Gene, 32 (1984) 439–457
Elsevier

GENE 1152

The *ori* sequences of the mitochondrial genome of a wild-type yeast strain: number, location, orientation and structure

(*Saccharomyces cerevisiae*; DNA replication origins; *ars* sequences; petite mutation)

Miklos de Zamaroczy, Godeleine Faugeron-Fonty, Giuseppe Baldacci, Regina Goursot and Giorgio Bernardi

Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2, Place Jussieu, 75005 Paris (France) Tel. (1) 329-58-24; (1) 336-25-25, ext. 41.01

(Received April 30th, 1984)
(Revision received August 24th, 1984)
(Accepted September 4th, 1984).

SUMMARY

We have investigated the number, the location, the orientation and the structure of the seven *ori* sequences present in the mitochondrial genome of a wild-type strain, A, of *Saccharomyces cerevisiae*. These homologous sequences are formed by three G + C-rich clusters, A, B and C, and by four A + T-rich stretches. Two of the latter, *p* and *s*, are located between clusters A and B; one, *l*, between clusters B and C; and one *r*, either immediately follows cluster C (in *ori* 3–7), or is separated from it by an additional A + T-rich stretch, *r'*, (in *ori* 1 and *ori* 2). The most remarkable differences among *ori* sequences concern the presence of two additional G + C-rich clusters, β and γ, which are inserted in sequence *l* of *ori* 4 and 6 and in the middle of sequence *r* of *ori* 4, 6 and 7, respectively. Neglecting clusters β and γ and stretch *r'*, the length of *ori* sequences is 280 $\pm$ 1 bp, and that of the *l* stretch 200 $\pm$ 1 bp. Hairpin structures can be formed by the whole A-B region, by clusters β and γ, and (in *ori* 2–6) by a short AT sequence, *lp*, immediately preceding cluster β. An overall tertiary folding of *ori* sequences can be obtained. Some structural features of *ori* sequences are shared by the origins of replication of the heavy strands of the mitochondrial genomes of mammalian cells.

INTRODUCTION

The mitochondrial genomes of spontaneous cytoplasmic petite mutants of *S. cerevisiae* are exclusively made up of identical repeat units resulting from the tandem amplification of the DNA segment originally excised from the parental wild-type genome (Bernardi, 1979; Faugeron-Fonty et al., 1979). Therefore, each repeat unit of a petite genome is bound to contain an origin of replication, and, since petite genomes can arise from many regions of the wild-type genome, the latter should contain not only one, but several origins of replication (Prunell and Bernardi, 1977). Furthermore, since petite genomes contain a large number of origins of replication, because of the amplification of their repeat units, they should be very efficient in replication (de Zamaroczy et al., 1979; Goursot et al., 1980).

Interestingly, the conclusions just mentioned could not be reached from studies on EtBr-induced

Abbreviations: *ars*, autonomously replicating sequence; bp, base pairs; EtBr, ethidium bromide; kb, 1000 bp; ORF, open reading frame; *ori*, origin of replication.

petites, practically the only ones studied in other laboratories, simply because these genomes were poorly characterized, were often rearranged (Lewin et al., 1978), and were supposed to contain novel sequences as the result of extensive reshuffling of the excised segments (Gordon and Rabinowitz, 1973; Fauman and Rabinowitz, 1974; Gordon et al., 1974; Lazowska et al., 1974). As a consequence, the model in which every petite mutant DNA contained a wild-type replication origin was considered unnecessarily complex (Locker et al., 1979; Bos et al., 1980), and complicated schemes were proposed both for the mechanism of formation of petite genomes and for their replication (Borst et al., 1976). Concerning the latter, hypotheses ranged (Sanders et al., 1976) from the translocation of the supposedly single replication origin of the wild-type genome onto the nascent petite genome (within or without the amplified region), to the use of secondary origins, and to the involvement in replication of the unusual pyrimidine of unknown structure found in petite RDIA (Borst et al., 1976).

When the sequence of a repeat unit from a spontaneous petite, a-1/1R/Z1, was first determined (Gaillard and Bernardi, 1979), we looked for an origin of replication and noted three G + C-rich clusters (henceforth called GC clusters: A, B, C in Figs. 1 and 4). These sequences had been postulated to be present in replication origins (Prunell and Bernardi, 1977), because their structures suggested that they could be specifically recognized by proteins (Bernardi, 1968). Clusters A and B were also found, together with a 35-bp flanking stretch, in the repeat units of petites b and a *-1/7/8 (Fig. 1), which were derived from two different regions of the wild-type genome (de Zamaroczy et al., 1979; Goursot et al., 1980). Further investigations (Bernardi et al., 1980) showed that such homology was present in at least five regions (later expanded to seven) of the wild-type genome, and that it extended over at least 260 bp to include cluster C.

Three observations reinforced our working hypothesis that these sequences (which we called *ori* sequences *) were indeed the postulated origins of replications. The first one (de Zamaroczy et al., 1979; 1981; Goursot et al., 1980; Bernardi et al., 1980; Blanc and Dujon, 1980) was that crosses of highly suppressive petites with wild-type cells practically only produced diploids harboring the mitochondrial genome of the petite used in the cross; this indicated that the genome of these "supersuppressive" petites could compete with the genome of wild-type cells, in all likelihood because it contained multiple copies of the origin of replication. Incidentally, this observation eliminated a different explanation for the phenomenon of suppressivity based on destructive recombination between the wild type and the petite genome (Cohen et al., 1970; Michaelis et al., 1973; Deutsch et al., 1974; Perlman and Birky, 1974; Slonimsky and Lazowska, 1977). The second observation (Bernardi et al., 1980; de Zamaroczy et al., 1981) was that both the potential secondary structure formed by GC clusters A and B and the primary structure of GC cluster C were very similar to structures found in the bona fide replication origin of the heavy strand of HeLa cell mitochondrial DNA (Crews et al., 1979). The third observation was that *ori* sequences contained sites for the initiation of DNA transcription which were known to be overlapping or contiguous with *ori* sequences in other genomes and, in particular, in the mitochondrial genome of HeLa cells (Cantatore and Attardi, 1980). More recently, direct evidence has been obtained for RNA-primed bidirectional DNA replication starting at *ori* sequences (Baldacci and Bernardi, 1983; Baldacci et al., 1984).

Here we present investigations on the number, the location, the orientation and the structure of the *ori* sequences present in the mitochondrial genome of a wild-type strain, A. Preliminary reports on this work were published (Bernardi et al., 1980; 1983; de Zamaroczy et al., 1981; Bernardi, 1982a). In the following paper, we examine the *ori* sequences as found in the mitochondrial genomes of a number of wild-type strains (Faugeron-Fonty et al., 1984).

---

* The alternative use of *rep* to indicate active *ori* sequences (Blanc and Dujon, 1980), as found in supersuppressive petites, should not be followed for two reasons. The first one is that *ori* is a term generally used for origins of DNA replication (Szybalski et al., 1970; Hiraga, 1976), whereas *rep* is a term used for genes coding for proteins playing a role in DNA replication (see, for instance, Broach et al., 1982). The second reason is that structural and not functional criteria are normally used to define genes and regulatory sequences; the name of a gene does not change whether the gene is active or inactivated by a deletion or an insertion; the same criterion is used for regulatory sequences, like promoters.

MATERIALS AND METHODS

(a) Yeast strains

All petites used in this work were spontaneous cytoplasmic petites (Faugeron-Fonty et al., 1979; 1983; de Zamaroczy et al., 1981; Goursot et al., 1982; Marotta et al., 1982; Mangin et al., 1983), derived from wild-type *S. cerevisiae* strains D-243-2B-R1 and C-982-19d, also called A and B, respectively, in our laboratory (Bernardi et al., 1970). As in previous work, "wild type" refers here to the mitochondrial genome. Fig. 1 shows the restriction maps of the repeat units of the mitochondrial genomes of these petites. Culture conditions were as described by Faugeron-Fonty et al. (1979).

(b) Mitochondrial DNA

Mitochondrial DNA was purified by centrifugation in a CsCl density gradient, using a method modified from Lang et al. (1977). Restriction enzyme degradations, gel electrophoresis, nick-translation of DNA probes, transfer of DNA to nitrocellulose filters, and filter hybridization were performed essentially as described (Faugeron-Fonty et al., 1979).

The primary structure of DNA was determined according to the chemical method of Maxam and Gilbert (1977), using both 5′ and 3′ end-labeling. For details, see the legends to Figs. 1 and 4. Sequence comparisons were made using the computer program of Dumas and Ninio (1982). Cloning of the *Hae*III fragment carrying *ori*6 from the genome of

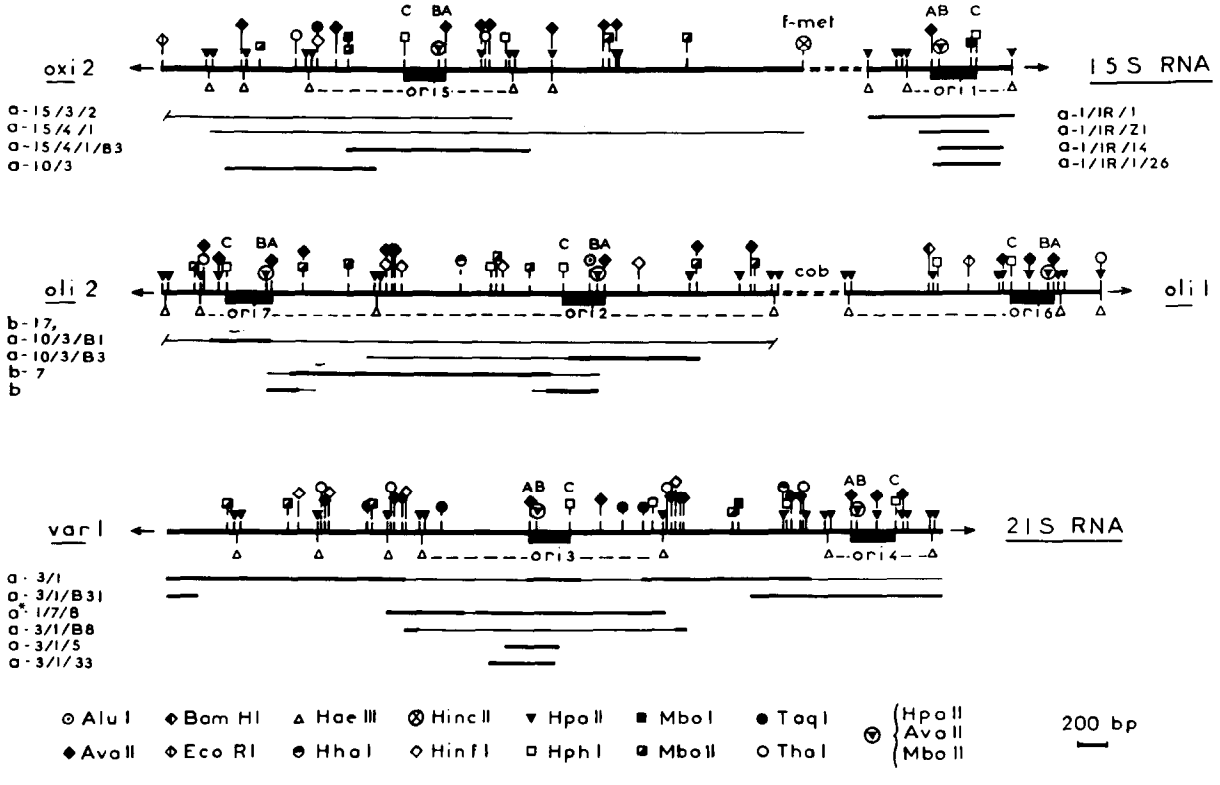

Fig. 1. Restriction maps of *ori* sequence regions in the mitochondrial genome of wild-type yeast. The left-to-right direction on the restriction maps corresponds to the clockwise direction on the genome map of Fig. 3, namely to the direction of transcription of most genes. Indicated are genes proximal to the genome segments shown, *ori* sequences and their corresponding *Hae*III fragments, and GC clusters A, B, C. Symbols for restriction sites are specified; for additional sites, see Fig. 4. Restriction maps were established on a series of petite genomes derived from wild-type strains A (for the *ori*5,-1, -3–4 regions), and B (for the *ori*2–7 region), or obtained from the sequence (Nobrega and Tzagoloff, 1980) of a petite, DS400/A12, derived from strain D273-10B (for the *ori*6 region). The repeat units of petite genomes, represented by horizontal lines, were largely or totally sequenced (Gaillard and Bernardi, 1979; Gaillard et al., 1980; Goursot et al., 1982; de Zamaroczy, M., and Bernardi, G., in preparation; sequenced regions are shown by thicker lines). The repeat units for petites a-15/3/2, a-15/4/1, b-17 and a-10/3/B1 are represented only in part; their lengths are 4200, 4450, 17000 and 4500 bp, respectively (Faugeron-Fonty et al., 1983; Marotta et al., 1982); break-off points are indicated by slashes. The order of the two *Hae*III fragments of petite a-1/1R/1 is not known.

petite a-3/1/5/B1 (de Zamaroczy et al., 1981) was done by oligo(C)-oligo(G) tailing in pBR322.

