

# A transposon with an unusual arrangement of long terminal repeats in the green alga *Chlamydomonas reinhardtii*

Anil Day, Michèle Schirmer-Rahire,  
Michael R.Kuchka, Stephen P.Mayfield<sup>1</sup> and  
Jean-David Rochaix

Departments of Molecular and Plant Biology, University of Geneva,  
30, Quai Ernest Ansermet, CH-1211 Geneva 4, Switzerland

<sup>1</sup>Present address: Department of Molecular Biology, Scripps Clinic and  
Research Foundation, 10666 North Torrey Pines Road, La Jolla, CA  
92037, USA

Communicated by J.D.Rochaix

We have isolated a 5.7-kbp dispersed moderately repeated DNA sequence (TOC1) from the mutant OEE1 gene of the *Chlamydomonas reinhardtii* strain FUD44. The copy number (2 to over 30) and genomic locations of TOC1 elements vary widely in different *C.reinhardtii* strains. Our standard laboratory photosynthetic strain exhibits a high degree of TOC1 instability during short periods of mitotic growth. TOC1 appears to be a retrotransposon: it contains LTRs and an oligonucleotide stretch that corresponds to a conserved pentapeptide of reverse transcriptase. TOC1 is an unusual retrotransposon: it is not flanked by a target site duplication in the OEE1 gene, the left end of TOC1 only contains a fraction of the LTR the remainder of which is present at its right end and TOC1 does not start with a 5' TG and end with a 3' CA. In most cases, TOC1 excision leaves behind a complete solo LTR sequence (577 bp) and in one case a deleted solo LTR sequence (191 bp). Solo LTR sequences form a separate family of repeated sequences in most of the strains tested.

**Key words:** *Chlamydomonas reinhardtii*/excision/integration/  
photosynthetic mutant/retrotransposon

## Introduction

Transposition, a recombination mechanism that does not rely on substantial sequence similarity between target and donor DNAs, is the distinguishing, although not necessarily the most frequent, mode of movement of transposable elements. Movement of transposable elements between different genomic locations is often associated with new mutations. This property of these elements has important evolutionary consequences and has also been very useful for their isolation from a wide range of species. Representative elements have been found in the plant, bacterial and animal kingdoms. In most species, transposable elements are part of the repetitive fraction of the genome.

Recently it has become clear that transposable elements can be divided into two groups based on a fundamental difference in transposition mechanism. In the first group, which includes prokaryotic and most plant transposable elements, the transposition intermediates appear to be

composed of DNA. The second group is comprised of elements (retrotransposons) that appear to transpose via an RNA intermediate (Boeke *et al.*, 1985; Weiner *et al.*, 1986). Within this group two subgroups have been delineated (Weiner *et al.*, 1986). Members of the first subgroup share structural similarities with the integrated DNA stage (proviral DNA) of animal retroviruses (Varmus, 1982; Finnegan, 1983) and have been called retrotransposons (Boeke *et al.*, 1985) or viral-like retrotransposons (Weiner *et al.*, 1986). The common origin of retroviruses and retrotransposons (Temin, 1980; Baltimore, 1985) is strengthened by the observation that two retrotransposons, *copia* from *Drosophila melanogaster* (Shiba and Saigo, 1983) and Ty from *Saccharomyces cerevisiae* (Garfinkel *et al.*, 1985), have been found associated with viral-like particles. The second subgroup of retrotransposons (non-viral retrotransposons) is composed of a heterogeneous mixture of transcribed dispersed repetitive DNA sequences that appear with varying degrees to be passive substrates of the cell's transpositional machinery (Weiner *et al.*, 1986). Non-viral retrotransposons produce a target site duplication of variable length and are often terminated by a 3' oligo(dA) stretch (Weiner *et al.*, 1986).

We have been using biochemical and genetic approaches to study the components required for photosynthesis in the green alga *Chlamydomonas reinhardtii*. Genetic analyses identify components necessary for photosynthesis that are not accessible to biochemical fractionation methods. For example, a number of nuclear loci that affect chloroplast gene expression have recently been characterized in our laboratory (Choquet *et al.*, 1988; Kuchka *et al.*, 1988). The isolation of these loci may shed some light on their function. Transposon tagging (Bingham *et al.*, 1981) provides a route for isolating nuclear loci that encode unidentified products.

