A Family of Repetitive Palindromic Sequences Found in *Neurospora* Mitochondrial DNA Is Also Found in a Mitochondrial Plasmid DNA*

(Received for publication, August 27, 1982)

Frank E. Nargang‡ and John B. Bell
From the Department of Genetics, University of Alberta, Edmonton, Alberta, Canada T6G 2E9

Lori L. Stohl and Alan M. Lambowitz
From the Edward A. Doisy Department of Biochemistry, Saint Louis University School of Medicine, St. Louis, Missouri 63104

*This work was supported by Grant U0011 from the Natural Sciences and Engineering Council of Canada and by Grant GM 23961 from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Recipient of an establishment grant from the Alberta Heritage Fund for Medical Research.

Neurospora mtDNA contains a repetitive, 18 nucleotide palindromic sequence (5'-CCCTGCAAGTCTGCAGGG-3') that contains two closely spaced *Pst*I sites (CTGCAG) in the arms of the palindrome (Yin, S., Heckman, J., and RaiBhandary, U. L. (1981) Cell 26, 325-332). In the present study, DNA sequence analysis was carried out to determine whether *Pst*I palindromes are present in an apparently distinct genetic element, the 3.6-kilobase mitochondrial plasmid from *Neurospora crassa* strain Mauriceville-1c (FGSC 2225). The plasmid contains a cluster of closely spaced *Pst*I sites extending over a 0.4-kilobase region (Collins, R. A., Stohl, L. L., Cole, M. D., and Lambowitz, A. M. (1981) Cell 24, 443-452). The DNA sequence shows that the cluster consists of eight *Pst*I sites organized in five palindromic elements. Two of the elements are identical with the canonical sequence found in mtDNA, whereas the remaining three elements differ from the canonical sequence by a few nucleotides. The occurrence of the *Pst*I palindromes in two otherwise unrelated DNA species is consistent with the hypothesis that they are related to mobile DNA sequences that either propagate or were once capable of propagating within mitochondria.

Both *Neurospora crassa* and yeast mtDNAs contain short GC-rich palindromic sequences which are repeated many times throughout the genomes (Prunell and Bernardi, 1977; Cosson and Tzagoloff, 1979; Yin et al., 1981). In *Neurospora*, the repetitive sequence consists of a highly conserved 18 nucleotide palindrome, 5'-CCCTGCAAGTCTGCAGGG-3', that contains two closely spaced *Pst*I sites (CTGCAG) in the arms of the palindrome (Yin et al., 1981). This core sequence is generally preceded on the 5' side by pyrimidines (mostly C's) and followed on the 3' side by purines (mostly G's) to give an extended palindromic structure with the core sequence at the apex (Yin et al., 1981). The yeast repetitive elements consist of less conserved GC-rich clusters, some of which are imperfect palindromes of 20 to 24 nucleotides (Prunell and Bernardi, 1977; Cosson and Tzagoloff, 1979). The yeast GC-rich clusters show virtually no sequence homology with the *Neurospora Pst*I palindromes, but the two elements appear to be structurally related and could have a common function.

The *Neurospora Pst*I palindromes are found flanking genetic elements (e.g. the small and large rRNA genes, protein genes, and tRNA genes) as well as within the large rRNA gene near opposite ends of the 2.3-kilobase intervening sequence (Yin et al., 1981). The yeast GC-rich clusters are most often located outside of genes, but are sometimes found within genes (e.g. the small rRNA and *var-1* genes; Sor and Fukuhara, 1982; Hudspeth et al. 1982). Tzagoloff et al. (1980) and Yin et al. (1981) noted that transcripts containing the GC-rich elements have the potential to form double-stranded structures either by intra- or interstrand base pairing. They proposed that these structures might serve as signals for RNA processing enzymes analogous to *Escherichia coli* RNase III.

Certain wild type *Neurospora* strains contain intramitochondrial plasmid DNAs in addition to the standard mtDNA (Collins et al., 1981; Stohl et al., 1982). The mitochondrial plasmids appear to constitute a distinct class of genetic elements. They show virtually no sequence homology with the standard mtDNA in DNA-DNA hybridization experiments, and they achieve high copy number without suppressing behavior toward wild type mtDNA. In these respects, they are clearly distinguished from defective mtDNA species found in *Neurospora* stopper mutants and in other fungi (Bertrand et al., 1980; de Vries et al., 1980). Restriction enzyme mapping showed that the 3.6-kilobase mitochondrial plasmid from *N. crassa* strain Mauriceville-1c (FGSC 2225) contains two unusual restriction site clusters: one consisting of six *EcoRI* sites in a 1-kilobase region and the other of five or more *Pst*I sites in a 0.4-kilobase region (Collins et al., 1981). In the present work, DNA sequence analysis shows that the latter cluster contains eight *Pst*I sites organized in five palindromic elements. Two of the elements are identical with the canonical 18-nucleotide sequence found in the mtDNA, whereas the remaining three elements differ from the canonical sequence by a few nucleotides.

