most eukaryotes, the rDNAs are clustered in long tandem repeats on only one or a limited number of different chromosomes. Although the total number of these rDNA repeats appears to be maintained at a level appropriate for each organism, gene with such a repeated structure are in general thought to be unstable because of a high frequency of recombinational events. Thus, it might be expected that organisms have developed systems to regulate recombination within rDNA repeats.

In the yeast *Saccharomyces cerevisiae*, 200 copies on average are tandemly arrayed in a central position on the longest chromosome (XII). Recombinational events within the rDNA repeats in normal growing yeast cells appear to be mostly mediated by a *FOB1*-dependent system. *FOB1* is the gene required for replication fork blocking activity at replication fork barrier (RFB) sites, rDNA region-specific recombination and expansion/contraction of rDNA repeats. The latter two activities are likely to be triggered by double-strand breaks at the RFB sites and repair of the breaks via gene conversion. Thus, this *FOB1*-dependent recombination mechanism apparently contributes to the maintenance of average copy number of rDNA. However, in *Δfob1* cells, the repeats are still maintained without any fluctuation of copy number. This suggests that there is another, so far unidentified system acting to prevent contraction of the number of repeats.

In order to explore such a putative second system, we collected a number of mutants in which the rDNA copy number decreased drastically under *Δfob1* conditions. Among these, we found condensin gene mutants, suggesting that, in addition to condensation and separation of chromosomes in M phase, condensin plays an important role in maintaining the repeated structure of rDNA. Each of the genes encoding a condensin subunit is known to be essential for growth, but here isolated condensin mutations are not strict but are of the leaky type. Analyzing double mutants and examining specific interactions between condensin and rDNA regions revealed that (1) in the double mutants, the copy number of rDNA was dramatically decreased, (2) condensin complexes associated with the RFB region in a *FOB1*-dependent manner (Figure 1B and C), (3) the association between condensin and the RFB was established during S phase and was maintained until anaphase, (4) double mutants showed slow growth which may be caused by defects in the separation step of the long rDNA array in anaphase. These results strongly suggest that *FOB1*-dependent condensin association with the RFB region is required for efficient segregation of rDNA repeats.

Recently we found that an RNA polymerase I (PolI) defective mutation suppressed the dramatic reduction of

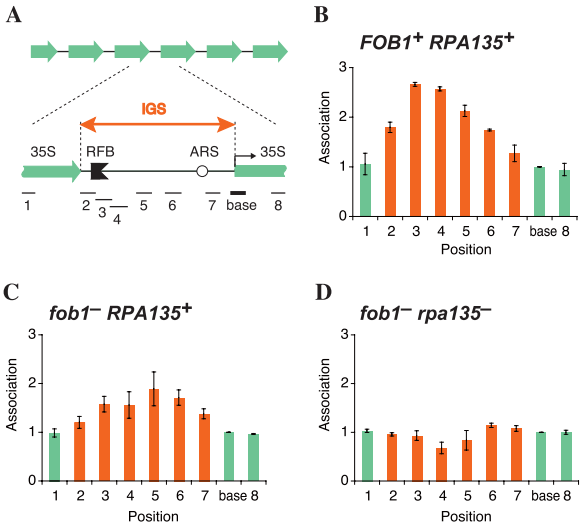

Figure1. Condensin relative association patterns with rDNA in a wild-type strain, *fob1*, and *fob1 rpa135* (PolI defective) mutants.

(A) The structures of the rDNA tandem array and an enlarged intergenic spacer sequences (IGS) between the 35S coding regions are shown. Within the IGS, the replication fork barrier (RFB) site and autonomous replication sequence (ARS) are located, respectively. The positions of the PCR fragments (1~8 and base fragment) used for chromatin immuno- precipitation (ChIP) assay are indicated as short bars under the map. Condensin association with rDNA in (B) wild-type strain, (C) *fob1* strain and (D) *fob1 rpa135* (PolI defective) double mutant.

the rDNA copy number in the condensin and *fob1* double mutant. Because PolI is an rDNA specific transcription enzyme, ΔpolI defective mutants are non-viable. However, if the defective cells carry a plasmid in which rDNA is inserted downstream of a Gal-dependent promoter, they become viable in the presence of galactose. Using this ΔpolI mutant, we examined effect of PolI enzyme on condensin association with rDNA. Under *fob1* conditions, Figure 1 C and D show that PolI enzyme seems to force condensin from transcribing to non-transcribing (IGS) region. This characteristic change of pattern suggests that constant PolI transcription throughout the cell cycle prevents condensin from associating with the transcribing region of rDNA. Thus, it would be expected that in the triple (*fob1*, condensin and PolI) mutant, the partially defective condensin which is uniformly associated with the rDNA region, would facilitate successful separation between the replicated long rDNA region in anaphase.