RESULTS

(a) The number of *ori* sequences

Screening of the mitochondrial genomes of many independent spontaneous petites derived from wild-type strain A revealed that the majority of them contained at least one *ori* sequence, as judged from diagnostic restriction sites located in GC clusters A, B and C (Fig. 1). The exceptions concerned *ori*$^-$, *ori*$^\circ$ and *ori*^{r} petites, in which *ori* sequences were partially deleted, absent, or rearranged; these cases have been studied elsewhere (de Zamaroczy et al., 1981; Goursot et al., 1982; Faugeron-Fonty et al., 1983; Mangin et al., 1983). Restriction mapping showed that the *ori* sequences found in different petites belonged in seven classes, showing different flanking sequences (Fig. 1). When petite genomes containing different *ori* sequences were used as ^{32}P-labeled *ori* probes and hybridized to *Hae*III digests of the mitochondrial genome of wild-type strain A, seven common hybridizing fragments were found (Fig. 2) having the same size as those carrying different *ori* sequences (Fig. 1). This experiment confirmed that the mitochondrial genome of strain A contained seven *ori* sequences, which were called *ori*1 to *ori*7 in the order in which we discovered them.

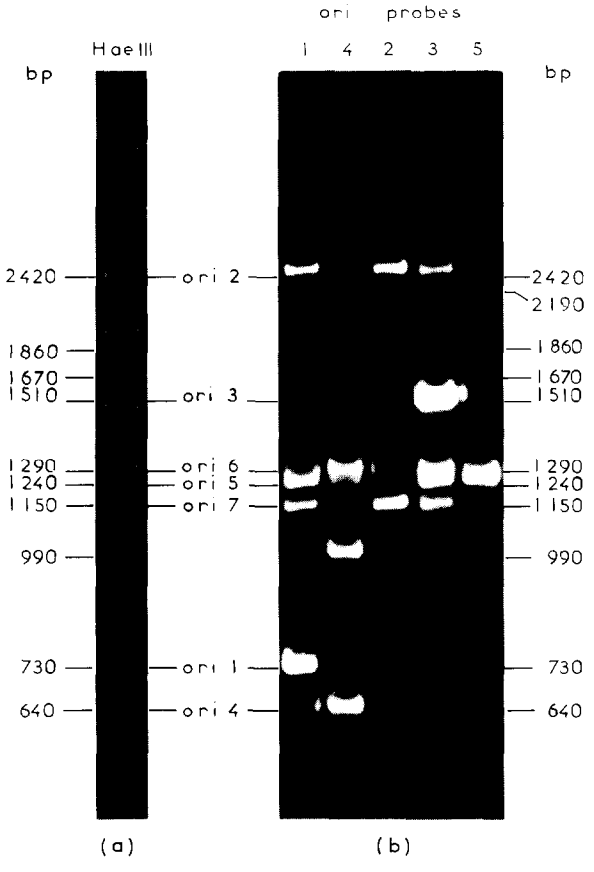

Fig. 2. Hybridization of ^{32}P labelled probes for *ori* 1, 2, 3, 4, 5 on *Hae*III restriction fragments from mitochondrial DNA of wild-type strain A (see RESULTS, section **a** and MATERIALS AND METHODS). (a) Electrophoretic pattern on 2% agarose gel of the *Hae*III digest stained with EtBr. The sizes (in bp) of the fragments carrying *ori* sequences are indicated. (b) Autoradiogram patterns of ^{32}P-labelled mitochondrial DNAs from petite mutants a-1/1R/Z1 (*ori*1), a-3/1/B31 (*ori*4), b (*ori*2); a-3/1/B8 (*ori*3), a-15/4/1/B3 (*ori*5) hybridized on *Hae*III fragments from strain A after transfer from agarose (a). The following should be noted. (i) The *ori*1 probe, consisting of *ori*1 and about 100 bp of flanking AT sequences, revealed all *ori* sequences (but only weakly *ori*3) except for the two containing cluster β, *ori*4 and 6. Likewise, *ori*3 hybridized on all *ori* sequences (but only very weakly on *ori*I), except for *ori*4 and 6. In contrast, the *ori*4 probe hybridized on the *ori* sequences containing cluster β and /or γ, namely *ore*4, 6 and7. The combined use of *ori*I and ori4 probes allows, therefore, the detection of all *ori* sequences and provides information on the presence of cluster β and γ. The other two probes showed more specific hybridization patterns apparently because the influence of flanking sequences. The *ori*5 probe practically only hybridized on itself. The *ori*2 probe hybridized on itself and also on *ori*7, with which it shared the flanking sequence preceding cluster A. It should be noted that the *ori*2 probe, (the genome of petite b), contains, in fact, a hybrid *ori*2-*ori*7 sequence, formed by *ori*2 except for part of the A-B region which derived from *ori*7; the A-B regions of *ori*2 and *ori*7 are, however, identical (Marotta et al., 1982; de Zamaroczy et al., 1983; and this work). (ii) Some other bands appear in addition to the seven hybridization bands due to *ori* sequences: (1) the 2190-bp band corresponds to the repeat unit of petite b-7, the product of very frequent excisions taking place between *ori*2 and *ori*7 (Marotta et al., 1982); this band is not visible in this figure; (2) the 990-bp band corresponds to the *Hae*III fragment separating the *Hae*III fragments carrying *ori*3 and *ori*4 (see Fig. 1); part of this fragment is present in the repeat units of petite genomes carrying *ori*3 or *ori*4; (3) the 1860 and 1670-bp bands, only produced with the DNA of petite a-3/1/B31, are due to nonspecific hybridizations of this probe, which contains not only the five GC clusters of *ori*4 (see Fig. 4), but also four additional GC clusters.

(b) The location and orientation of *ori* sequences

The *ori* sequences were mapped and oriented on the genome of strain A by hybridizing *ori* probes with

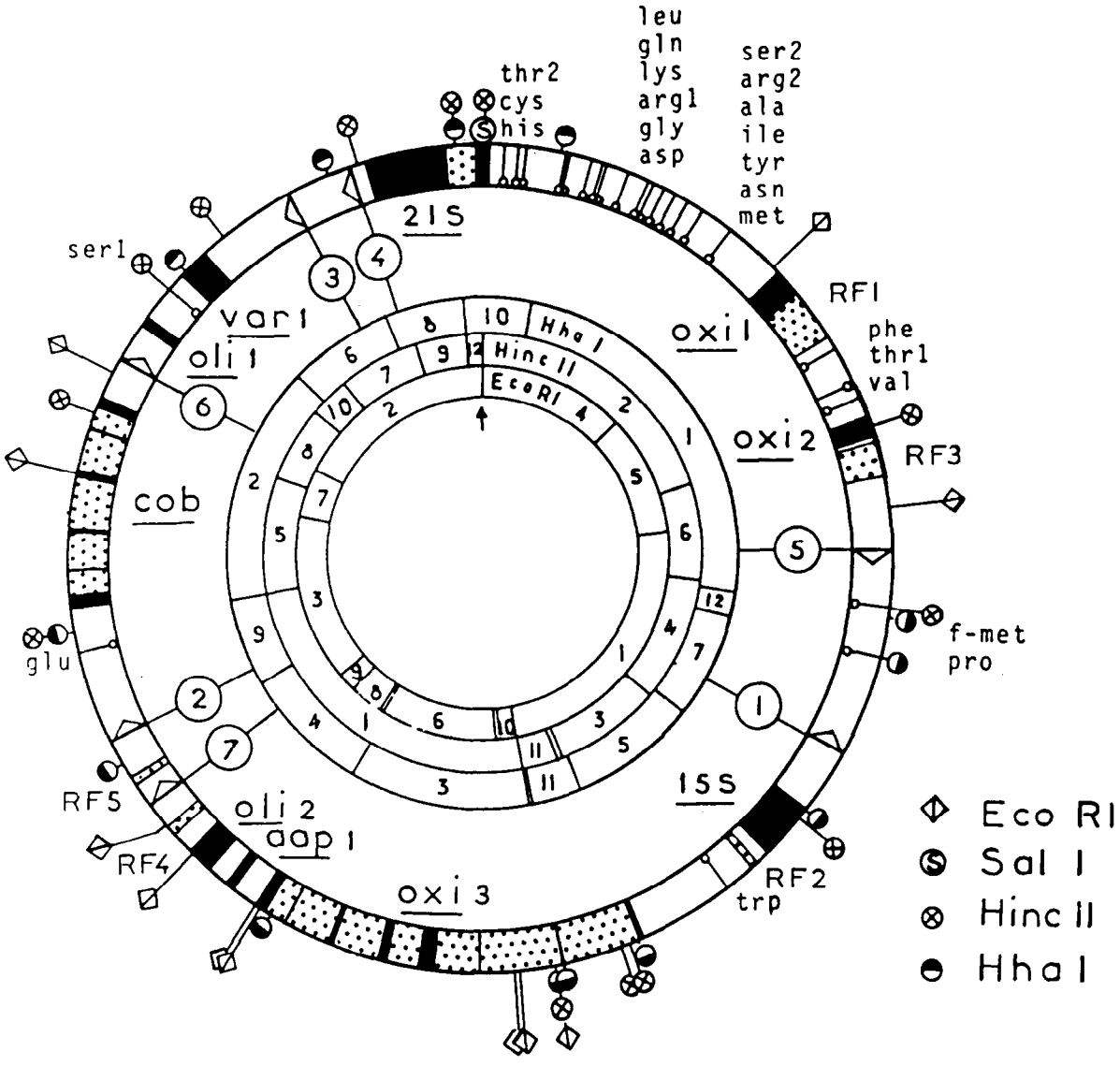

Fig. 3. Restriction maps of the mitochondrial genome of wild-type strain A and location of its seven *ori* sequences. *Eco*RI, *Sal*I, *Hinc*II, *Hha*I restriction sites were mapped on the mitochondrial genome of wild-type strain A. Numbers indicate restriction fragments in decreasing size order. The upward arrow corresponds to the *Sal*I site, used as the map origin. Circled numbers indicate the location of *ori* sequences (cluster C); open arrowheads point in the direction cluster C → cluster A. Black areas correspond to mitochondrial genes or their exons; radial lines indicate tRNA genes, (the *thr*l gene is not encoded by the same DNA strand as all other mitochondrial genes); dotted areas correspond to intervening sequences and intergenic open reading frames (RF1 to RF5; see Colin et al., 1984).

*Hha*I, *Hinc*II and *Eco*RI + *Sal*I fragments (Fig. 3) and with other shorter restriction fragments; *ori*5, 7, 2 and 6 are oriented with cluster C → cluster A in the clockwise direction on the standard map (Fig. 3; this is the direction of transcription of most genes); *ori*1, 3 and 4 have the opposite orientation.

(c) The primary structure of *ori* sequences

The primary structure of six *ori* sequences and their flanking regions (Fig. 4) were determined on the repeat units of 14 spontaneous petites. *Ori*1, -3, -4 and -5 were sequenced in petite genomes derived from strain A, *ori*2, -6 and -7 in petite genomes derived from strain B; the latter are located on *Hae*III fragments having the same size and the same position on the genome map in strains A and B (Faugeron-Fonty et al., 1984).