**MATERIALS AND METHODS**

*Isolation and Cloning of Mitochondrial Plasmid DNA—Mitochondrial plasmid DNA was isolated from the Mauriceville-1c strain of *N. crassa* (Fungal Genetics Stock Center 2225) as described by Collins et al. (1981). To facilitate large scale isolation of mitochondrial plasmid DNA, the 2.6-kilobase *EcoRI* fragment of the plasmid, which contains the entire *Pst*I cluster, was cloned into the single *EcoRI* site of the *E. coli* plasmid pBR322. The recombinant plasmid is designated pLS53.*

*Isolation of Plasmid DNA from *E. coli*—Plasmid DNA was isolated from chloramphenicol-amplified *E. coli* cultures by equilibrium centrifugation in CsCl-ethidium bromide gradients as described by Kahn et al. (1979).*
DNA Sequencing—DNA sequence was determined by the dideoxy method (Sanger et al., 1977). The M13 phage derivatives mp7, mp8, and mp9 (obtained from Bethesda Research Labs) were used for the production of single-stranded template DNA. Restriction enzymes and T4 DNA ligase were obtained from New England Biolabs (Beverly, MA), Bethesda Research Labs, and Boehringer Mannheim and were used in accordance with the supplier's instructions. Bacterial cells were transformed with the replicative form of M13 DNA using the procedure of Norgard et al. (1978). The DNA sequence was analyzed using the computer programs of Staden (1977, 1978, 1979, 1980a, 1980b).

RESULTS AND DISCUSSION

Fig. 1 shows the nucleotide sequence of a 681-base pair HindIII fragment which contains the entire PstI cluster of the Mauriceville plasmid. The entire sequence of each strand has been determined. The cluster contains eight PstI sites which are present in five separate, but related, sequence elements (Fig. 2). Two pairs of PstI sites, beginning at positions 107 and 454, are present in the canonical 18-nucleotide sequence, 5'-CCCTGCAGTACTGCAGGG-3' (Fig. 2a and e). Another pair of PstI sites is found in an 18-nucleotide sequence which begins at nucleotide 323 and which differs from the canonical sequence at the 2nd and 17th positions (Fig. 2b). Interestingly, the substitutions preserve the perfect palindromic structure.

Fig. 2. Summary of the PstI elements and flanking sequences. The elements shown in a and e contain the canonical 18-nucleotide sequence. PstI sites are enclosed in boxes. Arrows indicate nucleotides that differ from the canonical sequence and/or where insertion of a specified nucleotide(s) would restore the sequence. The Δ beneath the underlined bases in b indicates that deletion of those four nucleotides would also restore the canonical sequence.
of the 18-base pair sequence. The remaining two PstI sites are present in sequences which have diverged further from the canonical sequence. One PstI site is found in a 17-nucleotide sequence which begins at position 153 and which differs from the canonical sequence in three positions (Fig. 2b). The other single PstI site is found in a 16-nucleotide sequence which begins at position 276 and which differs from the canonical sequence in two positions (Fig. 2c). In both of the latter cases, the changes abolish one of the paired PstI sites in the canonical sequence. The PstI elements in the plasmid DNA are generally preceded on the 5' side by pyrimidines and followed on the 3' side by purines, but with the exception of the element at position 276, the flanking sequences do not include runs of C's and G's, as they do in mtDNA (Fig. 2; cf. Yin et al., 1981).

The sequence in Fig. 1 was scanned by computer and found to contain no other sequences closely related to the canonical PstI palindromic sequence. However, the sequence does contain a relatively large direct repeat. The 162 nucleotides from positions 66 to 228 are identical with the 185 nucleotides from positions 229 to 397 in a total of 143 positions (Fig. 3). Each component of the repeat contains two PstI elements. Thus, the presence of two of the elements could simply reflect a duplication event.

The sequence contains a single ATG methionine codon at position 535 to 538 on the bottom strand (Fig. 1). This ATG begins a reading frame of 35 amino acid residues. There are also two relatively long reading frames that do not begin with ATG codons. One begins with the CCC codon at position 68 on the top strand and extends for 176 amino acids to the TAA termination codon at 596 to 598. The other begins with the AGA codon at position 636 on the bottom strand and extends for 190 amino acids to the TAA termination codon at 66.

Although these reading frames contain no ATG codons, both contain several isoleucine codons (ATC, ATA, ATT) which initiate certain unassigned reading frames in mammalian mtDNAs (Anderson et al., 1981; Bibb et al., 1981). The longest reading frame which might originate outside the cluster begins at positions 1 to 3 on the top strand and extends inward for 99 amino acids. Sequence data (not shown) extending beyond the HindIII site at positions 1 to 6 indicate that a TAA termination codon occurs in this reading frame four codons upstream. There are no ATG codons within this reading frame although, again, several isoleucine codons are present. The significance of these reading frames cannot be judged at the present time.

There are also several inverted repeat regions in the sequence. The largest is centered around the canonical PstI palindromic which begins at nucleotide 107. It extends from nucleotide 88 to 141 and contains only six mismatched bases. Although each of the other PstI elements also exists as either perfect or imperfect inverted repeats, the palindromes are not extended by the flanking sequences.