In higher eucaryotes, it is well known that M phase-specific repression of gene expression (mitotic repression) occurs, though the reason for this remains unknown. Our results suggest that mitotic repression allows condensin to associate uniformly with whole chromosomes, which ensures their successful condensation and subsequent separation. In any event, genetic and molecular analyses of highly specific chromosomal regions, such as the rDNA regions, provide useful data which help us to understand the nature of normal (non-rDNA) chromosomal regions..

II. *E. coli* with a linear genome

There are two types of chromosomal structure, linear and circular. While all chromosomes in eukaryotes are linear, those in almost but not all prokaryotes are circular. Thus, we can readily understand why almost all mitochondrial and chloroplast genomes are also circular, as both are inferred to be derived from some kind of bacteria. However, the reasons why the two main biological kingdoms have so distinctly different genome structures, or why there are, although extremely rare, exceptions (bacteria with linear genomes) remains unresolved. In order to address this, linearization of the circular genome of bacteria and comparison between circular and linear bacteria with identical genetic backgrounds seems a rational strategy.

We therefore attempted to linearize the circular genome of *Escherichia coli,* as there are in fact two types of exception bacteria with naturally linear genomes. One type, represented by the filamentous soil bacteria *Streptomyces* species, possesses a protein-designated ìt erminal proteinî (TP) that is covalently joined to the 5í ends of both termini of the genome. The second type, exemplified by the spirochete *Borrelia burgdorferi*, has telomeres with covalently closed hairpin structures at their termini. The ends of linear *Borrelia* chromosomes are similar to those of the linear *Borrelia* plasmid, the *E. coli* phage N15 and certain animal viruses such as the poxvirus. Especially N15 attracted our attention, because this phage is very similar to the λ phage in many aspects,

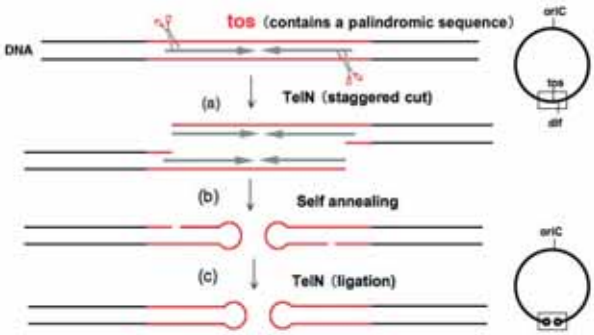

Figure2. Linearization of *E. coli* circular genome by TelN protein.

Locations of the *tos*, *oriC* and *dif* sites are shown on the *E. coli* circular genome in the right side figure. A palindrome sequence in the *tos* is represented by a pair of long arrows. (a) TelN cleavages the *tos* sequence at staggered positions, shown by two pairs of scissors. (b) Each single-stranded sequence exposed at the termini is self-annealed and (c) each gap left is sealed by TelN, producing a terminus with a hair-pin structure.

such as genome size, in having cohesive ends and so on. However, the lysogenization process is different; while λ phage lysogenizes by integration of the circular genome into the host chromosome, N15 lysogenizes by converting a circular genome into a linear one. For the linearization, the mechanism of which has been analyzed mainly by Russian investigators, two components - the *tos* site and protelomerase (TelN) protein - are required. Thus, in order to linearize the *E. coli* chromosome, we inserted the *tos* sequence into a replication termination region of the *E. coli* genome and then supplied TelN protein by introducing a plasmid with the telN gene or lysogenizing the N15 phage itself. In either case, the *E. coli* circular genome was confirmed to be linearized probably by the mechanism shown in Figure 2. Thus, *E. coli* with a linear genome is viable. In addition, we found that the linear state and whole genome structure are very stable, and that the linear genome clone is pure, probably not contaminated with circular genome cells. There were no appreciable differences between cells with linear and circular genomes in growth rates, cell and nucleoid morphologies, genome-wide gene expression (with a few exceptions), or DNA gyrase- and topoisomerase IV-dependency.

One structural difference to be expected is that the circular, but not the linear genome can form a dimer. With a circular chromosome, a dimer is formed through an odd number of recombination events between the sister chromosomes, but with linear genomes no dimer can be produced in this way even if sister-chromosomal recombination occurs. In *E. coli*, the dimer chromosome is monomerized by *dif* site-specific recombination catalyzed by *cer*-specific recombinase C and D (XerCD) and FtsK proteins. Thus, on a wild-type background, *dif* or *xerCD* mutants show slower growth, producing more elongated cells. We found that under *dif*-defective conditions, only cells with a circular genome but not cells with a linaer genome developed abnormal phenotypes.

