The endpoints of an inversion in wheat chloroplast DNA are associated with short repeated sequences containing homology to *att-lambda*

Christopher J. Howe

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK

Summary. The endpoints of an inversion in wheat chloroplast DNA are shown to be associated with copies of a short repeated sequence. Recombination across the repeats in an inverted configuration may have been responsible for the inversion, although they are currently in a direct orientation owing to a second inversion. The repeated sequence contains an element homologous to the core of the bacteriophage lambda *att-site,* which can function as such in vivo.

Key words: Chloroplast DNA - Repeated sequences -Recombination - *att*

Introduction

A number of lines of evidence indicate that genetic recombination does occur in chloroplasts. Perhaps the most direct is the demonstration of recombination between markers in the *Chlamydomonas* chloroplast genome, and gene conversion has also been shown to occur in *Chlamydomonas* (Gillham 1978). Another example is shown by those chloroplast genomes which are composed of two single copy sequences separated by a pair of large inverted repeated sequences. The chloroplast DNA of several such plants is composed of two types of molecule having the single copy sequences in different relative orientations. These two forms are probably interconverted by recombination across the inverted repeats (Palmer et al. 1984). This recombination must be frequent in the lifetime of the plant, or stochastic fluctuations might be expected to cause an imbalance in the proportions of the two types of molecule. Intermolecular recombination, perhaps between the repeats, may be responsible for the generation of circular dimers of chloroplast DNA (Kolodner and Tewari 1979).

A related example of recombination is that of the evolutionary rearrangement of chloroplast DNA. Although the organisation of the chloroplast DNA in higher plants is in general remarkably constant between a wide range of species, many legumes show a large number of rearrangements, and in other plants there are simpler alterations (Palmer and Thompson 1982). These include an inversion of some 20 kbp in the genomes of a number of monocotyledonous plants, including maize, wheat and barley but not *Spirodela* (Palmer and Thompson 1982; Howe et al. 1983; Oliver and Poulsen 1984; de Heij et al. 1983). This region includes the genes for the alpha and proton-translocating subunits of ATP synthase. In mung bean an inversion of 50 kbp has occurred, one end of which is close to one of the endpoints of the 20 kbp inversion of wheat and maize (Palmer and Thompson 1982). An apparently very similar inversion has taken place in *Oenothera,* which is well separated from mung bean in evolution (Herrmann et al. 1983). This suggests that certain regions of the chromosome may be particularly recombinogenic. Little is known about the mechanism of recombination, but examination of the DNA sequence at those points where it has taken place may shed light on this problem.

It is not possible to identify accurately the endpoints of the frequent inversions of the two single copy regions relative to one another, as they are within the large inverted repeated sequences which are typically at least 20 kbp in size. It should however be possible to define more accurately the endpoints of the occasional inversions in the single copy regions. If these inversions have taken place relatively recently, it is quite likely that there will be readily discernible vestiges of whatever sequences, if any, have caused them. This is particularly likely to be so for the inversion in wheat, since it is not present in all monocots. Therefore it was decided to study this inversion, with particular regard to the presence of repeated sequences, by DNA sequence analysis. The ability of a putative recombinogenic sequence to act in vivo in a site-specific recombination system is demonstrated.

Materials and methods

Preparation and digestion of DNA. Plasmids pTac(P4) and pTac(P10) contain respectively the fourth and tenth largest *PstI* fragments of wheat *(Triticum aestivum* cv. Mardler) chloroplast DNA inserted into the *PstI* site of pBR322. The fragments are 12.6 kbp and 5.2 kbp respectively in size (Bowman et al. 1981). Plasmid DNA was prepared as described previously (Howe et ai. 1982). Restriction enzymes were obtained from B.R.L. (UK) and Amersham International and used according to the manufacturers' instructions. Electrophoresis in agarose and polyacrylamide gels and recovery of DNA from gels by electrophoresis onto DEAE-cellulose paper was carried out as described previously (Howe et al. 1982).

Nucleic acid hybridisation. Restriction fragments were transferred to PALL "Biodyne A" membrane (Pall Process Filtration Ltd., Portsmouth, UK) according to the manufacturers' instructions. It was found that polyacrylamide gels shrank significantly during the processing prior to blotting, so it was necessary to record the dimensions before and after processing. Nick translation and hybridisation were carried out as described previously (Howe et al. 1983) except that the hybridisation temperature used was 32 °C (T_m -30 °C). After washing, blots were subjected to autoradiography for 2 days with a Cronex Lightning Plus intensifying screen and Kodak Xomat-S film.