A large insertion in the oxygen-evolving enhancer 1 (OEE1) protein gene (Mayfield *et al.*, 1987) of the photosynthetic mutant FUD44 indicated that *C.reinhardtii* may contain an active transposable element. The OEE1 protein is required for oxygen evolution by the water-splitting complex that supplies photosystem II with electrons. FUD44 is deficient in the OEE1 protein and will only grow on a medium containing a reduced carbon source such as acetate. Revertants of FUD44 that are able to fix CO<sub>2</sub> and as a result grow on minimal medium have been isolated at a frequency of 10<sup>-5</sup>/cell. Unlike FUD44, these revertants accumulate detectable amounts of OEE1 mRNA (Mayfield *et al.*, 1987).

Here we describe the characterization of the mutant and revertant OEE1 protein genes that we have isolated from FUD44 and two revertants of FUD44. The mutation in FUD44 results from the insertion of a 5.7-kb moderately repeated sequence into the second intron of the OEE1 gene. Reversion results from incomplete excision of the element from the OEE1 locus. Although the repeated element (TOC1, Transposon Of *Chlamydomonas*) appears to fall into

the retrotransposon class of mobile elements, its long terminal repeat (LTR) arrangement is unlike that of any other retrotransposon: the left end of TOC1 only contains a fraction of the LTR (217-bp repeat) the remainder of which (237-bp repeat) is present at its extreme right end and is separated from the complete right LTR (tandem 237-/217-bp repeats) by a unique 123-bp sequence. In addition, we propose that a variation of the retroviral integration mechanism is responsible for TOC1 breaking the 5' TG 3' CA rule observed by most retrotransposon LTRs.

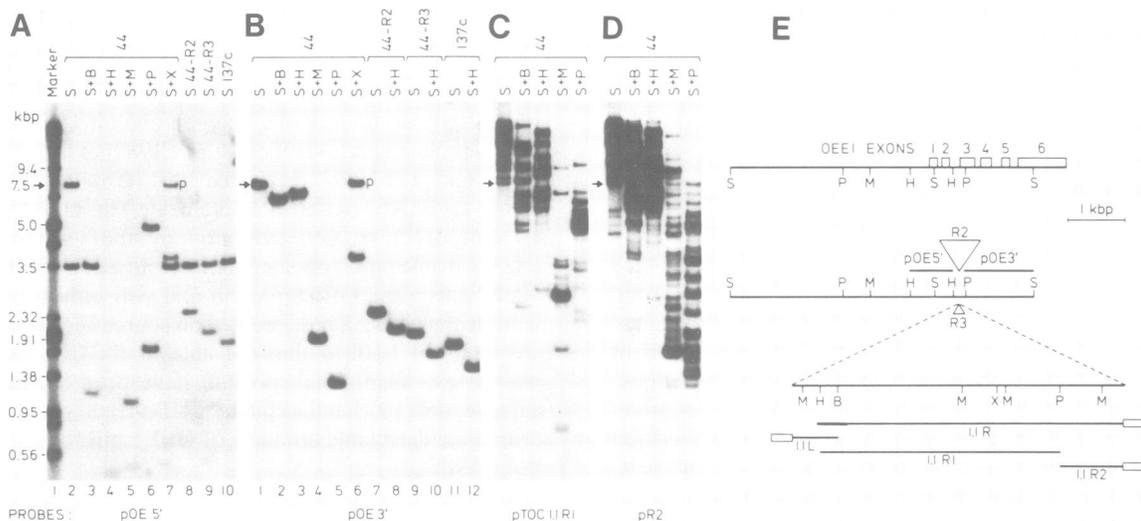
## Results

### Isolation of the mutant and revertant OEE1 protein genes

The region containing the alteration in the OEE1 gene from the mutant strain FUD44 lies within a 193-bp *HindIII*–*PstI* fragment that contains 144 bp of intron 2 followed by 49 bp of exon 3 (Mayfield *et al.*, 1987). The size of the insertion in the mutant OEE1 gene was determined by using *SalI*, an enzyme that does not cut within the foreign insert, and hybridizing *SalI* digests of FUD44 DNA on Southern blots with probes that lie 5' and 3' to the point of insertion. The wild-type 1.8-kbp *SalI* fragment from 137c (Figure 1A, lane 10 and 1B, lane 11) bearing most of the OEE1 gene is replaced by a 7.5-kbp *SalI* fragment in digests of FUD44 DNA (Figure 1A, lane 2 and 1B, lane 1). Two classes of reversion events have been detected. In 38 out of 39 revertant (14 spontaneous revertants and 25 revertants obtained by UV mutagenesis) DNAs examined, the mutant 7.5-kbp band is replaced by a 2.4-kbp *SalI* fragment. The spontaneous revertant FUD44-R2 is an example of this class of revertants (Figure 1A, lane 8, and 1B, lane 7). Digests of DNA from the spontaneous revertant FUD44-R3, the only representative of the second class of revertants, contain a 2.0-kbp OEE1 *SalI* fragment (Figure 1A, lane 9 and 1B, lane 9).