The *ori* sequences are presented in Fig. 4 using the same orientation of clusters A, B, C, regardless of their actual orientation on the genome map (Fig. 3). The GC content of *ori* sequences is 18.7%, on the

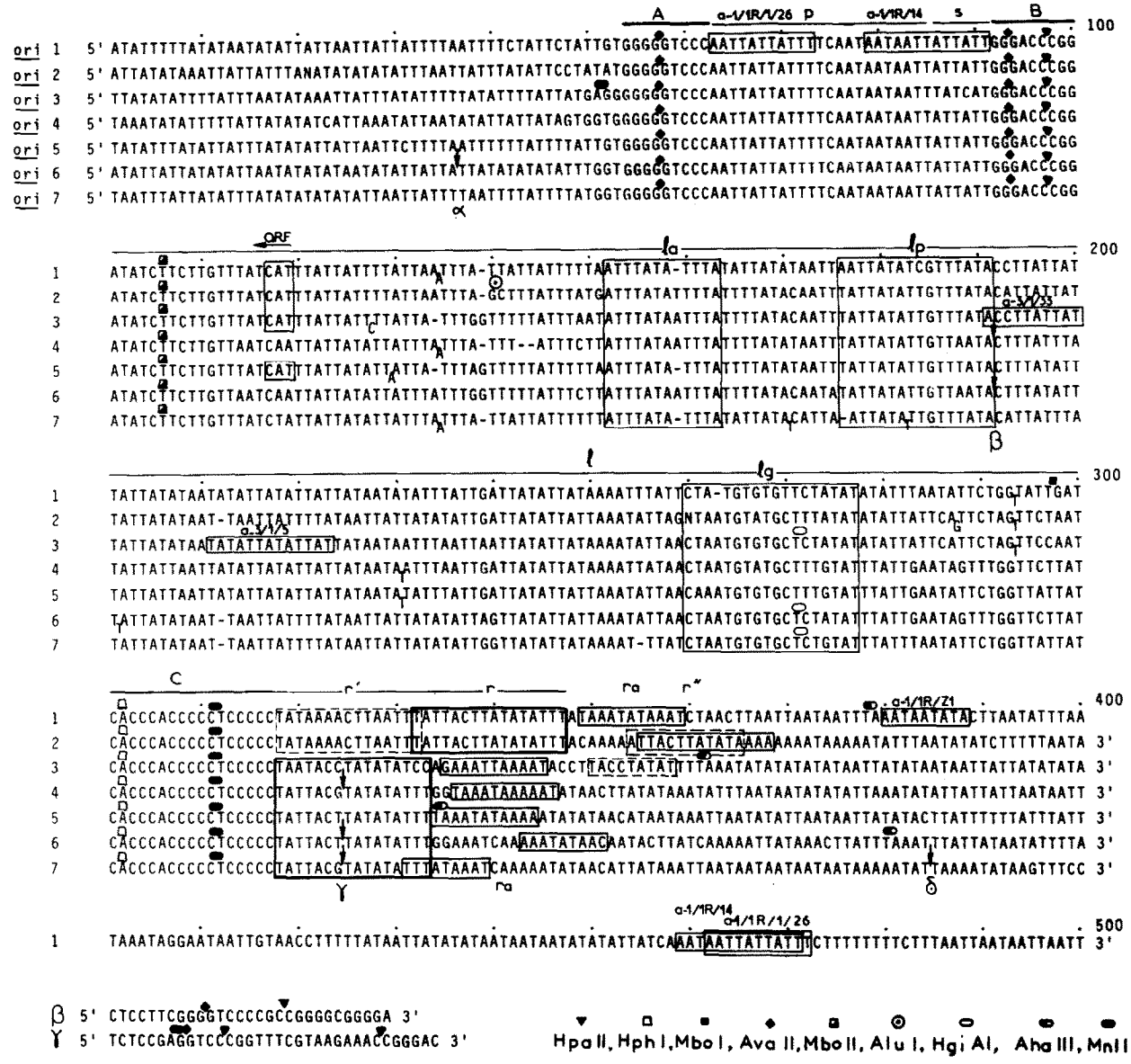

Fig. 4. Primary structure of *ori* sequences and their flanking regions. The *ori* sequences were determined for the repeat units of 14 different petite mutants: a-1/1R/1, a-1/1R/Z1, a-1/1R/14, a-1/1/R/1/26 (*ori*1); b, a-10/3/B3 (*ori*2); a*-1/7/8, a-3/1/5, a-3/1/33 (*ori*3); a-3/1/B31 (*ori*4); a-15/4/1/B3 (*ori*5); a-3/1/5/B1 (*ori*6); a-10/3/B1 (*ori*7). Site *ori*2 was originally sequenced on petite b (de Zamaroczy et al., 1981), since discovered to be a hybrid *ori* sequence (like that of petite b-7); for this reason the region preceding cluster A of *ori*2, as previously presented (de Zamaroczy et al., 1981), belonged in fact to *ori*7 (see also legend of Fig. 2). Regions of *ori* sequences are indicated by thick lines for GC clusters A, B and C and thin lines for AT stretches *p*, *s*, *l* and *r*. The positions of GC clusters *β* in *ori* 4 and 6, and *γ* in *ori* 4, 6 and 7 are given, as well as their sequences in *ori* 4 and 6 (bottom lines); *r* sequences are indicated by heavy line boxes; *r'* and *r"* sequences by broken line boxes. Other boxes indicate sequences *l*a, *l*p, *l*g, *r*a, the excision sequences of the repeat units of petites a-1/1R/1/26, a-1/1R/14, a-3/1/33, a-3/1/5, and a-1/1R/Z1 and the initiation triplet of the open reading frames of *ori* 1, 2, 3 and 5. The left flanking sequence of *ori*6 includes GC cluster *α* (see Nobrega and Tzagoloff, 1980), and the right flanking sequence of *ori*7 another GC cluster, *δ*, (see Macino and Tzagoloff, 1980); downward arrows indicate their locations. The primary structures of *α* and *δ* are given in Fig. 5. N indicates a modified C in sequence *l*g of *ori*2. Another modified nucleotide is found in the left flanking sequence of *ori*2.

average (neglecting additional clusters β and γ and sequence r'), about the same as the whole genome (Bernardi et al., 1970). They are characterized by extremely specific features, GC clusters and A + T-rich stretches (henceforth called AT stretches), which are described below (Figs. 4–6).

(*i*) GC clusters A, B and C are identical in all *ori* sequences; GC cluster β was only found in *ori*4 and 6, at exactly the same location and with an identical sequence; GC cluster γ was only found in *ori*4, 6 and 7, again at exactly the same location and with an identical sequence, except that three nucleotides were missing in the case of *ori*7 (Fig. 5); in contrast with the other GC clusters which had a GC content of 82–89%, the γ cluster only had a GC content of 59%: two additional GC clusters, α and δ, exist in the flanking sequences of *ori*6 and *ori*7, respectively, as present in the petite genomes studied by Nobrega and Tzagoloff (1980) and Macino and Tzagoloff (1980); their positions are indicated in Fig. 4 and their sequences are given in Fig. 5. GC cluster α is an *ori*^{s} sequence, as defined by Goursot et al. (1982). Restriction sites corresponding to clusters α and δ were also found in our petites carrying *ori*6 and -7.

(*ii*) The complementary palindromic AT stretch *p* following cluster A is identical in all *ori* sequences; the short AT sequence *s* preceding cluster B is identical in all *ori* sequences, except for *ori*3, where it presents four differences (Figs. 4 and 7).

(*iii*) The long AT sequence *l* bridging clusters B and C is a complex region characterized by a constant length (neglecting cluster β, if present): 200 bp in *ori*2, 5, 6; 201 bp in *ori*3; 199 bp in *ori*1, 4; and 198 in *ori*7. Region *l* is essentially made up of largely overlapping, short, direct and inverted repeats, the latter being complementary or not. These features are identical to those found in AT spacers of intergenic and intervening sequences (Bernardi and Bernardi, 1980; Bernardi 1982b; 1983).

The GC content of the *l* region is 10 ± 1%. The distribution of A and T on the two strands is asymmetrical, the ratio T/A being 1.7 ± 0.2 in the strand represented in Fig. 4; oligo(T) stretches are longer and more abundant than oligo(A) stretches on the same strand.

As shown in Fig. 6, 143 positions of *l* stretch are constant, being occupied by the same base in all *ori* sequences (126 positions), or at least in six of them (17 positions). Of the 57 variable positions, 46 are occupied by two nucleotides only; 34 of these positions are occupied by A or T, 9 by A or G and C or T and three present a deletion. The other 11 positions are occupied by three nucleotides; the two most frequent nucleotides are A or T in seven of these positions.

The first half of the *l* sequence (positions 101–197) comprises: (1) a stretch, between positions 101 and 151, which is very rich in T (T/A = 2.6 ± 0.2) and extremely conserved, particularly in its first 32 bases; this accounts for the fact that an 80-bp region encompassing clusters A–B, and this stretch was the first to be seen as an identical sequence at three different locations on the mitochondrial genome (de Zamaroczy et al., 1979); (2) a highly conserved

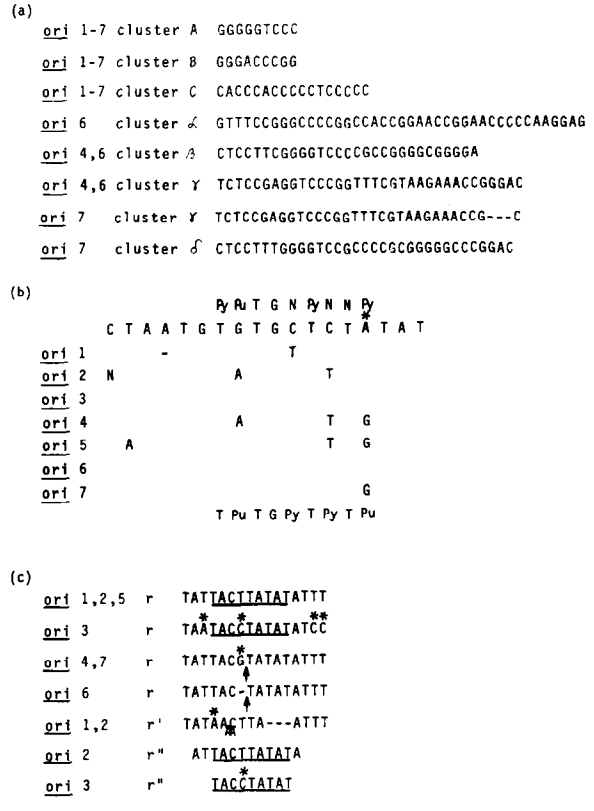

Fig. 5. Primary structure of specified regions in *ori* sequences. Base changes, deletions and insertions are indicated, by asterisks, dashes and subscripts, respectively. (a) GC clusters of *ori* sequences. (b) *lg* sequences; the two top lines give the consensus sequence of the cleavage site of *E. coli* DNA gyrase (Morrison et al., 1980), and the "majority" *lg* sequence; the bottom line gives the consensus sequence of the *lg* nonanucleotides; N indicates a modified C. (c) *r*, *r'* and *r"* stretches; arrows indicate the position of insertion of cluster γ; the nonanucleotide common to other transcription starts (Baldacci and Bernardi, 1982; Christianson and Rabinowitz, 1983) is underlined.