Construction and identification of clones for DNA sequencing. The 400 bp *SalGI-PstI* and other smaller fragments of P10 were cloned into the *SalGI*, *PstI* and *HincII* sites of M13 strains mp8 and mp9 as appropriate. The 3.2 kbp *BamHI* fragment of P4 containing the homology to P10 was digested with *Sau3A* and *AluI* and the fragments cloned into the *BamHI* and *HincII* sites of M13mp8. A number of single stranded DNA preparations were obtained and $1 \mu l$ samples spotted onto "Biodyne A" membrane. This was treated according to the manufacturer's instructions and probed with nick-translated P10 DNA. Clones showing a strong hybridisation signal were sequenced, along with *Sau3A-PvuII* and *Alu* subclones from the 3.2 kbp *BamHI* fragment. DNA sequence determination was carried out by the chain termination procedure, using $(\alpha)^{35}$ S]thio)dATP (Amersham International, 7.8 mCi/ml, 410 Ci/mmol in aqueous solution) and buffer gradient gels as described (Biggin et al. 1983).

Integration of bacteriophage lambda. The 400 bp *SalGI-PstI* fragment of P10 cloned in pUC8 was introduced into the *E. coli* strain *HfrH* \triangle attB.B' by transformation in the presence of 100 mM CaCl₂ at 4° C, selecting for the presence of ampicillin resistance. Media used were as described elsewhere (Schrenk and Weisberg 1975). Stocks of phage strains *hcI857 \$7 xis6 nin5 b515 b519, Xc h80 del9* and *Xclb2* were prepared on E. *coli* CSH25 by confluent lysis (Schrenk and Weisberg 1975). The transformed *HfrH∆attB.B* strain was prepared for lysogenisation (Schrenk and Weisberg 1975) and aliquots containing ca. 5 x 10^8 bacteria were mixed with ca. 5 x 10^8 $\lambda cI857 S7$ *xis6 nin5 b515 b519* phage and incubated in a final volume of 0.5 ml at 32 °C for 15 min. Hundredfold dilutions were made and 0.4 ml samples of these were taken. To these were added

150 μ g ampicillin, 15 μ mol MgSO₄ and ca 5 x 10⁸ of each of the killing phage. Then 3 ml molten soft agar was added and the whole poured onto an LB agar plate containing ampicillin at 50 μ g/ml and MgSO₄ at 5 mM. Control plates with bacteria or one or more phage omitted were also included. The frequency of lysogenisation was found to be about 1%. After incubation at 30 °C for 20 h, separate pools of lysogens were induced by homogenising the soft agar overlay into LB medium with ampicillin followed by incubation at 32° C, 37° C and 32° C. The final phage titre was found to be ca 10^9 pfu/ml. Phage able to transduce ampicillin resistance were selected by addition of ca. 10⁸ phage to ca. 5 x 10⁷ *HfrH* \triangle attB.B' bacteria (not carrying the pUC8-chloroplast DNA plasmid) prepared for phage infection as described (Schrenk and Weisberg 1975). This strain was used since the deletion of the primary attachment site reduces the likelihood of transducing phage reintegrating, thus enhancing their stability as plasmids. The mixture was incubated in a final volume of 0.5 ml for $15'$ at 32° C, 1 ml LB broth was added and, after a further hour at 32 °C, cells were spun down in an Eppendorf microcentrifuge, washed with fresh LB broth and plated onto LB agar with ampicillin before overnight incubation at 32 °C. Controls were included which lacked bacteria or phage. Ampicillin resistant colonies were picked off and grown in LB medium with ampicillin. DNA was prepared and samples digested with *HindlII,* followed by electrophoresis in a 6% polyacrylamide gel. When a novel 350 bp *HindlII* fragment was detected, a larger digest was carried out and the fragment purified, cloned and sequenced.