We initially attempted to clone the OEE1 gene from partial *Sau3a* digests of FUD44 DNA using the  $\lambda$  EMBL4/*Escherichia coli* NM539 (Kaiser and Murray, 1985) vector/host cloning system. Using an OEE1 cDNA probe that spans the point of insertion, six different recombinant phages were isolated from the FUD44/ $\lambda$  EMBL4 library. The inserts of all six phages stopped short (3') of the alteration in the mutant OEE1 gene. To overcome the possibility that the ends of the intact foreign element are unstable in *recA*<sup>+</sup> *E.coli* cells we decided to clone the mutant OEE1 gene under *recA*<sup>-</sup> conditions. Furthermore, the altered region of the mutant OEE1 gene from FUD44 was cloned as two overlapping restriction fragments in the plasmid pAT153: a 1.2-kbp *SalI*–*BamHI* fragment (pTOC1.1L, Figure 1A, lane 3) and a 6.7-kbp *HindIII*–*SalI* fragment (Figure 1B, lane 3). The two fragments have a 380-bp overlap. A *HindIII*–*PstI* fragment of 5.4 kbp (pTOC1.1R) was subcloned by partial *PstI* digestion of the 6.7-kb *HindIII*–*SalI* fragment. The insertions within the OEE1 genes from FUD44-R2 and R3 were cloned as 2.0- and 1.6-kbp *HindIII*–*SalI* fragments, respectively (Figure 1B, lanes 8 and 10). *HindIII*–*PstI* fragments of 0.77 kbp (pR2) and 0.38 kbp (pR3) were subcloned from the 2.0- and 1.6-kbp *HindIII*–*SalI* fragments.

The cloned fragments bearing the mutant OEE1 gene from FUD44 appear to be faithful representatives of the altered OEE1 gene since the restriction sites mapped in the cloned fragments (Figure 1E) are present and co-linear with those predicted from digests of FUD44 DNA (Figure 1A and B); this procedure verifies the *BamHI*, *HindIII*, *PstI*, *XhoI* and outermost *MluI* sites in the restriction map of the cloned fragments. The subclones derived from the 5.4-kbp *HindIII*–*PstI* fragment (pTOC1.1R) are defined in Figure 1E. A 4.1-kbp subcloned *HindIII*–*PstI* fragment (pTOC1.1R1) that lies internal to the insertion sequence hybridizes to multiple bands in a number of different digests of FUD44 DNA on Southern blots (Figure 1C). The



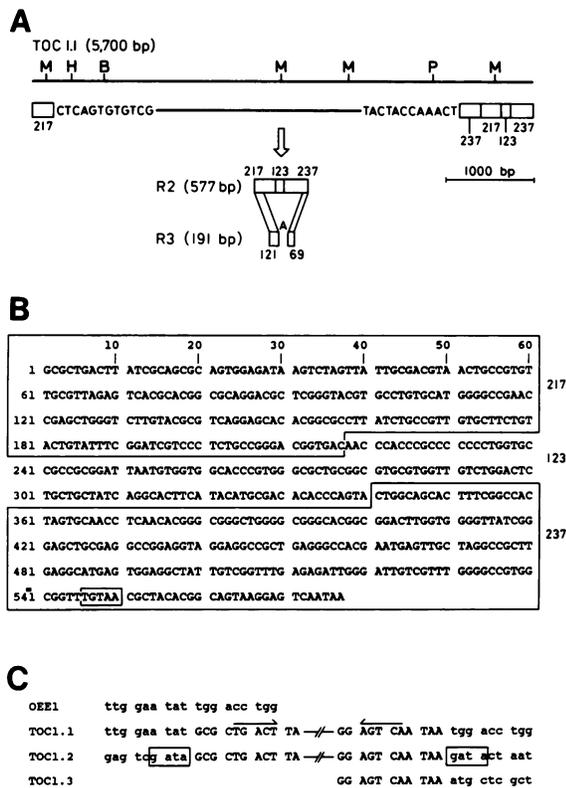
**Fig. 1.** Restriction enzyme analysis of the OEE1 genes from FUD44, FUD44-R2, FUD44-R3 and 137c. The hybridization probes used are indicated below each blot and their locations are shown in E. pOE3' was subcloned as an *EcoRI*–*SalI* fragment from a recombinant phage isolated from a *Sau3A* FUD44/ $\lambda$  EMBL4 library. pOE5' is a 0.7-kb fragment cloned from the 5' end of the 137c OEE1 gene. Restriction enzymes sites: B = *BamHI*, H = *HindIII*, M = *MluI*, P = *PstI*, S = *SalI*, X = *XhoI*. p = partial *XhoI* digestion product; 2.5  $\mu$ g of digested DNA were loaded per lane. The probe pR2 is a 770-bp *HindIII*–*PstI* fragment from FUD44-R2 that is composed of a 577-bp insertion into the wild-type 193-bp *HindIII*–*PstI* fragment of OEE1. The position of the mutant 7.5-kbp OEE1 *SalI* fragment in digests of FUD44 is indicated by arrows in A, B, C and D. In E, the OEE1 sequences in the recombinant pTOC plasmids containing the 1.2-kbp *SalI*–*BamHI* (1.1L), 5.4-kbp *HindIII*–*PstI* (1.1R), 4.1-kbp *HindIII*–*PstI* (1.1R1) and 1.36-kbp *PstI*–*PstI* (1.1R2) fragments are drawn as boxes.

