A hairpin plastid genome in barley

T.H.N.Ellis and A.Day¹

John Innes Institute, Colney Lane, Norwich, UK

¹Present address: Department of Molecular Biology, University of Geneva, 30 quai Ernest Ansermet, 1211 Geneva 4, Switzerland

Communicated by D.A.Hopwood

A linear plastid genome of a barley albino plant derived from anther culture has been characterized. The monomeric genome is 21 kb long. This molecule has a hairpin structure. We discuss the possible origins of this genome.

Key words: barley/genome/hairpin/plastid

Introduction

The organization of the higher plant chloroplast genome has been well conserved in evolution (Palmer, 1985). Generally the genome is represented by an ~130 kb genome which has a circular restriction map. This map is divided into three regions: two single copy regions of ~70 and 20 kb, separated by an inverted repeat of about 20 kb.

Recent studies of this genome in albino plants derived from anther culture (Day and Ellis, 1984, 1985) have catalogued a series of gross rearrangements, most of which are deletions. These variant plastid genomes (the albino plants lack chloroplasts) exist in a variety of physical forms. Two deletion derivatives of the wheat plastid genome are likely to have arisen as a consequence of a single recombination event between the two single copy regions. Any such recombination event generates a deletion, and the consequence of this mechanism has been proposed as the basis of the evolutionary stability of the higher plant plastid genomes (Day and Ellis, 1984, 1985). A comparison of the chloroplast genomes of *Chlamydomonas reinhardtii* and *C. eugametos* (Lemieux and Lemieux, 1985) shows many rearrangements within each single copy region, but only one rearrangement mixes sequences from the two single copy regions. This may occur as a consequence of a double recombination event, or some more complex process, without generating a deletion of the circular molecule.

Circular deleted genomes were not predicted variants of the plastid genome, but their occurrence in these albino plants does not require us to invoke any novel process occurring in the plastid genome. However, several variant plastid genomes have been found to be linear molecules (Day and Ellis, 1985). The origin of these molecules is not easily understood. This paper describes some relevant properties of one such linear barley plastid genome.

Results and discussion

Possible origins of linear plastid DNAs

We have considered some possibilities for the generation of linear DNA molecules from a circular molecule. The most obvious mechanism is breakage as a consequence of shear. In the case of the linear molecules in the albino plant BA2a the ends of the linear plastid DNA were at a defined point (Day and Ellis, 1985). This indicates that if the linears are derived by breakage then

the break point must be at a unique site. Alternatively the linear plastid DNAs may derive from the action of an endonuclease in the cell lineage leading to the formation of pollen, which is active in a manner analogous to the restriction methylation hypothesis of Sagar and Ramanis (1973). Thus the linears would be a consequence of pollen development and not reflect more general properties of plastid DNA.

Three other possible mechanisms occurred to us, and these are illustrated in Figure 1. The first is where a crossover occurs between two replicating molecules, in which replication is via a rolling circle mechanism (Kolodner and Tewari, 1975); the crossover occurs between inverted repeat sequences of opposite polarity with respect to the free end (Figure 1a).

The second invokes a bi-directional replication mechanism, where recombination occurs between the replicated and unreplicated copy of the inverted repeat sequence. Resolution of this structure in the absence of further recombination generates a linear DNA. A second recombination event would generate a rolling circle form (Figure 1b, ix).

The third process involves neither recombination nor the large inverted repeat sequences. If a replication fork reaches a short inverted repeat sequence then at the centre of the sequence a strand exchange can occur because the next few bases added are homologues of either the newly replicated strand or the unreplicated DNA (Figure 1c). Strand switching at the replication fork has also been suggested as the mechanism that generates the extrachromosomal rDNA of *Tetrahymena* (Yao and Gall, 1977). Since strand switching produces snapback DNA it is familiar as a process to avoid during *in vitro* labelling by nick-translation.