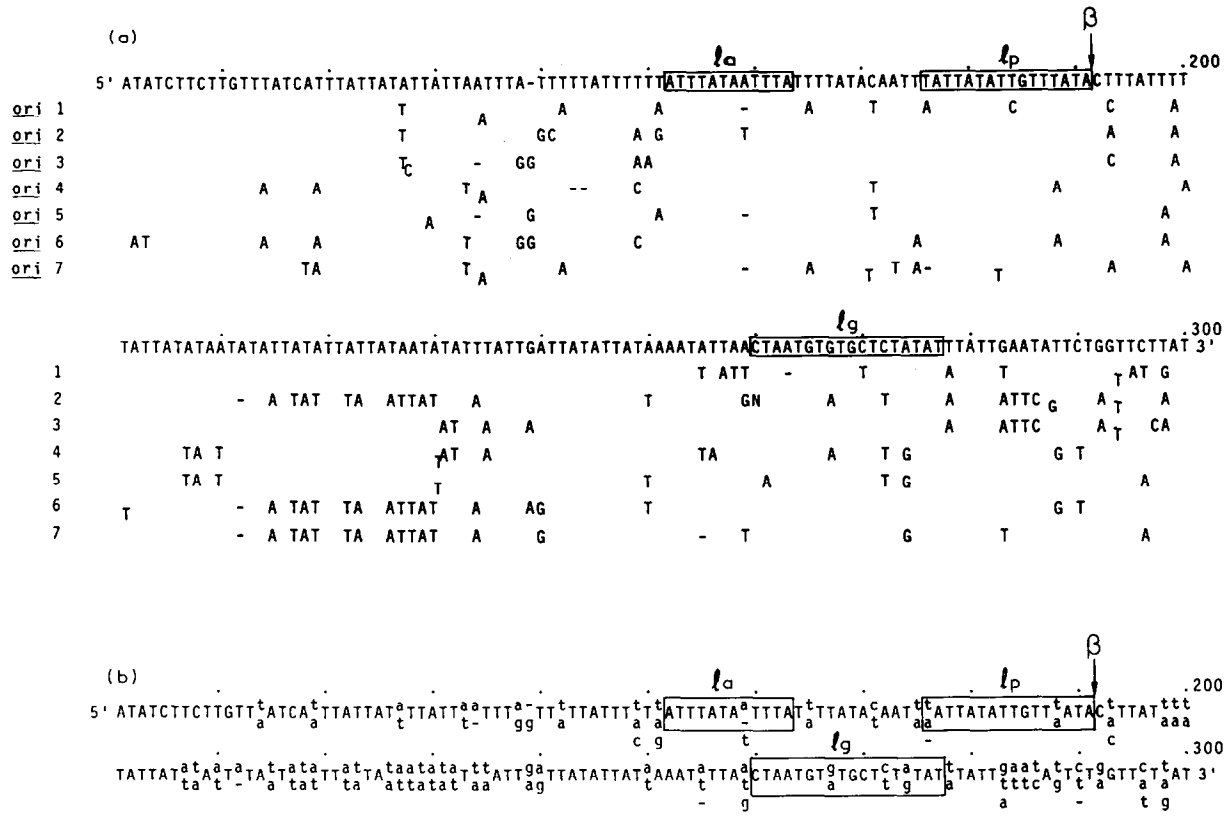

Fig. 6. AT stretch *l* of *ori* sequences. (a) The sequences in Fig. 4 are presented as a "majority sequence" (top lines); differences found in different *ori* sequences, sequences *l*a, *l*p, *l*g and the position of insertion of cluster *β* are indicated. Numbering is as in Fig. 5. The sequences between positions 200–230 of *ori* 2, 6 and 7 seem to be the result of inverting those present in *ori* 1, 3, 4 or 5; this might take place by making use (see Faugeron-Fonty et al., 1983) of inverted repeats present in *ori*1 (positions 201-206 and 224-230), *ori*3 (positions 200-207 and 224-231), *ori*4 (positions 200–206 and 225–231) or *ori*5 (positions 201–206 and 225–230). (b) The sequences in Fig. 4 are presented as a "consensus sequence". Capital letters indicate conserved nucleotides as present in all *ori* sequences (or in only six of them for 17 positions), lower case letters indicate variable nucleotides, the upper ones being those most frequently found. Deletions are indicated by dashes. Insertions are not shown.

24-bp sequence starting at position 152; its first 11-bp sequence, *l*a, shows a perfect match, in *ori*1, 5 and 7, with the *ars* site of the mitochondrial DNA of *Candida utilis* (Tikhomirova et al., 1983) and with one of the two *ars* sites of yeast sequence ARS2 (Tschumper and Carbon, 1982); the other *l*a sequences differ from these *ars* sites, ATTTATATTTA, by the addition, after position 7, of a T (in *ori*2) or an A (in *ori*3, 4 and 6); in all cases, the *l*a sequence was followed by a conserved base, T, also present in the *C. utilis ars* site; (3) a conserved 16-bp sequence, *l*p, immediately preceding the insertion position of cluster *β*; in *ori*2–6 this sequence can form a hairpin loop (see section **d**).

The second half of the *l* sequence (positions 198–300) contains a 50-bp highly conserved stretch (positions 233–282), which comprises an 18-bp sequence, *l*g, (Fig. 5), very largely made up of alternating purines and pyrimidines; this sequence contains a nonanucleotide (previously called cluster B′; Bernardi, 1982) which differs in its last position only from the consensus sequence of the cleavage site of DNA gyrase from *E. coli* (Morrison et al., 1980). It is of interest that even this more variable second half of sequence *l* contains long stretches which are identical in two or more *ori* sequences (see, for example, the 33-bp stretch between positions 198 and 230 of *ori*1, 3 and 5, or the 38-bp stretch between positions 201 and 238 of *ori*2, 6 and 7). Moreover, some of the sequence differences (positions

200–230) in the second half of the *l* stretch of *ori*2, 6 and 7 may be the result of local inversions (see legend to Fig. 6).

(*iv*) The AT stretch *r* (Fig. 5) is a 16-bp sequence which was originally identified because it contains a site of transcription initiation (Baldacci and Bernardi, 1982). Sequence *r* is perfectly conserved in *ori*1, 2, 5 and 6; in contrast, it presents a base change in *ori*3, 4 and 7, at position 7, the position which immediately precedes the insertion of cluster γ; in *ori*3, sequence *r* presents three additional base changes. Sequence *r* is located immediately after cluster C, except in *ori*1 and *ori*2, where it is separated from it by a 14-bp AT sequence, *r'*, showing some homology with *r* (Fig. 5). In *ori*2 and 3, sequence *r* is followed by a sequence which is a partial, perfect repeat of sequence *r* (Baldacci and Bernardi, 1982) and contains a transcription initiation site. This sequence, *r''*, is 12 bp long in *ori*2, and 9 bp long in *ori*3 (Figs. 4 and 5). Additional, similar, but shorter repeats (not indicated in Fig. 4) are found in the region following sequence *r* in the other *ori* sequences, except for *ori*7. Finally, as in the case of cluster β, the insertion of cluster γ takes place within a very highly conserved region.

(*v*) The A + T-rich sequence *r*a occupies a region following the *r* sequence, with which it shows some overlap in the case of *ori*7. In *ori*1, this sequence is identical (Baldacci and Bernardi, 1983) to the *ars* sites of *C. utilis* and yeast *ARS2*, already mentioned; in *ori*5, *r*a is identical with the *ars* site of yeast HMR left (Broach et al., 1982). Sequence *r*a is the perfect (in *ori*1) or partial (in the other *ori* sequences) inverted complement of sequence *l*a (a palindrome in *ori*1, 5, 7, and is followed by an A-rich region about 30 bp long. It should be noted that patchy complementary inverted homology is found between the flanking sequences of *l*a and *r*a, particularly on the side of clusters C and β.

(*vi*) The regions preceding cluster A exhibit, at approximately the same positions, 0–50, sequences which are the direct repeats of sequences following cluster B (positions 120–180; these repeats are not identical in different *ori* sequences). The regions following sequence *r*a contain a number of stretches which are identical in different *ori* sequences, although they are not located at exactly the same positions; two such sequences are the 21-bp sequences shared by *ori*5 and *ori*7 (positions 345–367).

If 200-bp stretches flanking the *ori* sequences are examined, one can notice a difference in the T/A ratio, which is (on the strands of Fig. 4 and neglecting GC clusters) 1.4 ± 0.2 in the sequences preceding cluster A, and 0.9 ± 0.2 in those following cluster C; the latter ratio is that found, on the average, in intergenic sequences.

(d) Potential secondary structures in *ori* sequences

As previously observed (Gaillard and Bernardi, 1979), the complementary inverted repeats of clusters A and B can be folded in a hairpin structure (with a side loop corresponding to sequence *s*) comprising the whole A–B cluster region (Bernardi et al., 1980; de Zamaroczy et al., 1981). The stem-and-loop structure of this 'A-B fold' (Fig. 7) is perfectly conserved in all *ori* sequences.

A case of special interest is that of the *ori*1 sequences present in the genomes of two petites, a-1/1R/1/26 and a-1/1R/14 (Bernardi et al., 1980; de Zamaroczy et al., 1981; see Fig. 1). In the former, cluster A and part of the AT stretch *p* were deleted during the excision process; in the latter, the excision also involved part of AT stretch *s* (de Zamaroczy et al., 1983). The positions of the excision sequences are indicated in Fig. 4. In both cases, an identical "replacement A–B fold" can be formed using the 11-nucleotide excision sequence and the preceding 23 nucleotides in the case of a-1/1R/1/26, or the 13-nucleotide excision sequence and the preceding 20 nucleotides in the case of a-1/1R/14 (Fig. 7). This stem-and-loop structure is shorter in size and has a shorter side loop compared to the A–B fold of intact *ori* sequence. Interestingly, these replacement A–B folds, located only 26 (in a-1/1R/1/26) or 7 (in a-1/1R/14) nucleotides upstream of the normal A–B fold, comprise nucleotides present on subsequent repeat units of the petite genomes; in other words, the "restored" *ori*1 sequences are hybrid.

Both clusters β and γ may be folded as shown in Fig. 7. Another secondary structure deserving mention here is the small inverted palindrome *l*p immediately preceding cluster β (Fig. 7). This comprises a stem formed by 5 bp in *ori*4 and 6, or by 4 bp in *ori*2, 3 and 5; in the case of *ori*1 and *ori*7, two and three mismatches, respectively, in the stem make this structure most unlikely.

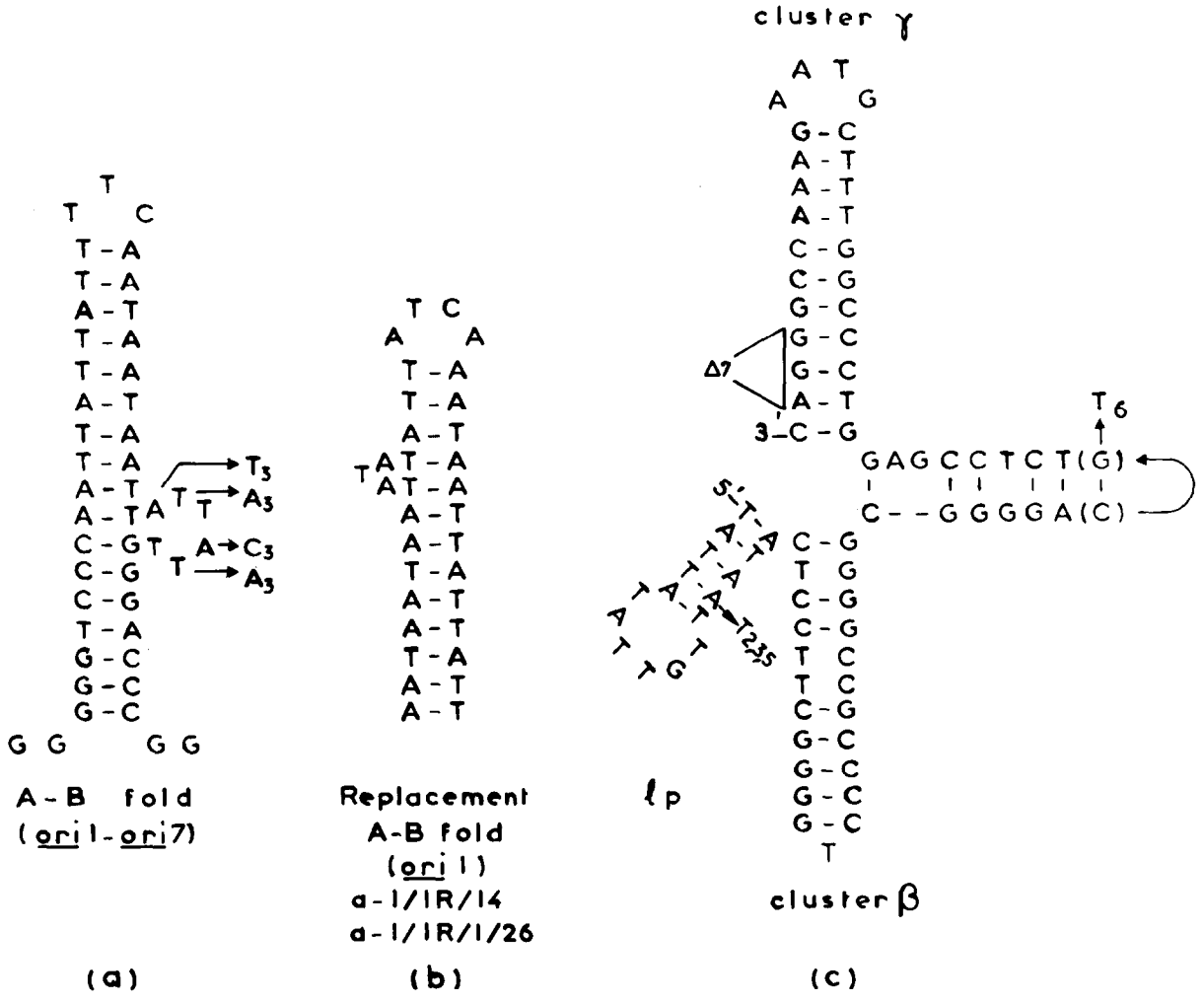

Fig. 7. Potential secondary structures in *ori* sequences. (a) A-B fold (hairpin) region of *ori* 1-7. (b) Replacement A-B fold (hairpin) as found in *ori*1 of petites a-1/1R/14 and a-1/1R/1/26. (c) *lp* sequence, β and γ clusters; the arrangement shown is that for the three secondary structures in *ori*4 and *ori*6 (see DISCUSSION, section i, and Fig. 10). Base changes (arrows) and deletions (Δ) are present in some *ori* sequences (indicated by subscripts).