insertion therefore corresponds to a moderately repeated family of DNA sequences that we have named TOC1 (see Discussion). In Figure 1C, TOC1 bands of <5.7 kbp are predominantly found in *Mlu*I and *Pst*I digests of FUD44 DNA, which indicates that most TOC1 elements contain internal *Mlu*I and *Pst*I sites. The cloned 0.77-kbp *Hind*III–*Pst*I OEE1 fragment from FUD44-R2 (pR2) must also contain a repetitive element since it hybridizes to multiple bands in digests of FUD44 DNA; under these hybridization conditions, the 193 bp of OEE1 sequences present in pR2 only hybridize to their single copy complementary sequences in genomic DNA (not shown).

**Sequence analyses of the altered regions of the OEE1 genes from FUD44, FUD44-R2 and FUD44-R3**

TOC1 integrates within the 193-bp *Hind*III–*Pst*I fragment of the wild-type OEE1 gene. Sequencing inwards from the 5' *Hind*III site and 3' *Pst*I sites revealed that the point of TOC1 insertion in FUD44 lies 12 bp 3' to the *Hind*III site.

A summary of the sequence data obtained by sequencing the ends of TOC1.1L and TOC1.1R, the entire 0.77-kbp and 0.38-kbp insertion-containing *Hind*III–*Pst*I fragments of FUD44-R2 and FUD44-R3, respectively, is shown in Figure 2A and B. Figure 2 reveals that TOC1 does contain LTRs but their arrangement is novel. The left end of TOC1 appears to only contain a fraction of the LTR, the remainder of which is found at the right end. There are no sequence differences between the left and right 217-bp repeats or the two 237-bp repeats. The 123-bp sequence that separates the far-right 237-bp repeat from the right 217-bp repeat appears to be present only once in TOC1.1. This is based on sequencing the inserts in pTOC1.1L and pTOC1.1R2 (see Figure 1E), and the absence of detectable hybridization between pR2, which contains the 123-bp sequence (see Figure 2A), and the insert of pTOC1.1R1 (not shown). The 577-bp and 191-bp insertions within the OEE1 genes from FUD44-R2 and FUD44-R3, respectively, result from internal deletions of the TOC1 element. The 237-, 217- and 123-bp sequences of TOC1.1 are identical to those found in R2. The terminal TOC1 sequences present in the OEE1 gene of FUD44-R3 differs from the [217-bp][123-bp][237-bp] LTR unit of TOC1 and the insertion in the OEE1 gene from FUD44-R2 in that it contains an adenosine residue in place of bases 122–508 of the LTR unit.