Of the three mechanisms illustrated in Figure 1 neither the first nor the second can simply account for the linear barley plastid DNAs because none of the deleted plastid genomes have homology with the large inverted repeat sequence, but either mechanism could generate intermediates in the formation of these linear DNA molecules.

Of all the postulated mechanisms only the one illustrated in Figure 1c requires that at least one end of the linear molecules be distinct from the sheared ends of the linear DNAs of these preparations. We therefore investigated the nature of the ends of the linear DNAs of the barley albino plant BA10.

The mode of replication of plastid DNA is an important factor influencing the mechanisms which can generate linear DNAs. If plastid DNA replication lacks discontinuous DNA synthesis, and proceeds by a displacement mechanism (see Clayton, 1982; Kolodner and Tewari, 1975), generating large tracts of single-stranded DNA, then strand switching may be favoured. Furthermore if the plastid DNA polymerase is a strand displacement polymerase, then the hairpin molecule illustrated in Figure 1c may be displaced by a second round of DNA synthesis.

The ends of the linear plastid DNAs

Figure 2 illustrates an experiment to investigate the nature of the ends of the linear DNAs of BA10. The untreated DNA of this plant contains several linear DNA molecules that hybridize to pHvc P1, the major classes being 21 and 43 kb in length, and

Fig. 1. Generation of linear DNA molecules. (A) Recombination between two rolling circle replicative forms. Inverted repeat sequence is indicated by a long arrow. Polarity of small single copy region is indicated by an open triangle. (B) The consequences of recombination between inverted repeats during replication. (i) Replication begins. (ii) and (iii) Recombination occurs between a replicated and the unreplicated copy of the large inverted repeat (IR) sequence (Futcher, 1981). (iv–vi) The replication fork at the bottom of (iii) reaches the one at the top. This could happen at any time after the crossover but is simplest to draw before several rounds of replication are completed. (vii) No further recombination events happen to (vi) generating a linear. (viii) and (ix) Recombination again between IR sequences generates a standard rolling circle. (C) (i) Replication is initiated. (ii–iv) Strand switching. This is drawn as occurring due to the looping out of the a–a' hairpin, but this is not necessarily the mechanism by which the strand switching occurs. (v) The whole structure after DNA synthesis has proceeded. (vi) The hairpin molecule can be displaced. (vii–x) This molecule can replicate after the open end is sealed. (xi) Recombination between a linear monomer and a circular dimer generates a linear trimer. Recombination within a trimer can generate a dimer.

probably represent monomer and dimer of the same molecule. These molecules are not susceptible to the action of DNA ligase without the addition of both polynucleotide kinase and T4 DNA polymerase. These molecules must therefore contain at least one frayed end.

Figure 3A illustrates a second experiment. One sample of BA10 DNA has been loaded normally, and a second has been loaded in a solution containing 0.2 M NaOH. Under these conditions duplex DNA will melt at room temperature and enter the gel as single strands, and subsequently migrate as single stranded molecules. However, the 21-kb molecule has migrated to essentially the same position. The slight difference is attributable to the inevitable difference in the ionic strengths of the loading wells. The behaviour of these molecules suggests that after alkali

denaturation, the 21-kb duplex is reformed. These molecules must therefore be hairpins, because as shown in Figure 2, at least one end is frayed. The 43-kb molecules are also visible in this track, but have a reduced intensity. Presumably these molecules are nicked a little, and not many intact 43-kb duplexes reform on entering the gel. Random nicking will form a series of molecules which are partly single stranded and partly double stranded, as well as molecules which are entirely single stranded. These molecules, and similar products derived from the 21-kb hairpin contribute to the added background in the alkali loaded track.