The presence of PstI palindromic sequences in two otherwise unrelated DNA species, Neurospora mtDNA and the Mauriceville plasmid, could be accounted for by common ancestry, by a common mechanism for intramolecular repetition of DNA sequences, or by some type of recombination event (e.g., transposition). Our findings are consistent with the idea that the PstI elements are related to mobile DNA sequences which either propagate within the genome or arise as a result of recombination. This hypothesis has been advanced previously by Yin et al. (1981) for the Neurospora PstI palindromes and by Sor and Fukuhara (1982) for the yeast 6C-rich clusters.

The divergence of three of the five PstI elements on the Mauriceville plasmid contrasts with the virtually total conservation of the canonical sequence on mtDNA (see Yin et al., 1981). One possible inference is that the PstI palindromes have some function on mtDNA which may be shared by some, but not all of the PstI elements on the plasmid. In this regard, it is interesting that only the outermost PstI elements in the cluster are identical with those on mtDNA (Fig. 1). In the absence of experimental evidence, the list of possible functions could encompass virtually all aspects of gene expression, including RNA processing as suggested by Yin et al. (1981). Viewed from the perspective of the mitochondrial plasmids, there is an interesting correlation between the PstI palindromes and the concentration of transcripts detectable by hybridization. Thus, of the three mitochondrial plasmids which have been described, only the Mauriceville plasmid, which contains clustered PstI sites, gives rise to a readily detectable transcript (Collins et al., 1981; Stohl et al., 1982). In the mtDNA, the PstI palindromes are found flanking rRNA, tRNA, and protein genes, and the majority are found in the rRNA-tRNA region (Yin et al., 1981; Browning and RajBhandary, 1982). An interesting possibility is that the palindromic GC-rich sequences enhance transcription of particular mtDNA regions, either by serving as enzyme recognition sites or by a localized effect on DNA structure (cf. the 72-base pair direct repeat of SV40 DNA; Dynan and Tijan, 1982). Further insight into the function of the PstI elements may come when the Mauriceville plasmid has been more fully characterized with respect to genetic function and location of genetic elements such as the DNA replication origin and promoters.

Acknowledgments—We thank D. Vig for technical assistance, R. Vincent for helpful discussion, and P. Fedorak, M. Holmes, K. Morgan, and K. Roy for assistance with computer programming.

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. (1981) Nature (Lond.) 290, 457-465

Bertrand, H., Collins, R. A., Stohl, L. L., Goewert, R. R., and Lambowitz, A. M. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 6032-6036

Bibb, M. J., Van Etten, R. A., Wright, C. T., Walberg, M. W., and Clayton, D. A. (1981) Cell 25, 167-180

Browning, K. S., and RajBhandary, U. L. (1982) J. Biol. Chem. 257, 5253-5256

Collins, R. A., Stohl, L. L., Cole, M. D., and Lambowitz, A. M. (1981) Cell 24, 443-452

Cosson, J. and Tzagoloff, A. (1979) J. Biol. Chem. 254, 42-43
Holland Biomedical Press, Amsterdam

Dynan, W., and Tijan, R. (1982) Trends Biochem. Sci. 7, 124–125
Hudspeth, M. E. S., Ainley, W. M., Shuman, D. S., Butow, R. A.,
and Grossman, L. I. (1982) Cell 30, 617–626
Kahn, M., Kolter, R., Thomas, C., Figurski, D., Meyer, R., Remaut,
E., and Helinski, D. R. (1979) Methods Enzymol. 68, 268–280
Norgard, M. V., Keem, K., and Monahan, J. J. (1978) Gene 3, 279–
282
Prunell, A., and Bernardi, G. (1977) J. Mol. Biol. 110, 53–74
Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad.
Sci. U. S. A. 74, 5463–5467
Sor, F., and Fukuhara, H. (1982) Nucleic Acids Res. 10, 1625–1633
Staden, R. (1977) Nucleic Acids Res. 4, 4037–4051
Staden, R. (1978) Nucleic Acids Res. 5, 1013–1015
Staden, R. (1979) Nucleic Acids Res. 6, 2601–2609
Staden, R. (1980a) Nucleic Acids Res. 8, 817–825
Staden, R. (1980b) Nucleic Acids Res. 8, 3673–3694
Stohl, L. L., Collins, R. A., Cole, M. D., and Lambowitz, A. M. (1982)
Nucleic Acids Res. 10, 1439–1458
Tzagoloff, A., Nobrega, M., Akai, A., and Macino, G. (1980) Curr.
Genetics 2, 149–157
Yin, S., Heckman, J., and RajBhandary, U. L. (1981) Cell 26, 325–332
A family of repetitive palindromic sequences found in Neurospora mitochondrial DNA is also found in a mitochondrial plasmid DNA.
F E Nargang, J B Bell, L L Stohl and A M Lambowitz

J. Biol. Chem. 1983, 258:4257-4260.

Access the most updated version of this article at http://www.jbc.org/content/258/7/4257

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/258/7/4257.full.html#ref-list-1