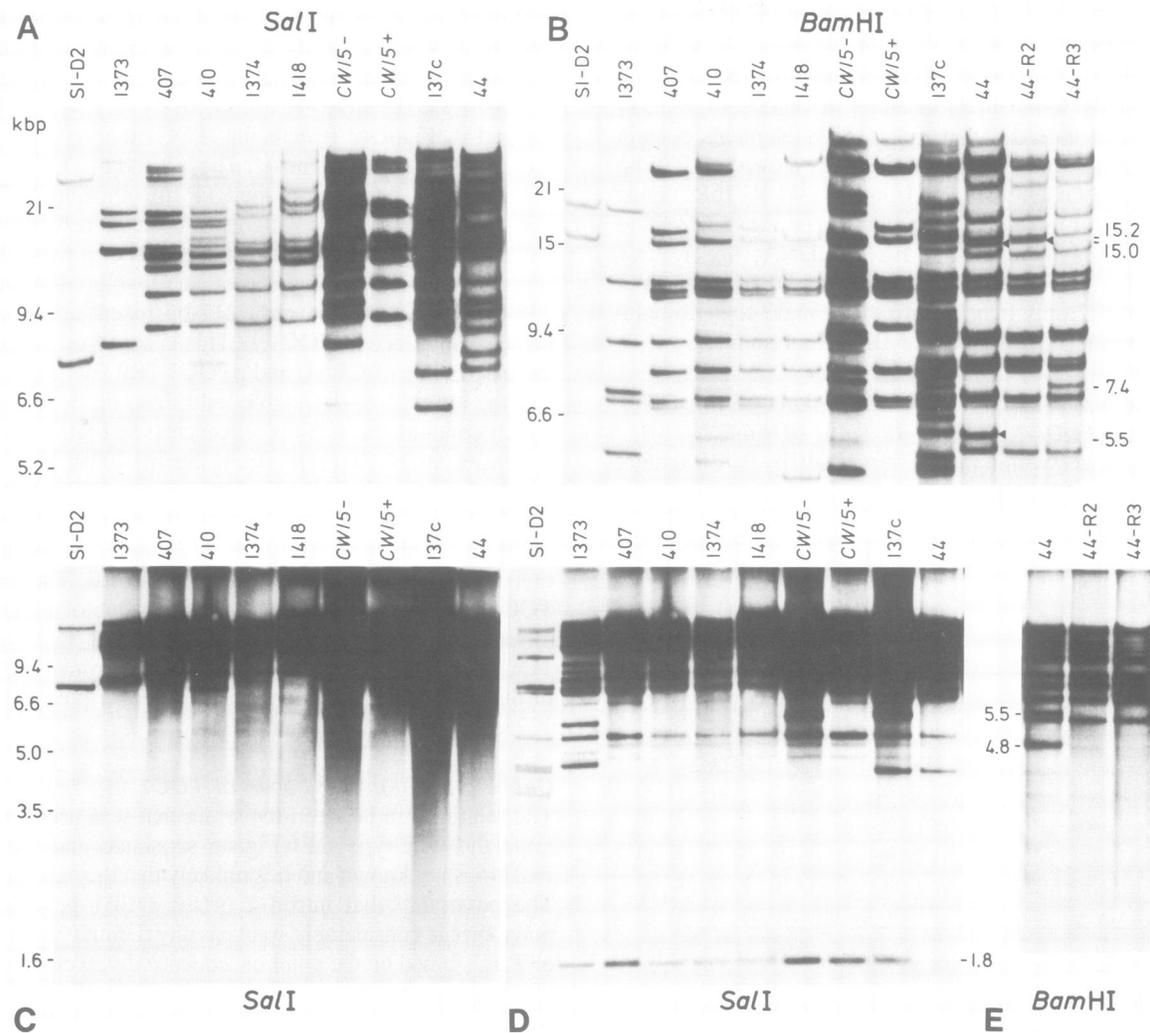
**Fig. 2.** Sequence analysis of the cloned OEE1 fragments from FUD44, FUD44-R2 and FUD44-R3. (A) The structure of the foreign DNA inserts in the OEE1 genes of FUD44 (TOC1.1), FUD44-R2 and FUD44-R3. The 237-, 217- and 123-bp sequences of TOC1.1 are identical to those found in R2. Within TOC1.1 there are no sequence differences between the left and right 217-bp repeats or the two 237-bp repeats. The insertion within the OEE1 gene of FUD44-R3 contains an adenosine residue in place of bases 122–508 found in the insertion within the OEE1 gene of FUD44-R2. The terminal 12 bases of the internal region of TOC1.1 are shown. Restriction enzyme sites located within TOC1.1 include B = *Bam*HI, H = *Hind*III, M = *Mlu*I, P = *Pst*I. (B) DNA sequence of the [217-bp][123-bp][237-bp] unit or solo LTR found in the OEE1 gene of FUD44-R2. TGTA, a likely component of the *C. reinhardtii* polyadenylation signal, is boxed (Silflow *et al.*, 1985; Goldschmidt-Clermont and Rahire, 1986). (C) Defining the terminal bases of TOC1. TOC1 sequences are shown in upper-case letters while flanking regions are represented in lower-case letters. A five-base inverted repeat sequence in TOC1 is indicated by arrows. A four-base duplication that flanks TOC1.2 is boxed.