Results

Localisation of the endpoints and associated repeated sequences

In the experiments of Howe et al. (1983), to map the wheat ATP synthase alpha subunit gene, hybridisation to spinach chloroplast DNA indicated that the wheat chloroplast DNA *PstI* fragments P4 and P10 contained the endpoints of a large inverted region. A *SalGI-PstI* digest of the wheat chloroplast DNA fragment P10 cloned in pBR322 was transferred from an acrylamide gel onto "Biodyne A" membrane and probed with nicktranslated purified P4 fragment. The result (Fig. 1) showed the presence of homology between P4 and P10 which was restricted to a 400 bp *SalGI-PstI* fragment of P10. The DNA sequence of this fragment was determined (Fig. 2). A blot of an agarose gel containing the products of a *BamHI* digest of the P4 fragment cloned in pBR322 was probed with purified P10 fragment. The result indicated (Fig. 1) that the homology was restricted to a single 3.2 kbp *BamHI* fragment, which previous studies had shown to contain one endpoint of the 20 kbp inversion (Howe et al. 1983). This suggested that the cross-hybridising regions are associated with the endpoints. The former are marked in Fig. lB. The 3.2 kbp *BamHI* fragment was purified, digested with *Sau3A* and AluI and the resulting fragments cloned into M13mp8. A dot blot of 120 single-stranded DNA preparations from

C. J. Howe: Endpoints of an inversion in wheat chloroplast DNA 141

Fig. IA, B. Cross-hybridisation between fragments P4 and P10. A *Track a* shows a polyacrylamide gel of the products of digestion with *PstI* and *SalGI* of the plasmid pTac(P10) and *track b* shows the results of probing a blot of this digest with nick-translated purified P4 DNA. *Track c* shows an agarose gel of the products of digestion with *BamHI* of the plasmid pTac(P4) and *track d* shows the results of probing a blot of this digest with nick-translated purified P10 DNA. The regions cross-hybridising are shown in relation to the rest of the chloroplast genome in B, *aPstI* restriction map also showing the large inverted repeats and the position and polarity of the ATPsynthase alpha subunit gene

this cloning was probed with the nick-translated P10 DNA and those preparations which hybridised were sequenced. Various other clones were constructed as shown in Fig. 2A, to confirm the sequences shown in Fig. 2B and 2C. The sequences were compared with sequences from tobacco chloroplast DNA (Ohme et al. 1984; Deno et al. 1984), which is known to be colinear with chloroplast DNA from spinach and a number of other dicots (Fluhr and Edelman 1981), using the program DIAGON on a BBC Microcomputer.

Identification of the site of recombination

The region sequenced from the P10 fragment contains the second exon of the tRNA Gly (UCC) gene and the last few nucleotides of the intron. The sequence at the intron-exon boundary is identical to that in tobacco. There is also a putative initiator tRNA f-met (CAU) gene which shows high homology to the tobacco tRNA f-met (CAU) molecule. This is flanked by two copies of a repeated sequence of some 70 bp. This same 70 bp sequence is present in the region from P4 which was shown to cross-hybridise with P10. The average homology between these three copies is 85%. Another feature of this part of P4 is a sequence partially homologous to the f-met tRNA gene. This could form a similar structure, although there is less base-pairing possible in the

main stem and it may perhaps represent a pseudogene. It is much less homologous to the tobacco structure (72%) than is the gene in P10 (89%). It bears little homology to the wheat elongator met-tRNA gene which is elsewhere on the chromosome (Howe et al. 1985). This f-met tRNA-like structure is shown in Fig. 3, along with the predicted structure of the f-met tRNA from P10, and a tRNA Arg (UCU) from P4, identified by its 100% homology to the tobacco equivalent. Also contained in the region of $P4$ sequenced is the 3' end of the gene for the alpha subunit of ATP synthase. A simplified map of the P4 and P10 regions is presented in Fig. 4 (bottom line). Comparison with the position of the corresponding genes in tobacco allows identification of the endpoints of the wheat inversion. Split tRNA Gly and tRNA f-met genes are over 20 kbp apart in tobacco (Ohme et al. 1984; Seyer et al. 1981), but close together in wheat P10, separated only by one copy of the 70 bp repeated sequence. This indicates that the 70 bp repeat in P10 occupies one inversion endpoint. The other endpoint would be expected to be close upstream from the tRNA Arg gene in P4, since this gene is well separated from the tRNA Gly gene in wheat, but is the next gene to it in tobacco. This position is occupied in wheat by another copy of the 70 bp repeated sequence. This suggests that the inversion was caused by recombination across a repeat which has since been preserved. It is possible that the recombination also led to a duplication of the f-met