Figure 3B illustrates the same experiment after treatment of the sample with S1 nuclease. In this case neither the 21- nor the 43-kb hairpins can reform after loading in alkali. This confirms that the 21-kb molecule in the alkali loaded (but not S1 treated)

Fig. 2. Effect of T4 ligase, T4 DNA polymerase and T4 polynucleotide kinase on plastid DNA molecules of BA10. M, Marker track: λ DNA, λ *Xho*I, λ *Hind*III and λ *Eco*RI digests. Fragment sizes are indicated in kb. U, Untreated BA10 DNA; P, BA10 DNA + T4 DNA polymerase; L, BA10 DNA + T4 DNA ligase; K, BA10 DNA + polynucleotide kinase. T4 DNA ligase reactions were carried out at 37°C for 20 min in ligation buffer containing 7% polyethylene glycol (Pheiffer and Zimmerman, 1983). Sequential reactions were carried out in the order indicated, i.e. PL = polymerase then ligase, KL = kinase then ligase, KP = kinase then polymerase, KPL = kinase then polymerase, then ligase. In these sequential reactions buffers were changed after phenol extraction and ethanol precipitation. The reaction products have been run out on a 0.8% w/v agarose gel and electrophoresed for 115 h at 15 V/cm, transferred to nitrocellulose and hybridized with nick-translated pTac P10, and a small amount of radioactive phage λ DNA.

track is not simply the single-stranded product of melting the 43-kb molecule. Furthermore the constraint on these molecules, which allows immediate renaturation is sensitive to S1 nuclease, and is thus most likely to be a hairpin structure rather than a crosslink or tightly bound protein.

When BA10 DNA is digested with *Sst*I only three major DNA fragments homologous to pHvc P1 are generated; these are ~4.1, 8.3 and 17 kb in length (Table 2 in Day and Ellis, 1985 and Figure 4). A simple restriction map for the monomer and dimer of the major BA10 plastid genome can be derived from these data; the 4.1-kb *Sst*I fragment is internal to the 21-kb genome, and two 8.3-kb fragments are derived from the ends. The 17-kb molecule is internal to the dimer alone (see Figure 5).

From Figure 4 it can be seen that the 17-kb *Sst*I fragment is susceptible to digestion with S1 nuclease. No new fragment size class is generated upon S1 digestion, suggesting that there is a

Fig. 3. Effect of neutral and alkaline loading on BA10 DNA. (A) BA10 DNA untreated; (B) BA10 DNA digested with S1 nuclease. n, Neutral loading; a, loading in buffer containing 0.2 M NaOH. Sizes of double-stranded DNAs are indicated. 0.8% agarose gel, 0.75 V/cm for 240 h. The mobility of marker fragments is indicated, sizes are given in kb. The hybridization is detected as homology to nick-translated pHvc P1.

Fig. 4. *Sst*I digests of BA10 DNA. n, Neutral loading; a, alkaline loading as Figure 3; g, glyoxylated DNA (see Carmichael and McMaster, 1980; ss, single-stranded DNA; ds, double-stranded DNA; -, *Sst*I digest; +, *Sst*I digest subsequently digested with S1 nuclease. Relevant fragments are arrowed, approximate sizes of ds and glyoxylated (ss) DNA are indicated. 0.8% agarose gel, 1.5 V/cm for 16 h. The hybridization probe is nick-translated pHvc P1.

single-stranded nick between the two monomeric units in the 43-kb dimer. Furthermore this explains why the 43-kb dimer itself is susceptible to digestion by nuclease S1 (see Figure 3).

When *Sst*I digested DNA is loaded on a neutral gel in 0.2 M NaOH, only the 8.3-kb fragment is unaltered in its mobility. This fragment largely disappears after digestion by S1 nuclease (Figure 4). This suggests that the 8.3-kb fragments contain the region of foldback, and thus run as double-stranded DNA after alkaline loading. After S1 digestion, the 8.3-kb fragment is largely replaced by a faster running DNA species, when loaded in alkali. Some fragments can be seen at this position in the track which has not been digested with S1 nuclease. Presumably these derive in part from the 17-kb *Sst*I fragment and in part from the 8.3-kb fragments which have a ligatable end (Figure 2).