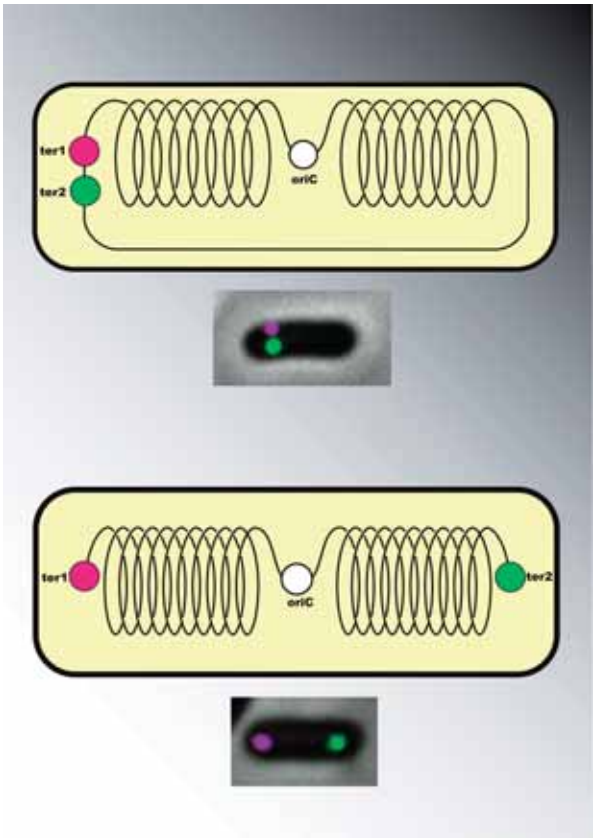

Figure3. Models showing locations of the two terminus sites (*ter1* and *ter2*) on the circular and the linear genomes in new born cells.

The *ter1* and *ter2* sites are located at ñ 20 kb and + 20 kb remote from the *tos* site.

In *Streptomyces* cells, each end of the linear genome was found to be associated with the other, probably through the two TPs. Thus, we tried to locate the position of each terminus of the *E. coli* linear genome in cells. To this end, we investigated the position of fluorescence foci of cyan and yellow fluorescent protein derivatives of the LacI and TerR repressors binding to their operator arrays inserted into two sites corresponding to the two termini of the linear genome. Microscopy indicated that while the two sites of the circular genome were close or overlapping in any cell cycle phase, those of the linear genome were separated and located at each end of a new born cell (Figure 3).

The termini of the linear gnome we constructed here are much closer to the *dif* site, the site directly opposite to the replication origin *oriC* site. If the ends are farther away from *dif*, does any phenotypic change occur? To examine this, the *tos* site was moved to five different positions on the chromosome and the growth of these cells was examined after N15 lysogenization (linearization). Those strains with genome termini the more remote from the *dif* site showed greater growth deficiencies and in an extreme case where the terminus was closest to *oriC* (approximate distance 500kb), lethality. This correlation might be caused by unbalanced replication of a pair of chromosome arms of different lengths.

From above results, the following conclusion can be drawn. (1) There were no obvious distinct differences between *E. coli* strains with circular or linear genomes. Therefore, there are at least two further conclusions from this: (a) presence of circular and linear genomes may not be inevitable, but accidental; and (b) *E. coli* with a linear genome as constructed here can survive in nature. (2) These findings raise the possibility that the topology of linear and circular genomes is not so different within the cell. (3) Two terminus ends of linear genomes separate and locate at terminus ends of the cell. Because the DNA molecule cannot condense itself, certain protein(s) must be involved. (4) The linearization system developed here could be a powerful technique in genome technology.

These results will be published in EMBO Reports (Cui et al., 2007, *8*, 181-187).

Publication List:

Original papers

Hayashi, K., Morooka, N., Yamamoto, Y., Fujita,K., Isono, K., Choi, S., Ohtsubo, E., Baba, T., Wanner, B.L., Mori H., and Horiuchi, T. (2006). Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. Mol. Systems Biol. doi:10.1038 msb100049:E1-E5

Inagaki, S., Suzuki, T., Ohto, M., Urawa, H., Horiuchi, T., Nakamura, K., and Morikami, A. (2006). *Arabidopsis* TEBICHI, with helicase and DNA polymerase domains, is required for regulated cell division and differentiation in meristems. Plant Cell *18*, 879-892

Johzuka, K., Terasawa, M., Ogawa, H., Ogawa,T., and Horiuchi, H. (2006). Condensin loaded onto the replication fork barrier site in ribosomal DNA (rDNA) repeats during S-phase in a *FOB1*-dependent fashion to prevent contraction of a long repeats in *S. cerevisiae*. Mol. Cell. Biol. *26*, 2226-2236.

Riley, M., Abe, T., Arnaud, M.B., Berlyn, M.K., Blattner, F.R., Chaudhuri, RR., Glasner, J.D., Horiuchi, T., Keseler, I.M., Kosuge1, T., Mori, H., Perna, N.T., Plunkett III, G., Rudd, K.E., Serres, M.H., Thomas, G.H., Thomson, N.R., Wishart, D., and Wanner, B.L. (2006). *Escherichia coli* K-12: a cooperatively developed annotation snapshot ñ 2005. Nucleic Acids Res. *34*, 1-9.