**Defining the terminal bases of TOC1**

The relationship between the strain that was used to determine the wild-type OEE1 gene sequence and the strain FUD44 is not known and it is unlikely that they are isogenic. The possibility that intron 2 of the OEE1 gene may be polymorphic considerably weakens any definition of the ends of TOC1 based on the point of sequence divergence between the wild-type and mutant OEE1 genes. In order to be more certain of the terminal bases of TOC1, we cloned at random two more members of the TOC1 family from FUD44 (TOC1.2 and TOC1.3) as *Hind*III–*Sal*I fragments in pAT153. Both ends of TOC1.2 were sequenced since this element lacks an internal *Hind*III site. Only the right end of TOC1.3 was sequenced since the cloned fragment stops at an internal *Hind*III site located at the same position as the *Hind*III site in the TOC1 insertion at the OEE1 locus. Alignment of the terminal sequences of the TOC1 insertion in OEE1 (TOC1.1) with that of TOC1.2 and TOC1.3 shows that the sequences are identical up to the point at which OEE1 gene sequences are encountered (Figure 2C). This strongly suggests that the end points of TOC1 defined by a comparison of the wild-type OEE1 gene with the TOC1 disrupted OEE1 gene are correct. Figure 2C also shows that whilst TOC1.2 is flanked by a four-base duplication, TOC1.1 does not generate a target duplication at the OEE1 locus. It is conceivable that TOC1.1 produced a four-base duplication in OEE1 that was subsequently deleted, but in our opinion this is unlikely because in the absence of any apparent phenotypic consequences (the OEE1 gene will remain mutant) the deletion mechanism invoked would have to be extremely precise; it would have to remove the four-base duplication while preserving both the termini of TOC1 and the adjacent OEE1 sequences that are normally found juxtaposed in the wild-type gene.

**Copy number variation of TOC1 between strains**

The number of bands that strongly hybridize to an internal TOC1 probe (TOC1.1R1) in DNA digested with an enzyme



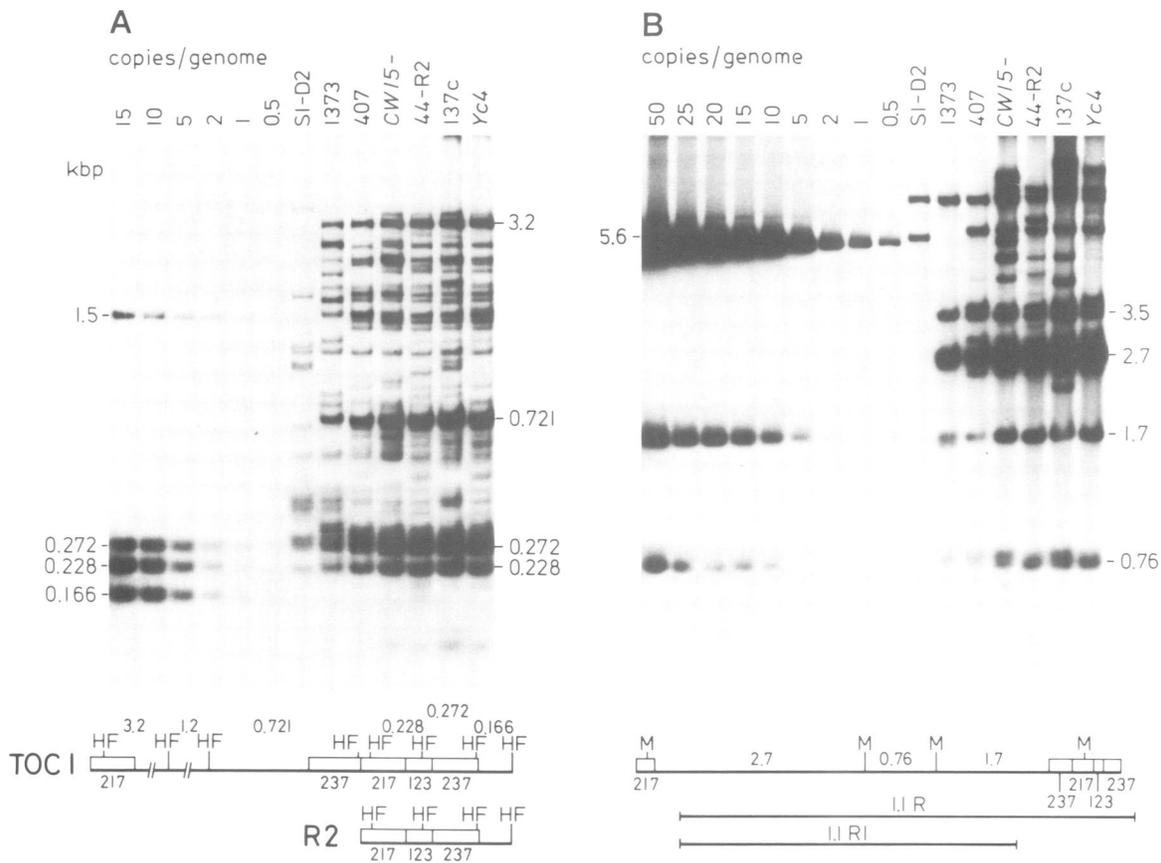
**Fig. 3.** Copy number variation of TOC1 between strains. pTOC1.1R1 was used as the probe in **A**, **B**, **C** and **E**. To obtain **D**, the filter used in **C** was reprobed with pR2. The restriction enzymes used are indicated above each blot. In **A** and **B**, digests were fractionated on a 0.8% agarose gel for 60 h at 1 V/cm. The blots probed in **C**, **D** and **E** were obtained from 1.1% agarose gels run for 16 h at 1.4 V/cm. The positions in kbp of marker restriction fragments are indicated on the left of panels **A**, **B** and **C**. cc407, cc410, cc1374 and cc1481 are wild-type strains of *C.reinhardtii*. 137c is our standard photosynthetic laboratory strain. cc1373 is an isolate of *C.smithii*. S1-D2 was obtained from P.Lefebvre. cw15 mt+ and mt- are opposite mating types of a cell-wall deficient strain. FUD44-R2 and R3 are revertants of FUD44.