When the BA10 *Sst*I digest (with and without S1 treatment) is glyoxylated prior to loading on the gel, essentially the same behaviour results as was seen for the alkaline loading. However

the 17-kb fragment is clearly seen, presumably replacing the 8.3-kb hairpin. Some 8.3-kb single-stranded DNAs can be seen before S1 treatment, but most of the 17-kb single strands are cut by S1 nuclease to generate two 8.3-kb molecules. A small amount of the 17-kb single strand is not cut by S1 nuclease. This could represent either 17-kb *Sst*I fragments, which do not contain a nick, or incomplete S1 digestion.

The 21-kb linear monomer appears to be derived entirely from the 20.5-kb *Pst*I fragment (P1) of barley plastid DNA because the probes pHvc P2, P3, P5, P6, P8, P9, P10 (Day and Ellis, 1985) and pTac B2 (unpublished result) do not hybridize to restriction enzyme digests of BA10 DNA (Day and Ellis, 1985). The P1 fragment of barley plastid DNA contains three internal *Xho*I sites; the largest *Xho*I/*Pst*I fragment from this region is 11.9 kb and derives from the P3 adjacent end of P1 overlapping T8 [see Day and Ellis (1985) for map location].

No *Xho*I sites have been found in the BA10 monomeric genome

Fig. 5. Restriction maps of BA10 pt DNA. Maps of monomer and dimer are shown. Fragment sizes in kb are indicated. T, *Sst*I site; S1, S1-nuclease sensitive site.

Fig. 6. Interconversion of theta and sigma replicative forms. (i) and (ii) Replication is initiated, strand switching occurs as in Figure 1c, i–iv. (iii) The region where strand switching occurred. The strands are labelled x, y and z for clarity. Z is the unreplicated DNA, and the hairpin has formed between x and y. (iv) The hairpin can melt out to form a four-stranded junction which has the structure of a recombination intermediate (Holliday, 1964). (v) The Holliday junction can be resolved by recombination. (vi) Resolution of the Holliday junction generates a rolling circle replicative form with a protected end. Generation of a rolling circle form from a structure like (ii) can explain how the pea chloroplast genome which does not have an inverted repeat organization may interconvert from a theta to sigma replicative form (Kolodner and Tewari, 1975, 1979). If the strand switching event took place at the opposite end of the replication bubble shown in (ii), then a rolling circle replicative form is again generated, but the fork proceeds around the genome in the opposite direction. Recombination between the tails of these two sigma forms will generate a linear DNA molecule (qv. Figure 1A) with both ends folded back on themselves (i.e. a self complementary circular single-stranded DNA).

[Table 2 in Day and Ellis (1985) where the *Xho*I fragment size was quoted as 27 kb], and the T8 fragment is reduced in size; suggesting that the BA10 plastid genome is probably derived from the 11.9-kb *Xho*I/*Pst*I fragment. We are thus forced to conclude that the BA10 monomeric genome contains some sequence duplication. The monomeric genome may be an inverted repeat sequence, with a dyad symmetry in the centre of the 4.1-kb *Sst*I fragment (Figure 5). This notion is supported by the observation that the two *Sst*I end fragments are the same size.

If the monomeric genome is an inverted repeat sequence, then we expect the S1-digested monomeric genome to run as a 10.5-kb duplex on a neutral gel when loaded in alkali (Figure 3), and we expect the 4.1-kb *Sst*I fragment to run as a 2.05-kb duplex DNA under the same conditions (Figure 4). Both these phenomena are observed, but single-stranded DNAs run close to the

position of double-stranded DNAs of half their size, on the gel systems we have employed (e.g. the 8.3-kb ss DNA runs near the 4.1-kb ds DNA on Figure 4).

Although this unfortunate property of our gel system makes us cautious in our interpretation, two more observations suggest that these DNAs (the '10.5-kb' S1-derived fragment and the '2.05-kb' *Sst*I-derived fragment) are in fact duplex molecules.