DISCUSSION

(a) Number, localization and orientation of *ori* sequences

These issues do not deserve any additional comment, except to say that strain A is different from most other laboratory strains in having a canonical *ori*1 sequence, instead of the highly rearranged *ori*1 * sequence, and in apparently lacking another *ori* sequence, *ori*8 (Faugeron-Fonty et al., 1984).

(b) General properties of *ori* sequences

The most striking feature of the primary structure of the seven *ori* sequences of the mitochondrial genome of wild-type yeast studied here is its extreme conservation. This is perfect for GC clusters A, B, C, β and γ and for AT stretches *p* and *s*, except for a 3-bp deletion in cluster γ of *ori*7, and for four changes in sequence *s* of *ori*3. In contrast, differences are found for AT-stretch *l* which shows an overall homology of 70% in different *ori* sequences. These differences provide, however, evidence for striking sequence constraints. As far as deletions and additions are concerned, they almost exactly compensate each other, accounting for the extreme size constancy ($\pm$ 0.5%) of the region in different *ori* sequences. The majority of substitutions are transversions which almost only involve, at about the same extent, A $\rightarrow$ T and T $\rightarrow$ A changes. The predominance of trans-

versions over transitions, by a factor higher than four, is in striking contrast with what is generally found. For example, 51 out of 53 base changes detected in *ori* sequences of the heavy strands of human mitochondrial genomes belonging to different individuals were transitions and only two were transversions (Greenberg et al., 1982).

(c) The *l*a and *r*a sequences

The *l*a sequences of *ori*1, 5 and 7 and the *r*a sequences of *ori*1 and 5 are identical to known *ars* sites and are embedded in A + T-rich stretches. They exhibit, therefore, the structural features required for autonomous DNA replication in yeast nuclei (Broach et al., 1982; Kearsey, 1983). This raises the question whether the mitochondrial and nuclear DNA replication systems of yeast really share structural features, an idea of some merit in view of the claim of shared subunits in nuclear and mitochondrial DNA polymerase (Scovassi et al., 1982). Sequences identical to known *ars* sites do not appear, however, to be required for mitochondrial DNA replication since they only exist in two out of the four *ori* sequences of the mitochondrial genome of strain A, which have been shown to be replicationally active in petite genomes, namely *ori*1 and *ori*5 (the other two such sequences are *ori*2 and *ori*3). The possibility remains open, however, that the other *l*a sequences and at least some of the *r*a sequences correspond to *ars* sites not yet found.

Another question is whether sequences identical to *ars* sites do or do not play a surrogate role in mitochondrial DNA replication. To answer this question, a computer search for ten *ars* sites (one from the mitochondrial DNA of *C. utilis*: Tikhomirova et al., 1983; nine from yeast nuclear DNA: Tschumper and Carbon, 1980; 1981; 1982; Stinchcomb et al., 1981; Broach et al., 1982; Kearsey, 1983) was done over 40 kb of mitochondrial DNA including *ori* sequences, intergenic sequences, closed reading frames from intervening sequences, the *var*1 gene, the 15S RNA gene, and part of the 21S RNA gene. The *C. utilis ars* site was found to be the most frequent site (32 copies), followed by those that differed the least from it, namely HMR right (21 copies), HMR left (15 copies), HO (5 copies), and HML right (4 copies). Both orientations were about equally represented. Only one copy of *ars* sites containing G and/or C was found (that of ARS2a). Among the five sequenced *ori*° petite genomes (Goursot et al., 1982), four showed several or only one (in the case of petite No. 4) *ars* sites, and one (petite No. 3) none. At least in the latter case, none of the *ars* sites explored is required for replication, but again the possibility remains open that one (or more) *ars* sequence(s) different from those explored is (are) present in the genome of petite No. 3. In any case, *ars* sites appear to be involved in the replication in the yeast nucleus of plasmids carrying *ori*° segments of mitochondrial DNA, such as that studied by Hyman et al. (1982).

(d) Sequence *l*g

As far as sequence *l*g is concerned, several points should be made. (i) Three of the ten transitions not accompanied by transversions of region *l* were the only changes found in different *ori* sequences in the nonanucleotides matching the cleavage site of *E. coli* DNA gyrase (Fig. 5). This suggests that an important structural requirement had to be preserved in the sequence. Along the same line, the consensus of the nonanucleotides from the seven *ori* sequences, T Pu T G Py T Py T Pu, is a strictly defined sequence, much more so than that of the cleavage site of *E. coli* DNA gyrase, Py Pu T G N Py N N Py. (ii) If the *l*g nonanucleotide corresponds to the cleavage site of a mitochondrial gyrase, this site is located within the *ori* sequence and not outside of it, as in the case of prokaryotic genomes (Gellert et al., 1979; Morrison and Cozzarelli, 1979; 1981). (iii) A computer search for *l*g sequences and for nonanucleotides matching the gyrase sequence over the 40 kb of sequences of yeast mitochondrial DNA (see section **c**, above) revealed no *l*g sequence and only one nonanucleotide (after the *cob* gene) reinforcing the idea that the latter is a highly specific sequence.

(e) Sequences *r*, *r*′, *r*″

Sequence *r* was originally shown (Baldacci and Bernardi, 1982) to be a site of transcription initiation in petite genomes carrying *ori*1, and to share a decanucleotide with the 17-bp transcription initiation sites of both mitochondrial rRNAs of *C. cerevisiae* and of the small rRNA of *Kluyveromyces lactis* (Osinga and Tabak, 1982; Christianson et al., 1982).

Subsequent work (Christianson and Rabinowitz, 1983) showed that, in fact, an octa- or nonanucleotide is shared by all transcription initiation sites of yeast mitochondrial genome identified so far.

As far as *ori* sequences are concerned, hybridization experiments with separated DNA strands have identified the template strand used in transcription as the strand containing the oligopyrimidine stretch of cluster C; and S1-nuclease degradation of RNA-DNA hybrids has indicated that transcription initiates at sequence *r* and proceeds in the cluster C → cluster A direction (Baldacci and Bernardi, 1982). More recent work (Baldacci and Bernardi, 1983; Baldacci et al., 1984) has shown that sequence *r* is the initiation site not only for long transcripts in petite genomes (which are absent in wild-type genomes), but also for short RNA primers; in contrast, nascent DNA chains copied on the "non-*r* strand" start within cluster C, following an RNA primer starting in sequence *l* just before cluster C, and are elongated towards sequence *r*. In contrast with the primers copied on the *r* strand, those copied on the "non-*r*" strand begin just after a nonconserved sequence located at the end of sequence *l* (see Fig. 6).

It should be noted that the T → C change in the central nonanucleotide of sequence *r* from *ori*3 has been found in one site of transcription initiation (the site for $tRNA^{glu}$ and *cob* genes; Christianson et al., 1983). In contrast with the central nonanucleotide of *r* sequence from *ori*6, those from *ori*4 and 7 do not correspond to any known transcription initiation site. The base changes found in these *r* sequences are likely to have followed the insertion of cluster *γ* inactivating the transcription initiation site.

Sequences *r′* and *r″* appear to be the result of duplications of sequence *r* (Baldacci and Bernardi, 1982). While *r′* is rather seriously rearranged , *r″* sequences are perfect, though shorter copies of the corresponding *r* sequence, and comprise the nonanucleotides of mitochondrial transcription starts. This raises two problems, namely whether the *ori* sequences carrying two tandem transcription initiation sites, *r* and *r″* (*ori*2 and *ori*3), play a more important role in the replication of the wild-type genome than the other two, and which sequence is actually used to start transcription. While no information is available on the first point, evidence has been obtained that transcription starts at sequence *r* and not at sequence *r″*, at least in case of *ori*2 (Colin et al., 1984).

(f) Potential coding regions in *ori* sequences

A very short ORF starting with an AUG triplet at position 119 (Fig. 4) and ending at position 89 with two tandem ochre codons corresponding to sequence *s*, is found in *ori*2. In the case of *ori*3, the four base changes in sequence *s* are responsible for the absence of the two ochre codons, and the open reading frame continues until position 38. The four *ori* sequences of the mitochondrial genome of strain A carrying these ORFs are those shown to be active in replication. The direction of transcription of this ORF is cluster C → cluster A and the template strand is that containing the oligopyrimidine of cluster C; these are properties already found in transcripts starting at the *r* sequence of *ori* sequences (Baldacci and Bernardi, 1982), which, in fact, cover the ORF under consideration here. The question remains open as to whether this short ORF, essentially corresponding to cluster B, is associated with the production of a small polypeptide, which would be 10 amino acids long in the case of *ori*1, 2, 5, and 27 amino acids long in the case of *ori*3. Both sequences show a codon usage similar to that found in other ORFs (Coruzzi et al., 1981) and in *var*1 (Hudspeth et al., 1982). The 10 amino acid sequence would be fmet-met-asn-lys-lys-met-ser-gly-ser-glu. Obviously, this could only be produced in wild-type cells where mitochondrial protein synthesis is active.

Other open reading frames exist in *ori* sequences. However, those beginning with an AUG triplet do not coincide in position in different *ori* sequences (with one exception located at positions 263–265); moreover, they would be transcribed from the strand containing the oligopurine sequence of cluster C in the direction cluster A → cluster C. No stable transcripts with these properties have been detected in petite genomes (Baldacci and Bernardi, 1982). Finally, a number of ORFs beginning with an AUA triplet also exist on both strands.

(g) The secondary structure of *ori* sequences

Several findings suggest a real existence for the "A–B fold". First of all, not only the A and B clusters, but also the AT stretch *p*, namely all the sequences required to form a stable stem, are perfectly conserved in primary structure; in contrast, sequence *s*, forming the side loop, can undergo

changes. Second, when a deletion removes cluster A and the following nucleotides, as in the *ori*1 sequences of petites a-1/1R/14 and a-1/1R/1/26, an identical replacement loop, with features and location similar to the A–B fold can be formed. This leads, however, to a decrease in replication efficiency as judged from suppressivity; this is 80% for a-1/1R/14 and a-1/1R/1/26, vs. >95% for a-1/1R/Z1, a petite carrying an intact *ori*1 sequence in an otherwise identical repeat unit (Figs. 1 and 4). Third, structures similar to the "A–B fold" are found in all mammalian mitochondrial *ori* sequences investigated so far (see section **h**, below).

The secondary structures proposed for clusters β and γ and for the AT stretch *lp* will be discussed below in section **i**.

(h) A comparison of mitochondrial *ori* sequences

An obvious question raised by the primary and secondary structural features of the *ori* sequences of the mitochondrial genome of yeast concerns their evolutionary conservation. A number of similarities were found between structures found in the *ori* sequences of the mitochondrial genome of yeast and those present on the heavy strands (but not on the light strands) of the mitochondrial genomes of mammalian cells.

(i) Secondary structures reminiscent of the A–B fold in their stem size, A:T/G:C pairs, and position of the side loop (Fig. 8), exist in the *ori* sequences of the heavy strand of mitochondrial DNAs from man (Crews et al., 1979; Anderson et al., 1981), rat (Kobayashi et al., 1980; Sekiya et al., 1980), mouse (Bibb et al., 1981) and calf (Anderson et al., 1982) and also in the putative *ori* sequence of *Aspergillus amstelodami* (Lazarus and Küntzel, 1981). In no case are these similarities in secondary structures accompanied by similarities in primary structures.

(ii) A nonanucleotide very similar to that of sequence *lg* is found in the heavy strand *ori* sequences of mammalian mitochondrial genomes (Fig. 9a). This sequence contains the conserved sequence CSB-1 (Wong et al., 1983).