that does not cut TOC1, e.g. *SalI* (Figure 3A), or DNA digested with enzymes that cut once, close to or beyond the ends of the probe, e.g. *BamHI* (Figure 3B) or *HindIII*, are an estimate of the copy number of TOC1 per genome.

Figure 3A and B shows that the copy number of TOC1 elements varies widely between different *C.reinhardtii* strains. This copy number variation is characteristic of mobile dispersed repetitive DNA sequences. The irregular spacing of bands bearing TOC1 elements is indicative of a dispersed organization. TOC1 copy number estimates of 2, 6, 11, 14–15, 10, 11, 22–24, 11–12, 35–40 and 23–24 were obtained for S1-D2, cc1373, cc407, c410, cc1374, cc1418, cw15 mt-, cw15 mt+, 137c and FUD44 respectively. For 137c and cw15 mt- the TOC1 copy number estimates were based on hybridization intensity of dots containing 137c and cw15 mt- DNA to pTOC1.1R1 since digests of these DNAs contain too many co-migrating bands. Strains cc407, cc410, cc1374 and cc1418 (from the *Chlamydomonas* stock centre at Duke University) were gathered

from a number of *C.reinhardtii* collections but it is unclear whether they represent independent isolates. 137c is our standard photosynthetic laboratory strain (Togasaki and Levine, 1970). S1-D2, which was recently isolated from the wild at Plymouth, Minnesota, mates freely with 137c and was kindly provided by P.Lefebvre (University of Minnesota). cc1373 is interfertile with *C.reinhardtii* but on the basis of body shape and slit chromatophore has been classified as a different species (Hoshaw and Ettl, 1966).

Examination of the *BamHI* digests of FUD44, FUD44-R2 and FUD44-R3 reveal that three bands of 4.8 (Figure 3E), 5.5 and 15 kbp (Figure 3B) in size present in FUD44 are absent in FUD44-R2 and R3. Only the absence of one of these bands (15 kbp) in digests of FUD44-R2 and FUD44-R3 is due to excision of TOC1 from the OEE1 locus; the 15-kbp band in the FUD44 lane hybridizes to pOE3' (results not shown, pOE3' is shown in Figure 1E). The new band that is slightly larger than 15 kbp in FUD44-R2 and the new band of 7.4 kbp in FUD44-R3 are not derived from