First the S1-digested BA10 DNA loaded in neutral buffer also contains some of this 10.5-kb molecule (Figure 3). This molecule has been seen in other experiments where the two alkaline loadings are well away from the neutral loadings, suggesting that this molecule does not derive from partial denaturation of the BA10 DNA due to its proximity to the sample well containing 0.2 M NaOH (data not shown). If the monomer is an inverted repeat sequence then we might expect it to exist in an equilibrium

between two hairpin structures (Figure 5B, only one of these, forms B, is cleaved by S1 into two 10.5-kb duplex molecules.

Secondly the *Sst* fragment running at the 2.05-kb position in Figure 4 is a single discrete band, while the 8.3-kb single-stranded molecule (arrowed) appears to form a wide band, and is resolved into two components as a consequence of strand separation in this gel system (e.g. see Domoney *et al.*, 1986). The lack of strand separation of the 2.05-kb fragment suggests that it is probably double stranded. The 2.05-kb *Sst* fragment is just detectable in neutral loadings of BA10 digests. This is consistent with the predominance of the A form of the monomer (Figure 5B) in BA10 DNA extract.

We thus infer that the 21-kb monomeric plastid genome of BA10 is an inverted repeat sequence. The map of the BA10 genome is similar to the genomes mapped in BA2a (Day and Ellis, 1985). In particular the 6.5-kb monomeric genome has both a truncated T8 fragment, and terminal *Sst*I fragments which are identical in size. If this genome is also an inverted repeat, then the origin of replication of the barley plastid genome can be mapped to a 3.3-kb region of the wild-type barley plastid DNA, corresponding to 2.1 kb of T8 (i.e. including S9 and S11) and 0.6 kb of T2 adjacent to T8 [for map location see Day and Ellis (1985)].

Conclusion

We have described plastid DNA molecules which have a hairpin structure. These linear molecules must have existed *in vivo*. The structure of the molecules suggests that they may have arisen by strand switching during DNA replication. Thus both the defective linear and circular plastid genomes of albino wheat and barley plants derived from another culture (Day and Ellis, 1984, 1985) can be seen as minor aberrations of processes which are normal events in plastid DNA, i.e. replication and recombination. We have argued elsewhere (Day and Ellis, 1984) that the recombination events which generate deleted circular plastid DNAs are responsible for the uniformity of plastid genome organization in higher plants.

The linear DNAs of these albino plants seem more of an aberration, and their presence difficult to rationalize. Of the various possible origins of these hairpin plastid genomes, the one illustrated in Figure 1c seems the best fitted to our observations. However, we have considered only simple models and in reality many more complex processes may occur.

It is possible that linear DNAs may be involved in normal replication; Figure 1b shows a replication scheme which interrelates theta- and sigma-type replication with a linear DNA. Both replication strategies have been proposed for chloroplast DNA (Kolodner and Tewari, 1975). If linear DNAs are involved in normal replication then recombination between linear and circular DNA will permit the resolution of interlocked plastid DNA circles. This may be an important mechanism for permitting the segregation of replicated plastid DNA.

The linear DNA molecules we have described may be generated from normal replicative forms of plastid DNA, by fairly simple processes. For example the formation of the hairpin structure shown in Figure 1c may represent a mechanism for converting the theta-type replicative form to a rolling circle form by arresting the movement of one of the replication forks, and thus also protecting the free end of the rolling circle (Figure 6). If in the scheme outlined in Figure 6, the strand switching occurred at the other end of the replication bubble in a proportion of molecules (Figure 6, ii) then the final sigma form will be like that shown in Figure 6, vi, but the end of the linear region will be at the opposite side of the replication origin from that shown.

Recombination within the region of homology between these two hypothetical structures would generate a linear DNA spanning the replication origin, and both ends of the molecule would be folded back on themselves as shown in Figure 1c, vii.