(iii) Several polypyrimidine stretches similar to that of cluster C are found on the heavy strand *ori*

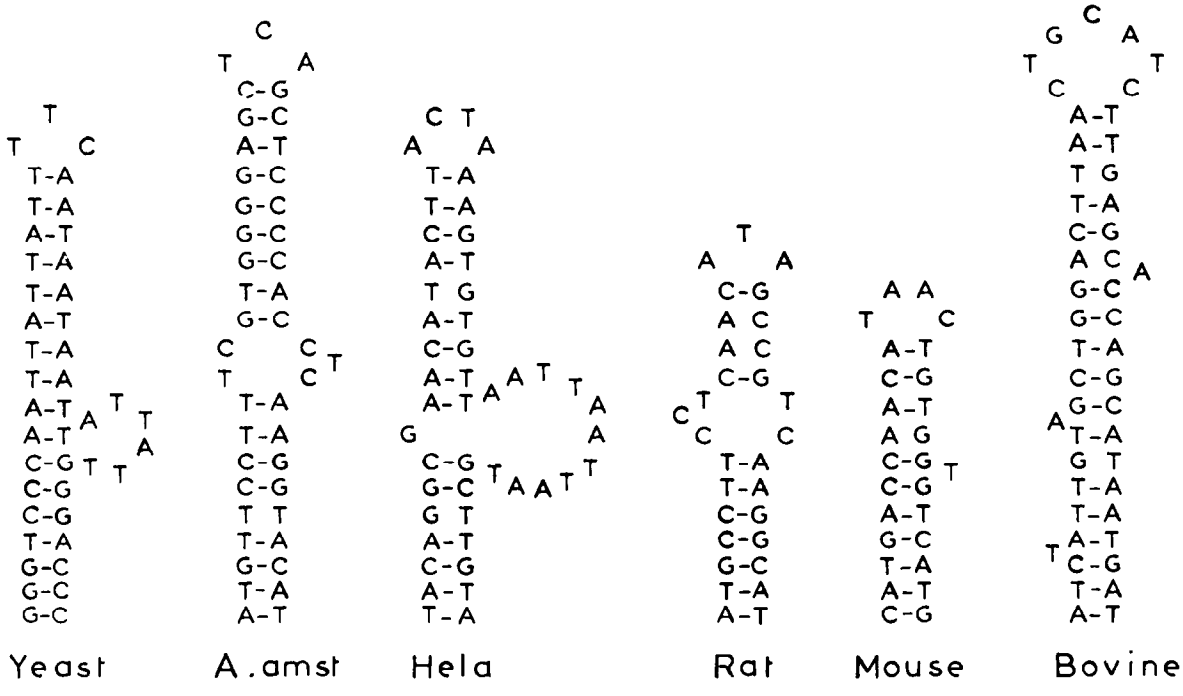

Fig. 8. Comparison of the secondary structure of the A-B hairpin fold of the mitochondrial *ori* sequences of yeast and of similar structures found in other mitochondrial *ori* sequences (heavy strand origins in the case of mammalian cells). See also DISCUSSION, section **h**. A. amst. stands for *A. amstelodami*.

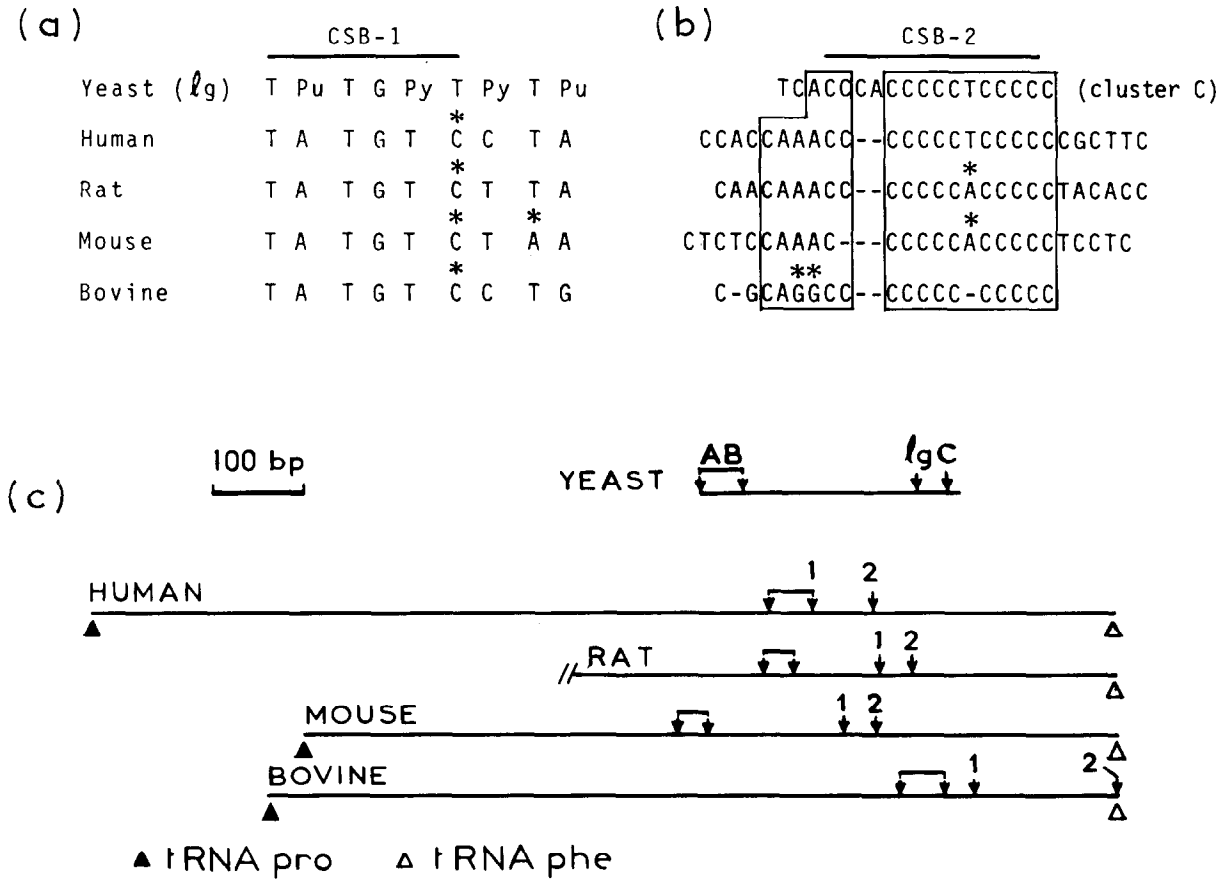

Fig. 9. Comparison of primary structures of yeast mitochondrial *ori* sequences with those found in mammalian mitochondrial *ori* sequences (heavy strand origins). Asterisks indicate differences. (a) The consensus sequence of the nonanucleotide present in sequence *lg* of yeast *ori* sequences is compared with the nonanucleotide comprising sequence CSB-1 (overlined) as present in mammalian *ori* sequences. (b) Cluster C of yeast *ori* sequences is compared with sequence CSB-2 of mammalian *ori* sequences. (c) The positions of A–B fold (or similar hairpin folds), the *lg* nonanucleotide (or sequence CSB-1), and cluster C (or sequence CSB-2) on the *ori* sequences of the mitochondrial genomes of yeast and mammalian cells (heavy strands) are compared; sequences CSB-1 and CSB-2 are indicated by 1 and 2, respectively; the position of tRNA pro and tRNA phe genes which delimit the *ori* region is shown.

sequences of mammalian mitochondrial genomes (Fig. 9b). One of them has been called sequence CSB-2 (Wong et al., 1983).

(iv) A comparison of the *ori* sequences from the mitochondrial genomes of yeast and mammalian cells revealed (Fig. 9c): a conservation of (1) the relative positions of the A–B fold, the *lg* nonanucleotide or CSB-1 and of cluster C or CSB-2; and (2) to some extent, the distance between the end structures, the A–B fold and the C cluster or CSB-2. This distance of 200 bp in yeast is 187 bp in mouse and 183 bp in calf; in man and rat, this distance is shorter, 76 bp and 134 bp, respectively; other polypyrimidine-polypurine clusters are, however, present in these genomes farther away from the A–B fold.

(i) A hypothetical tertiary structure for the *ori* sequences

The complementary inverted homology between sequences *la* and *ra* and the patchy homology between their flanking sequences raises the question as to whether this homology level is high enough as to cause base-pairing interactions between these two *ori* sequence regions. The answer to this question should be positive because the homology levels are certainly comparable with those found in sequences used in the excision of petite genomes (de Zamaroczy et al., 1983), a process involving base-pairing interactions and recombination.

Interactions between the first half of the *l* region

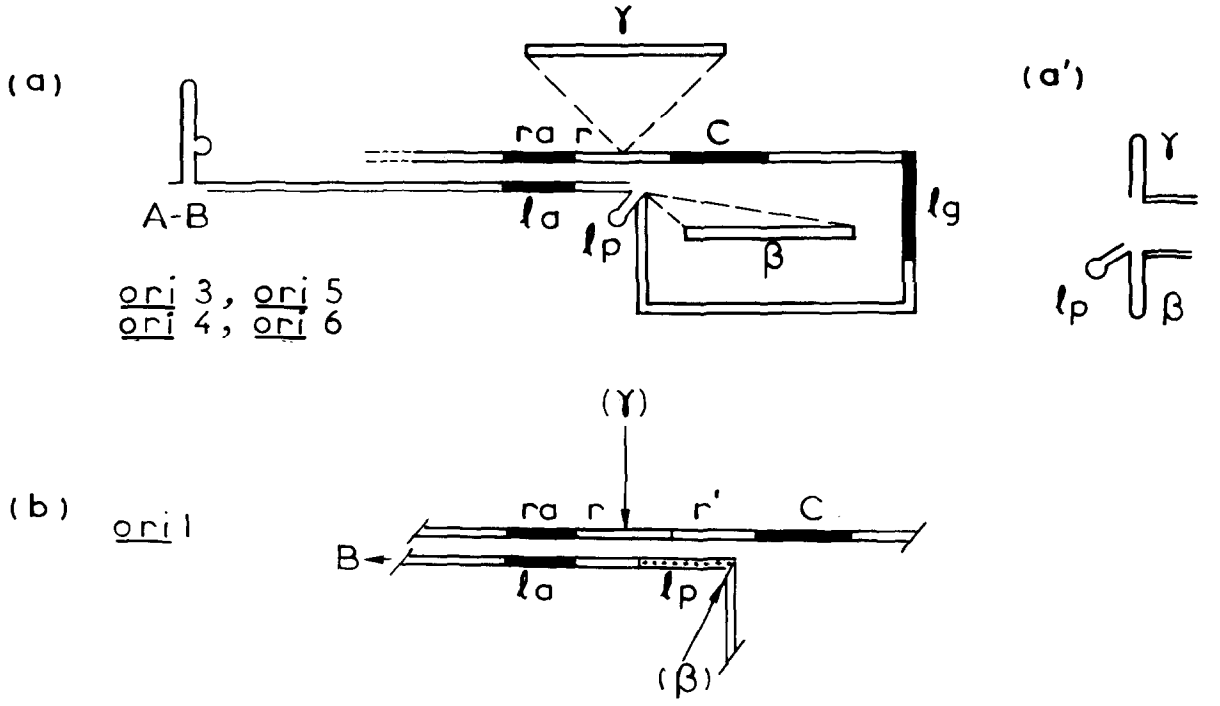

Fig. 10. Hypothetical superfolding of *ori* sequences from the mitochondrial genome of yeast (see also DISCUSSION, section i). (a, a') Superfolding of *ori*3 and -5; base-pairing interactions take place between sequences *l*a and *r*a as well as between flanking sequences; *l*p sequence is supposed to loop out. In *ori*4 and -6, the size of the loop (132 bp) between sequences *l*p and C would be considerably extended (by 70 bp) because of the insertion of clusters β and γ. This does not occur, however, if these clusters take the configuration shown in (a'); see Fig. 7 for more details. (b) Folding of *ori*1. In the case of *ori*1, *l*p cannot fold upon itself, but can interact with sequence *r'*; this interaction may also take place in *ori*2. The overall result is no change in the size of the loop. Only in the case of *ori*7, the loop would be larger by 30 bp, the size of cluster γ.

and the region following cluster C (or γ, if present) mediated by base pairing and by specific proteins would lead to an overall folding of the *ori* sequence, to the formation of a 132-bp double-stranded loop (Fig. 10), and to approaching considerably the critical A–B and C-*r* regions. Such a superfolding is supported by the general properties of sequence *l* and by the fact that remarkable differences among different *ori* sequences do not seem to cause changes in the size of the loop, except in *ori*7 (see legend to Fig. 10).