TTATTTGTTA CTGAAGAGAA AACCGTTCCA GCTG alpha

C

Fig. 3A-C. Predicted tRNA structures. The predicted structures (prior to modification) are shown of A a putative f-met tRNA from P10, B a putative arg (UCU) tRNA from P4 and C the f-met tRNA-like structure from P4. In A and B the -CCA is assumed to be added post-transcriptionally. A $-CCA$ triplet is already present in the DNA sequence for C

Fig. 4. Postulated evolution of wheat chloroplast DNA *(bottom line)* from a dicot-like ancestral form *(top line). Heavy arrows* indicate the position and direction of transcription of genes for gly (UCC), arg (UCU) and f-met (CAU) tRNAs (G, R, M) and the alpha subunit of ATP synthase. The *large light arrows* indicate the position of the main inverted repeat sequences and the *short light arrows* indicate the position and direction (arbitrarily defined) of the 70 bp repeats. Postulated crossovers are indicated between ancestral, intermediate and present forms, the extent of the latter being uncertain from these data

tRNA gene, and was thus more complicated than a simple crossover within homologous sequences. It may have involved an element of gene conversion.

However, this recombination would only result in inversion if the repeats were in an inverted configuration. Currently they are in a direct configuration. The direction of transcription of the wheat tRNA Gly gene (which has not been involved in the 20 kbp inversion)

with respect to the nearer arm of the large inverted repeat is also opposite to that of the tobacco gene. This indicates that the wheat tRNA Gly gene has been involved in a second inversion. The extent of this second inversion is not clear, but it has also involved the tRNAfmet gene placed close to the tRNA Gly gene by the first. Thus, before the second inversion, the repeated sequences would indeed be present in the inverted con-

[•] Fig. 2A-C. DNA sequence of cross-hybridising regions of P4 and P10. The sequencing strategy is shown in A. The sequence of part of P10 is shown in B and part of P4 in C. Regions corresponding to tRNAs, identified by homology with tobacco sequences (Ohme et al. 1984; Deno and S ugiura 1984) are *boxed* and *marked,* as are the repeated elements and the 3' terminus of the gene for the alpha subunit of ATP synthase. Sequences within the repeats which resemble the lambda *att* site are *underlined*

Fig. 5. Sequence around the *Natt*-site (Sanger et al. 1982) (boxed), part of the repeated sequence shown in Fig. 2 B and one λ -chloroplast DNA junction

figuration expected if they had been responsible for the first. This is summarised in Fig. 4. Because the precise extent of the second inversion is unknown, the exact size of the first cannot be determined.

While this manuscript was in preparation, the sequence of a region covering one endpoint (that shown in Fig. 2B) was also reported by Quigley and Weil (1985). They also postulate a multi-stage derivation of the wheat gene arrangement.