The role of linear DNAs in plastid biology of higher plants is very poorly understood. However, the existence of these molecules has now been established, and it would seem timely to begin a search for these DNAs in green plants.

Materials and methods

Plant material

The origin of the albino barley plants is described elsewhere (Day and Ellis, 1985), but it should be reiterated that the albino plants are very small (less than 5 g wet weight), and that they yield only about 50 µg of DNA for analysis.

Plasmids

pHvc P1 corresponds to the 20.7 (P1) *Pst*I fragment of barley chloroplast DNA cloned in the vector pAT 153 (Day and Ellis, 1985). The plasmid pTac P10 is a wheat chloroplast DNA fragment that maps to the P1 fragment of barley chloroplast DNA ~1.4 kb from the P1/P3 junction. This plasmid was a gift of C.M.Bowman and T.A.Dyer of the Plant Breeding Institute in Cambridge.

Nucleic acid manipulations

These were as previously described (Day and Ellis, 1985) except that the DNA hybridization reactions of Figures 3 and 4 were performed in 600 mM NaCl, 120 mM Tris-HCl pH 8.0, 8 mM EDTA, supplemented with 50 µg/ml heparin, 5 × Denhardt's solution (Denhardt, 1966), as well as 0.1% w/v sodium pyrophosphate and 10% w/v Dextran sulphate (Sigma), at a temperature of 65°C. (These gels had been washed in 0.12 M HCl prior to transfer in order to facilitate the transfer of hairpin DNAs to nitrocellulose.) Filters were washed in 0.1 × SSC at 65°C. The enzymes T4 DNA polymerase, T4 polynucleotide kinase and S1 nuclease were purchased from BRL (UK) Ltd, Cambridge. S1 digestion reactions were performed using ~50 units of the enzyme per µg of DNA, in 30 mM sodium acetate pH 4.6, 50 mM NaCl, 10 mM ZnCl₂.

Acknowledgements

We thank S.M. and A.K.Day for the safe custody of BA10 DNA, S.N.Covey and C.M.Thomas for useful discussions; B.Huang for her help with anther culture; C.Domoney and D.R.Davies for their comments on the manuscript; E.T.Dalzell for typing the manuscript; P.Scott and A.Davies for photographic work. A.D. was supported by an Agricultural Research Council Studentship.

References

- Carmichael, G.G. and McMaster, G.K. (1980) *Methods Enzymol.*, **65**, 380–391.
- Clayton, D.A. (1982) *Cell*, **28**, 693–705.
- Day, A. and Ellis, T.H.N. (1985) *Curr. Genet.*, **9**, 671–678.
- Day, A. and Ellis, T.H.N. (1984) *Cell*, **39**, 359–368.
- Denhardt, D.T. (1966) *Biochem. Biophys. Res. Commun.*, **23**, 641–646.
- Domoney, C., Ellis, T.H.N. and Davies, D.R. (1986) *Mol. Gen. Genet.*, **202**, 280–285.
- Futcher, A.B. (1981) D.Phil. Thesis, University of Oxford.
- Holliday, R. (1964) *Genet. Res.*, **5**, 282–304.
- Kolodner, R. and Tewari, K.K. (1975) *Nature*, **256**, 708–711.
- Kolodner, R. and Tewari, K.K. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 41–45.
- Lemieux, B. and Lemieux, C. (1985) *Curr. Genet.*, **10**, 213–220.
- Palmer, J.D. (1985) *Annu. Rev. Genet.*, **19**, 325–354.
- Pheiffer, B. and Zimmerman, S.D. (1983) *Nucleic Acids Res.*, **11**, 7853–7871.
- Sager, R. and Ramanis, Z. (1973) *Theor. Appl. Genet.*, **43**, 101–108.
- Yao, M.-C. and Gall, J.G. (1977) *Cell*, **12**, 121–132.

Received on 28 July 1986