(j) General considerations and conclusions

(*i*) The *ori* sequences of the mitochondrial genome of yeast share with *ori* sequences of other genomes the general property of being extremely complex structures fulfilling a number of roles required by DNA replication. The first to be demonstrated was that they contain a site for the initiation of RNA transcription (Baldacci and Bernardi, 1982; Baldacci et al., 1984), sequence *r*. This sequence as well as another one located on the opposite strand and on the other side of cluster C are used for RNA-primed bidirectional DNA replication (Baldacci and Bernardi, 1983; Baldacci et al., 1984). Other roles are, for the time being, only presumed on the basis of circumstancial evidence. These comprise a putative site for DNA gyrase within sequence *l*g: possibly, a gene largely corresponding to cluster B and directing the production of a small peptide in wild-type cells; one or more sites for membrane attachment; and sites for the termination of RNA transcription and DNA replication. The latter point is suggested by the finding of RNA transcripts and DNA nascent chains having the length of the monomeric units in petite genomes (Baldacci and Bernardi, 1982; 1983; Baldacci et al., 1984). While anchoring and termination sites are most likely to exist and only need to be identified, the feature in favor of the existence of the peptide-coding gene is that this is only present in active *ori* sequences. An attractive but purely speculative hypothesis is that this peptide is involved in the

cutting of the long RNA transcripts starting at sequence *r* which are present in petites but not in wild-type cells (Baldacci and Bernardi, 1982; Baldacci et al., 1984). As far as the putative gyrase sites are concerned, it should be noted that these appear to be associated with *ori* sequences from the mitochondrial genome of both yeast and mammalian cells.

(*ii*) Even if direct evidence for the lack of transcription is only available for *ori*4, several findings indicate that *ori* sequences carrying a γ cluster are inactive as starting points for both transcription and replication (Baldacci and Bernardi, 1982; 1983; Baldacci et al., 1984; Faugeron-Fonty et al., 1984). Inactive *ori* sequences do not show more sequence divergence, relative to a majority sequence, than active *ori* sequences. The maintenance in the mitochondrial genome and the conservation of the primary structure of *ori*4, 6 and 7 can be best understood by assuming: (1) that these sequences still fulfill other roles of *ori* sequences, the two most obvious ones having to do with unwinding and anchoring of genome units on the mitochondrial membrane, (2) that these roles require secondary and tertiary structures which in turn are only allowed by a large preservation of the primary structure.

(*iii*) The *ori* sequences of the mitochondrial genome of yeast resemble prokaryotic *ori* sequences in two main features. The first is the size, 280 bp in yeast, 245 bp in *E. coli* (the latter being, however, a minimal size not including the actual site of replication initiation; Tabata et al., 1983). The second is that both *ori* sequences are formed by external regions extremely conserved in primary structure and by an internal region in which base changes are tolerated, but not size changes (Tabata et al., 1983). Evident analogies in detailed features are seen between the *ori* sequences of the mitochondrial genome of yeast and the heavy-strand *ori* sequences of animal mitochondrial genomes (see section **h**, above). The mitochondrial genome of yeast makes use, however, of identical *ori* sequences in opposite orientations for its replication, whereas that of animal cells takes advantage of two different *ori* sequences for the replication of its strands (see Clayton, 1982, for a review). Moreover, the heavy-strand *ori* sequences exhibit a considerable variability in the size of the intermediary region in evolutionarily close species and also in their primary structure within a species (Greenberg et al., 1983). It is conceivable that these differences will be seen as less serious when a better understanding of the tertiary structure of these sequences will be available. A comparison of the structural features of mammalian heavy-strand origins with those of yeast *ori* sequences suggests that the former may be less finely tuned compared to the latter.

(*iv*) The mitochondrial genome of yeast is the first multireplicon genetic system on which detailed information is available. Several questions can then be asked. The first concerns the evolutionary origin of several *ori* sequences in the same genome. The most reasonable answer is that they arose by a mechanism of duplication followed by translocation events. In the case of *ori*2 and -7 and of *ori*3 and -4, which are present in the genome as tandem pairs separated by only 2000 bp, it is conceivable that one of the two tandem pairs was duplicated and translocated as a whole to generate the second one; if so, translocation was accompanied by an inversion and followed either by the addition of cluster β in *ori*4 or by its loss in *ori*7.

After duplication events took place, *ori* sequences incurred a number of independent changes. These belong in four classes. First, some *ori* sequences were changed by the insertion of cluster β (*ori*4, 6) and/or γ (*ori*4, 6, 7). Others underwent local tandem duplications of the *r* sequence (*ori*1 and 2 show sequence r'; *ori*2 and 3 contain r'' sequences). No *ori* sequences incurred internal deletions, even if these were possible in terms of available tandem repeats to be used as excision sequences (de Zamaroczy et al., 1983). In contrast, *ori* sequences appear to have undergone local inversions in sequence *l*. Finally, all *ori* sequences underwent point mutations, affecting the central *l* region.

ACKNOWLEDGEMENTS

We thank Dr. Jacques Ninio for his help in computer work, Dr. Renzo Marotta for his collaboration in the early stages of this work, Mr. Philippe Breton for the artwork and Ms. Martine Brient for typing this manuscript.

One of us (G. Baldacci) thanks EMBO for a long-term fellowship.

REFERENCES

Anderson, S., Bankier, A.T., Barrell, B.G., De Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G.: Sequence and organization of the human mitochondrial genome. Nature 290 (1981) 457–465.

Anderson, S., De Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.G.: Complete sequence of bovine mitochondrial DNA. J. Mol. Biol. 156 (1982) 683–717.

Baldacci, G. and Bernardi, G.: Replication origins are associated with transcription initiation sequences in the mitochondrial genome of yeast. EMBO J. 1 (1982) 987–994.

Baldacci, G. and Bernardi, G.: The replication of yeast mitochondrial DNA, in Schweyen, R.J., Wolf, K. and Kaudewitz, F. (Eds.), Mitochondria–1983: Nucleo–Mitochondrial Interactions. De Gruyter, Berlin, 1983, pp. 57–63.

Baldacci, G., Chérif-Zahar, B. and Bernardi, G.: The initiation of DNA replication in the mitochondrial genome of yeast. EMBO J. 3 (1984) 2115–2120.

Bernardi, G.: Mechanism of action and structure of acid deoxyribonuclease, in Nord, F.F. (Ed.), Advances in Enzymology. Interscience, New York, 1968, Vol. 31, pp. 1–49.

Bernardi, G.: The petite mutation in yeast. Trends Biochem. Sci. 4 (1979) 197–201.

Bernardi, G.: The origins of replication of the mitochondrial genome of yeast. Trends Biochem. Sci. 7 (1982a) 404–408.

Bernardi, G.: Evolutionary origin and the biological function of non-coding sequences in the mitochondrial genome of yeast, in Slonimski, P.P., Borst, P. and Attardi, G. (Eds.), Mitochondrial Genes. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982b, pp. 269–278.

Bernardi, G.: Genome instability and the selfish DNA issue. Folia Biol. 29 (1983) 82–92.

Bernardi, G. and Bernardi, G.: Repeated sequences in the mitochondrial genome of yeast. FEBS Lett. 115 (1980) 159–162.

Bernardi, G., Faures, M., Piperno, G. and Slonimski, P.P.: Mitochondrial DNA's from respiratory-sufficient and cytoplasmic respiratory-deficient mutant yeast. J. Mol. Biol. 48 (1970) 23–42.

Bernardi, G., Baldacci, G., Bernardi, G., Faugeron-Fonty, G., Gaillard, C., Goursot, R., Huyard, A., Mangin, M., Marotta, R. and de Zamaroczy, M.: The petite mutation: excision sequences, replication origins and suppressivity, in Kroon, A.M. and Saccone, C. (Eds.), The Organization and Expression of the Mitochondrial Genome. Elsevier, Amsterdam, 1980, pp. 21–31.

Bernardi, G., Baldacci, G., Colin, Y., Faugeron-Fonty, G., Goursot, R., Goursot, R., Huyard, A., Le Van Kim, C., Mangin, M., Marotta, R. and de Zamaroczy, M.: The origins of replication of the mitochondrial genome of yeast, in Nagley, P., Linnane, A.W., Peacock, W.J. and Pateman, J.A. (Eds.), Manipulation and Expression of Genes in Eukaryotes. Academic Press, New York, 1983, pp. 279–289.

Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton, D.A.: Sequence and gene organization of mouse mitochondrial DNA. Cell 26 (1981) 167–180.

Blanc, H. and Dujon, B.: Replicator regions of the yeast mitochondrial DNA responsible for suppressiveness. Proc. Natl. Acad. Sci. USA 77 (1980) 3942–3946.

Borst, P., Heyting, C. and Sanders, J.P.M.: The control of mitochondrial DNA synthesis in yeast petite mutants, in Bücher, Th., Neupert, W., Sebald, W. and Werner, S. (Eds.), Genetics and Biogenesis of Chloroplasts and Mitochondria. Elsevier, Amsterdam, 1976, pp. 525–533.

Bos, J.L., Heyting, C., Van der Horst, G. and Borst, P.: The organization of repeating units in mitochondrial DNA from yeast petite mutants. Curr. Genet. 1 (1980) 233–239.

Broach, J.R., Li, Y.-Y., Feldman, J., Jayaram, M., Abraham, J., Nasmyth, K.A. and Hicks, J.B.: Localization and sequence analysis of yeast origins of DNA replication. Cold Spring Harbor Symp. Quant. Biol. 47 (1983) 1165–1173.

Cantatore, P. and Attardi, G.: Mapping of nascent light and heavy strand transcripts on the physical map of HeLa cell mitochondrial DNA. Nucl. Acids Res. 8 (1980) 2605–2625.

Christianson, T. and Rabinowitz, M.: Identification of multiple transcriptional initiation sites on the yeast mitochondrial genome by in vitro capping with guanylyltransferase. J. Biol. Chem. 22 (1983) 14025–14033.

Christianson, T., Edwards, J.C., Levens, D., Locker, J. and Rabinowitz, M.: Transcriptional initiation and processing of the small ribosomal RNA of yeast mitochondria. J. Biol. Chem. 257 (1982) 6494–6500.

Christianson, T., Edwards, J.C., Mueller, D.M. and Rabinowitz, M.: Identification of a single transcriptional initiation site for the glutamic tRNA and *cob* genes in yeast mitochondria. Proc. Natl. Acad. Sci. USA 80 (1983) 5564–5568.

Clayton, D.A.: Replication of animal mitochondrial DNA. Cell 28 (1982) 693–705.

Cohen, D., Netter, P., Petrochilo, E. and Slonimski, P.P.: Mitochondrial genetics, I. Methodology and Phenomenology. Control of Organelle Development. Symp. Soc. Exp. Biol., vol. 24, Cambridge Univ. Press, London, 1970, pp. 449–496.

Colin, Y., Baldacci, G. and Bernardi, G.: A novel putative gene in the mitochondrial genome of yeast. Gene, submitted.

Coruzzi, G., Bonitz, S.G., Thalenfeld, B.E. and Tzagoloff, A.: Assembly of the mitochondrial membrane system: analysis of the nucleotide sequence and transcripts in the *oxi1* region of yeast mitochondrial DNA. J. Biol. Chem. 256 (1981) 12780–12787.

Crews, S., Ojala, D., Posakony, J., Nishiguchi, J. and Attardi, G.: Nucleotide sequence of a region of human mitochondrial DNA containing the precisely identified origin of replication. Nature 277 (1979) 192–198.

Deutsch, J., Dujon, B., Netter, P., Petrochilo, E., Slonimski, P.P., Bolotin-Fukuhara, M. and Cohen, D.: Mitochondrial Genetics, VI. The petite mutation in *Saccharomyces cerevisiae*; interrelations between the loss of the ρ factor and the loss of the drug resistance mitochondrial genetic markers. Genetics 76 (1974) 195–219.

de Zamaroczy, M., Baldacci, G. and Bernardi, G.: Putative origins of replication in the mitochondrial genome of yeast. FEBS Lett. 108 (1979) 429–432.

de Zamaroczy, M., Marotta, R., Faugeron-Fonty, G., Goursot, R., Mangin, M., Baldacci, G. and Bernardi, G.: The origins

of replication of the yeast mitochondrial genome and the phenomenon of suppressivity. Nature 292 (1981) 75–78.

de Zamaroczy, M., Faugeron-Fonty, G. and Bernardi, G.: Excision sequences in the mitochondrial genome of yeast. Gene 21 (1983) 193–202.