Integration of bacteriophage lambda

The 70 bp repeated sequence which appears to be associated with the main inversion contains a sequence which is remarkably similar to part of the 15 bp core sequence of the bacteriophage lambda attachment site 5'-GCTTTTTTATACTAA-3' (Landy and Ross 1977). Two of the three copies of the chloroplast DNA 70 bp repeat have 9 base pairs identical to the lambda core sequence, the third has 10. This homology is higher than in many documented secondary attachment sites on the *E. coli* chromosome (Nash 1981). Experiments were carried out to see if this sequence could act in vivo as an attachment site for lambda. The 400 bp *SalGI-PstI* fragment from P10 was chosen, since it contains 2 copies of the sequence in question, including the one most similar to the lambda attachment site. The fragment was cloned into pUC8 and then introduced by transformation into *E. coli* strain $HfrH\Delta at tB. B'$. The primary attachment site of lambda is deleted in this strain, increasing the frequency of lambda integration into secondary attachment sites (Shimada et al. 1972). The transformed strain was then infected with $\lambda cI857$ *\$7 xis6 nin5 b515 b519.* The last 3 mutations are deletions which amount to about 10 kbp and enhance the frequency of packaging of flanking host sequence into the lambda particle (Schrenk and Weisberg 1975). Lysogens were selected by superinfection with the strains *Xc h80 del9* and *XcI b2.* Lysogens are immune to killing by either of these phage, lambda resistant mutants are killed by the first, and wild-type cells are killed by both. A number of independent pools oflysogens were induced by heat inactivation of the $cI857$ repressor and the phage obtained were screened for the ability to transduce ampicillin resistance (as a result of having first integrated into the chloroplast DNA-pUC8 Ap^R hybrid plasmid). Several ampicillin resistant transductants produced from phage from different inductions were selected for further study. Plasmid DNA was isolated from them and analysed by restriction endonuclease digestion. The sizes of restriction fragments generated by digestion of DNA from the starting lambda strain with *HindlII* are 20.3, 9.46, 4.26, 3.83, 2.36, 0.59 and 0.15 kbp (Britton et al. 1983). Integration of lambda at the site closest in sequence to *art* should cause the 9.46 kbp restriction fragment to be replaced by one of 0.35 kbp and one of 12.2 kbp. Plasmid DNA isolated from ampicillin resistant transductants was digested with *HindlII* and subjected to polyacrylamide gel electrophoresis to screen for the appearance of the 0.35 kbp fragment. A number of fragments gave a restriction digest corresponding to the starting phage alone. It is not clear why such strains should be ampicillin resistant (the frequency of ampicillin resistant transductants being at least 1,000 times greater than the frequency of spontaneously occurring ampicillin resistant cells), but this effect has been reported before (Schrenk and Weisberg 1975). Some plasmids gave a restriction digest consistent with the integration of the phage into a site in the pUC8 vector. However, plasmids arising from 2 out of the 7 pools of lysogens gave a ca. 350 bp *HindlII* fragment. This fragment was purified, cloned into M13mp8 and sequenced. Part of the sequence is given in Fig. 5 and shows that lambda has indeed integrated at the exact site predicted by comparison with the normal integration mechanism. In both cases the *att* site used was the one closest in sequence to the phage *att* site. Interestingly, the integration appears to have caused the insertion of an extra T residue.

Discussion

A 20 kbp inversion in chloroplast DNA occurring during the evolution of the monocots has been analysed. The endpoints are marked by a 70 bp repeated sequence, and it seems likely that the inversion was brought about by recombination across these sequences, although one end may be associated with some duplication to yield the extra f-met tRNA "gene". If this repeat was responsible, and has been preserved in wheat, one would also expect to find it in other plants which show a similar gene organisation. Preliminary experiments using Southern hybridisation indicate that this is so in maize (C. J. Howe, unpublished), and sequencing studies in barley have indicated the presence of at least two copies of the 70 bp sequence (Oliver and Poulsen 1984). (It is not found in the corresponding regions of spinach or tobacco, which have not undergone the rearrangement.)

A second inversion has also happened subsequently, including one of the endopints of the first. If chloroplast DNA rearrangements generally are caused by recombination across repeated sequences, then this second inversion should be similarly demarcated. This is currently under investigation. It will be interesting to see if any monocots are found to have only the first inversion.

Since the repeated elements are now in a direct configuration, recombination across them would lead to fragmentation of the chloroplast DNA molecule, which would presumably result in loss of chloroplast viability, unless both fragments contained origins of replication. Large-scale deletions have been observed in the chloroplast DNA of wheat plants regenerated from pollen, and this has been proposed as a possible basis of maternal inheritance in chloroplasts (Day and Ellis 1984). Many of these deletions appear to include at least the region between the 70 bp repeats. There is, however, no evidence of high deletion rates in chloroplasts under normal circumstances, so perhaps high levels of recombination occur only at certain times during plant development. This is difficult to reconcile with the postulated frequent recombination across the two halves of the inverted repeat. In plant mitochondria, by contrast, there does appear to be continuous recombination across repeated sequences around the genome, generating a wide range of partially deleted molecules (Palmer and Shields 1984; Lonsdale et al. 1984).