Dumas, J.-P. and Ninio, J.: Efficient algorithms for folding and comparing nucleic acid sequences. Nucl. Acids Res. 10 (1982) 197–206.

Faugeron-Fonty, G., Culard, F., Baldacci, G., Goursot, R., Prunell, A. and Bernardi, G.: The mitochondrial genome of wild-type yeast cells, VIII. The spontaneous cytoplasmic "petite" mutation. J. Mol. Biol. 134 (1979) 493–537.

Faugeron-Fonty, G., Mangin, M., Huyard, A. and Bernardi, G.: The mitochondrial genomes of spontaneous *orir* petite mutants of yeast have rearranged repeat units organized as inverted tandem dimers. Gene 24 (1983) 61–71.

Faugeron-Fonty, G., Le Van Kim, C., de Zamaroczy, M., Goursot, R. and Bernardi, G.: A comparative study of the *ori* sequences from mitochondrial genomes of twenty wild-type yeast strains. Gene 32 (1984) 459–473.

Fauman, M.A., Rabinowitz, M.: DNA-DNA hybridization studies of mitochondrial DNA of ethidium bromide induced petites of yeast. Eur. J. Biochem. 42 (1974) 67–71.

Gaillard, C. and Bernardi, G.: The nucleotide sequence of the mitochondrial genome of a spontaneous "petite" mutant of yeast. Mol. Gen. Genet. 174 (1979) 335–337.

Gaillard, C., Strauss, F. and Bernardi, G.: Excision sequences in the mitochondrial genome of yeast. Nature 283 (1980) 218–220.

Gellert, M., Mizuuchi, K., O'Dea, M.H., Ohmori, H. and Tomizawa, J.: DNA gyrase and DNA supercoiling. Cold Spring Harbor Symp. Quant. Biol. 43 (1979) 35–40.

Gordon, P., Casey, J. and Rabinowitz, M.: Characterization of mitochondrial deoxyribonucleic acid from a series of petite yeast strains by deoxyribonucleic acid hybridization. Biochemistry 13 (1974) 1067–1075.

Gordon, P. and Rabinowitz, M.: Evidence for deletion and changed sequence in the mitochondrial deoxyribonucleic acid of spontaneously generated petite mutants of *Saccharomyces cerevisiae*. Biochemistry 12 (1973) 116–123.

Goursot, R., de Zamaroczy, M., Baldacci, G. and Bernardi, G.: Supersuppressive "petite" mutants of yeast. Curr. Genet. 1 (1980) 173–176.

Goursot, R., Mangin, M. and Bernardi, G.: Surrogate origins of replication in the mitochondrial genomes of *ori*° petite mutants of yeast. EMBO J. 1 (1982) 705–711.

Greenberg, B.D., Newbold, J.E. and Sugino, A.: Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene 21 (1983) 33–49.

Hiraga, S.: Novel F prime factors able to replicate in *Escherichia coli* Hfr strains. Proc. Natl. Acad. Sci. USA 73 (1976) 198–202.

Hudspeth, M.E.S., Ainley, W.M., Shumard, D.S., Butow, R.A. and Grossman, L.I.: Location and structure of the *var1* gene on yeast mitochondrial DNA: Nucleotide sequence of the 40.0 allele. Cell 30 (1982) 617–626.

Hyman, B.C., Cramer, J.H. and Rownd, R.H.: Properties of a *Saccharomyces cerevisiae* mitochondrial DNA segment conferring high-frequency yeast transformation. Proc. Natl. Acad. Sci. USA 79 (1982) 1578–1582.

Kearsey, S.: Analysis of sequences conferring autonomous replication in baker's yeast. EMBO J. 2 (1983) 1571–1575.

Kobayashi,M., Yaginuma, K., Seki, T. and Koike, K.: Nucleotide sequences of the cloned *Eco*A fragment of rat mitochondrial DNA, in Kroon, A.M. and Saccone, C. (Eds.), The Organization and Expression of the Mitochondrial Genome. Elsevier, Amsterdam, 1980, pp. 221–229.

Lang, B., Burger, G., Doxiadis, I., Thomas, D.Y., Bandlow, W. and Kaudewitz, F.: A simple method for the large-scale preparation of mitochondria from microorganisms. Analyt. Biochem. 77 (1977) 110–121.

Lazarus, C.M. and Kuntzel, H.: Anatomy of amplified mitochondrial DNA in "ragged" mutants of *Aspergillus amstelodami*: Excision points within protein genes and a common 215 bp segment containing a possible origin of replication. Curr. Genet. 4 (1981) 99–107.

Lazowska, J., Michel, F., Faye, G., Fukuhara, H. and Slonimski, P.P.: Physical and genetical organization of petite and grande yeast mitochondrial DNA, II. DNA-DNA hybridization studies and buoyant density determination. J. Mol. Biol. 85 (1974) 393–410.

Lewin, A., Morimoto, R., Rabinowitz, M. and Fukuhara, H.: Restriction enzyme analysis of mitochondrial DNAs of petite mutants of yeast: classification of petites and deletion mapping of mitochondrial genes. Mol. Gen. Genet. 163 (1978) 257–275.

Locker, G., Lewin, A. and Rabinowitz, M.: The structure and organization of mitochondrial DNA from petite yeast. Plasmid 2 (1979) 155–181.

Macino, G. and Tzagoloff, A.: Assembly of the mitochondrial membrane system: Sequence analysis of a yeast mitochondrial ATPase gene containing the *oli*-2 and *oli*-4 loci. Cell 20 (1980) 507–517.

Mangin, M., Faugeron-Fonty, G. and Bernardi, G.: The *ori*^{r} to *ori*$^+$ mutation in spontaneous yeast petites is accompanied by a drastic change in mitochondrial genome replication. Gene 24 (1983) 73–81.

Marotta, R., Colin, Y., Goursot, R. and Bernardi, G.: A region of extreme instability in the mitochondrial genome of yeast. EMBO J. 1 (1982) 529–534.

Martin, R.P., Bordonné, R. and Dirheimer, G.: The paromo mycin in the yeast mitochondrial genome, in Akoyounoglou, G., Evangelopoulos, E.A., Georgatsos, J., Palaiologos, G., Trakatellis, A. and Tsiganos, C.P. (Eds.), Cell Function and Differentiation, Part B. Liss, New York, 1983, pp. 355–365.

Maxam, A.M. and Gilbert, W.: A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74 (1977) 560–564.

Michaelis, G., Petrochilo, E. and Slonimski, P.P.: Mitochondrial Genetics, III, Recombined molecules of mitochondrial DNA obtained from crosses between cytoplasmic petite mutants of *Saccharomyces cerevisiae*: physical and genetical characterization. Mol. Gen. Genet. 123 (1973) 51–65.

Morrison, A. and Cozzarelli, N.R.: Site-specific cleavage of DNA by *E. coli* DNA gyrase. Cell 17 (1979) 175–184.

Morrison, A. and Cozzarelli, N.R.: Contacts between DNA

gyrase and its binding sites on DNA: features of symmetry and asymmetry revealed by protection from nucleases. Proc. Natl. Acad. Sci. USA 78 (1981) 1416–1420.

Morrison, A., Brown, P.O., Kreuzer, K.N., Otter, R., Gerrard, S.P. and Cozzarelli, N.R.: Mechanisms of DNA topoisomerases, in Alberts, B. (Ed.), Mechanistic Studies of DNA Replication and Genetic Recombination, ICN-UCLA Symposia on Molecular and Cellular Biology, Vol. 19. Academic Press, New York, 1980, pp. 785–807.

Nobrega, F.G. and Tzagoloff, A.: Assembly of the mitochondrial membrane system. DNA Sequence and organization of the cytochrome b gene in *Saccharomyces cerevisiae* D273-10B. J. Biol. Chem. 255 (1980) 9828–9837.

Osinga, K.A. and Tabak, H.F.: Initiation of transcription of genes for mitochondrial ribosomal RNA in yeast: comparison of the nucleotide sequence around the 5′-ends of both genes reveals a homologous stretch of 17 nucleotides. Nucl. Acids Res. 10 (1982) 3617–3626.

Perlman, P.S. and Birky Jr., C.W.: Mitochondrial genetics in baker's yeast: molecular mechanism for recombinational polarity and suppressiveness. Proc. Natl. Acad. Sci. USA 71 (1974) 4612–4616.

Prunell, A. and Bernardi, G.: The mitochondrial genome of wild-type yeast cells, VI. Genome organization. J. Mol. Biol. 110 (1977) 53–74.

Sanders, J.P.M., Heyting, C. and Borst, P.: The variability of the mitochondrial genome of *Saccharomyces* strains, in Bücher, Th., Neupert, W., Sebald, W. and Werner, S. (Eds.), Genetics and Biogenesis of Chloroplasts and Mitochondria. Elsevier, Amsterdam, 1976, pp. 511–517.

Scovassi, A.I., Torsello, S., Plevani, P., Badaracco, G.F. and Bertazzoni, U.: Active polypeptide fragments common to prokaryotic, eukaryotic and mitochondrial DNA polymerases. EMBO J. 1 (1982) 1161–1165.

Sekiya, T., Kobayashi, M., Seki, T. and Koike, K.: Nucleotide sequence of a cloned fragment of rat mitochondrial DNA containing the replication origin. Gene 11 (1980) 53–62.

Slonimski, P.P. and Lazowska, J.: Transposable segments of mitochondrial DNA; unitary hypothesis for the mechanism of mutation, recombination, sequence reiteration and suppressiveness of yeast "petite colonie" mutants, in Bandlow, W., Schweyen, R.J., Wolf, K. and Kaudewitz, F. (Eds.), Mitochondria–1977: Genetics and Biogenesis of Mitochondria. de Gruyter, Berlin, 1977, pp. 39–52.

Stinchcomb, D.T., Mann, C., Selker, E., Davis, R.W.: DNA sequences that allow the replication and segregation of yeast chromosomes, in Ray, D.S. and Fox, C.F. (Eds.), The Initiation of DNA Replication, ICN-UCLA Symposia on Molecular and Cellular Biology, Vol. 22. Academic Press, New York, 1981, pp. 473–488.

Szybalski, W., Bøvre, K., Fiandt, M., Hayes, S., Hradecna, Z., Kumar, S., Lozeron, H.A., Nijkamp, H.J.J. and Stevens, W.F.: Transcriptional units and their controls in *Escherichia coli* phage λ: Operons and Scriptons. Cold Spring Harbor Symp. Quant. Biol. 35 (1970) 341–353.

Tabata, S., Oka, A., Sugimoto, K., Takanami, M., Yasuda, S. and Hirota, Y.: The 245 base-pair *ori*^{c} sequence of the *E. coli* chromosome directs bidirectional replication at an adjacent region. Nucl. Acids Res. 11 (1983) 2617–2627.

Tikhomirova, L.P., Kryukov, V.M., Strizhov, N.I. and Bayev, A.A.: mt DNA sequences of *Candida utilis* capable of supporting autonomous replication of plasmids in *Saccharomyces cerevisiae*. Mol. Gen. Genet. 189 (1983) 479–484.

Tschumper, G. and Carbon, J.: Sequence of a yeast DNA fragment containing a chromosomal replicator and the *TRP1* gene. Gene 10 (1980) 157–166.

Tschumper, G. and Carbon, J.: Sequencing and subcloning analysis of autonomously replicating sequences from yeast chromosomal DNA, in Ray, D.S. and Fox, C.F. (Eds.), The Initiation of DNA Replication, ICN-UCLA Symposia on Molecular and Cellular Biology, Vol. 22. Academic Press, New York, 1981, pp. 489–500.

Tschumper, G. and Carbon, J.: Delta sequences and double symmetry in a yeast chromosomal replicator region. J. Mol. Biol. 156 (1982) 293–307.

Wong, J.F.H., Ma, D.P., Wilson, R.K. and Roe, B.A.: DNA sequence of the *Xenopus laevis* mitochondrial heavy and light strand replication origins and flanking tRNA genes. Nucl. Acids Res. 11 (1983) 4977–4995.

Communicated by J. Carbon.