The results suggest then, that there is a requirement for sequence homology for chloroplast recombination to take place, although there may also be a sequence specificity. Further study of other evolutionary rearrangements is required to see if the association with repeated sequences is a general feature. It is interesting that the putative recombinogenic repeated sequence should display sufficient homology to the recognition sequence of the site-specific recombination pathway for lambda integration for it to function as a lambda *art* site in vivo. It will be especially interesting to see if such a sequence is present at other sites of recombination in chloroplasts. The presence of specific integration sites in apparently non-coding regions of the chloroplast genome may be useful in genetic manipulation. A similar sequence also occurs in the \$2 episome of maize mitochondria close to the site of recombination with the mitochondrial chromosome (Levings and Sederoff 1983; Schardl et al. 1984). Experiments are in progress to test if the \$2 *att*like sequence can also function as such in vivo.

Acknowlegements. I am grateful to Mr. S. E. Haywood for technical assistance, and to Drs. C. M. Bowman, T. A. Dyer, J. C. Gray, M. C. Jones-Mortimer and D. C. McN. Moore for helpful discussions and advice. Dr. Jones-Mortimer kindly supplied the strains used for the phage integration experiments. This work was supported by the Nuffield Foundation and Corpus Christi College, Cambridge.

References

- Biggin MD, Gibson TJ, Hong GF (1983) Proc Natl Acad Sci USA 80:3963-3965
- Bowman CM, Koller B,Delius H, Dyer TA (1981) Mol Gen Genet 183:93-101
- Britton P, Boronat A, Hartley DA, Jones-Mortimer MC, Kornberg HL, Parra F (1983) J Gen Microbiol 129:349-358
- Day A, Ellis THN (1984) Cell 39:359-368
- Deno H, Sugiura M (1984) Proe Natl Acad Sci USA 81:405 408
- Fluhr R, Edelman M (1981) Nucleic Acids Res 9:6841-6853
- Gillham NW (1978) Organelle heredity. Raven Press, New York
- deHeij HT, Lustig H, Moeskops D-JM, Bovenberg WA, Bisanz C, Groot GSP (1983) Curt Genet 7:1-6
- Herrmann RG, Westhoff P, Alt J, Winter P, Tittgen J, Bisanz C, Sears BB, Nelson N, Hurt E, Hauska G, Viebrock A, Sebald W (1983) In: Ciferri O, Dure L (eds) Structure and function of plant genomes. Plenum Press, New York, pp 145-153
- Howe CJ, Bowman CM, Dyer TA, Gray JC (1982) Mol Gen Genet 186:525-530
- Howe CJ, Bowman CM, Dyer TA, Gray JC (1983) Mol Gen Genet 190:51-55
- Howe CJ, Fearnley IM, Walker JE, Dyer TA, Gray JC (1985) Plant Mol Biol 4:333-345
- Kolodner R, Tewari KK (1979) Proc Natl Acad Sci USA 76: $41 - 45$
- Landy A, Ross W (1977) Science 197:1147-1160
- Levings CS, Sederoff RR (1983) Proc Natl Acad Sci USA 80: 4055 -4059
- Lonsdale DM, Hodge TP, Fauron CM-R (1984) Nucleic Acids Res 12:9249-9261
- Nash HA (1981) Annu Rev Genet 15:143- 167
- Ohme M, Kamogashira T, Shinozoki K, Sugiura M (1984) Nucleic Acids Res 12:6741-6749
- Oliver RP, Poulsen C (1984) Carlsberg Res Commun 49:647- 673
- Palmer JD, Thompson WF (1982) Cell 29:537-550
- Palmer JD, Osorio B, Watson JC, Edwards H, Dodd J, Thompson WF (1984) In: Thornber JP, Staehelin A, Hallick RB (eds) Biosynthesis of the photosynthetic apparatus: molecular biology, development and regulation. Liss, New York, pp 273-283
- Palmer JD, Shields CR (1984) Nature (London) 307:437-440
- Quigley F, Weil JH (1985) Curr Genet 9:495-503
- Sanger F, Coulson AR, Hong GF, Hill DF, Petersen GB (1982) J Mol Biol 162:729-773
- Scbardl CL, Lonsdale DM, Pring DR, Rose KR (1984) Nature (London) 310:292-296
- Schrenk WJ, Weisberg RA (1975) Mol Gen Genet 137:101-107
- Seyer P, Kowallik KV, Herrmann RG (1981) Curr Genet 3:189- 204
- Shimada K, Weisberg RA, Gottesmann ME (1972) J Mol Biol 63:483-503
- Communicated by C. J. Leaver

Received July 1, 1985