**Fig. 4.** (A) Internal organization of LTRs. (B) Internal organization of TOC1 elements in various strains. Southern blots bearing *HinfI* digests of total DNA (2.5  $\mu$ g/lane) and dilutions of pR2 in A, and *MluI* digests of total DNA (2.5  $\mu$ g/lane) and dilutions of pTOC1.1R1 in B, were hybridized with  $^{32}$ P-labelled pR2 and pTOC1.1R1, respectively. In A, the 1.5-kbp *HinfI* band of pR2 (lanes 1–6) contains 80 bp (of which 12 bp are OEE1 sequences) that are complementary to the probe, the remainder consists of pUC18 sequences. In B, the 5.6-kbp band in *MluI* digests of pTOC1.1R1 (lanes 1–9) contains 2.3 kbp of the 2.7-kbp *MluI* fragment, pUC18 (2.69 kbp) and the right junction fragment of TOC1.1R (0.64 kbp) located beyond the far right *MluI* site. To estimate copy numbers, the *C. reinhardtii* haploid genome size was taken as 100 000 kbp. HF = *HinfI*, M = *MluI*. *HinfI* fragments within the 1.2-kbp internal region of TOC1 that do not hybridize to pR2 are not shown.

the OEE1 locus since pTOC1.1R1 is an internal probe and does not hybridize to the LTR ([217-bp][123-bp][237-bp] sequences present in the OEE1 genes of FUD44-R2 and R3.

**Are most long terminal repetitions associated with intact TOC1 elements?**

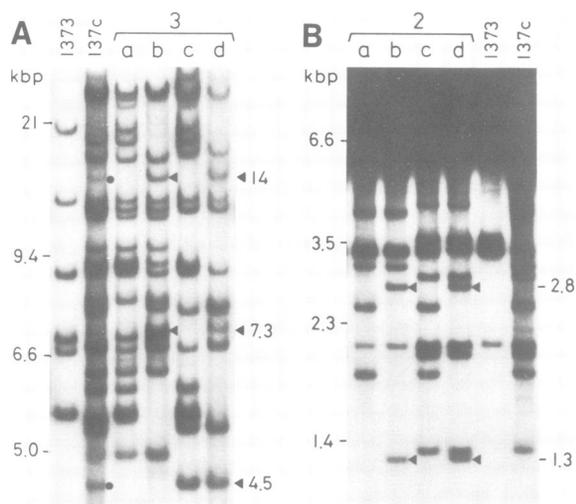
Of the two characterized TOC1 excision events from the OEE1 gene, one leaves an intact solo LTR ([217-bp][123-bp][237-bp] unit) in the OEE1 gene of FUD44-R2 while the other excision event leaves behind a deleted solo LTR in FUD44-R3 (Figure 2A). Since *SalI* does not cut within the intact TOC1 element, the number of solo LTRs in a given strain can be estimated by first probing a Southern blot bearing a *SalI* digest with a large internal TOC1 probe and then determining the number of additional bands that hybridize to a second LTR-specific probe. Strictly speaking this is an estimate of non-TOC1-associated LTRs rather than solo LTRs since the LTRs could flank the ends of a different family of repeated sequences.

Solo LTRs are present in all the strains examined (Figure 3D). They are most noticeable in strains bearing few TOC1 elements. However, even in digests of strains bearing many TOC1 elements, solo LTRs can be easily detected as *SalI* fragments of <5.7 kbp in size that hybridize to the LTR-

specific probe pR2 (Figure 3D) which are not detected with the internal TOC1 probe pTOC1-1R1, even after long exposure times (Figure 3C). The 193 bp of the OEE1 gene in the pR2 probe hybridizes to a 1.8-kbp OEE1 band in *SalI* digests from most strains. Within a given lane, solo LTR bands exhibit a great deal of variation in their hybridization intensity to pR2. S1-D2 and cc1373 contain 13 and 16 solo LTR units per genome, respectively. The strains cc407, cc410, cc1374, cc1418 and *cw15* mt+ each contain from 10 to 12 solo LTRs per genome. At least nine solo LTRs can be found in *SalI* digests of DNA from any one of *cw15* mt-, FUD44 and 137c, but the actual numbers of solo LTR bands which may be expected to be much higher than nine per genome are masked by the large number of TOC1 bands in *SalI* digests of DNA from these strains.

**Is the unusual LTR arrangement of the TOC1 element present at OEE1 shared by other TOC1 elements?**

Figure 4A and B shows the hybridization patterns of the LTR-specific probe, pR2, and the internal pTOC1.1R1 probe, obtained with fractionated *HinfI* and *MluI* digests of total DNAs, respectively. In Figure 4B, the relatively higher hybridization intensity of the 2.7-kbp fragment is partly explained by greater probe overlap. The 0.228- and



**Fig. 5.** Genetic analysis of TOC1 elements. Southern blots bearing *Bam*HI digests of DNA (2.5  $\mu$ g/lane) from the products of tetrads hybridized with pTOC1.1R1 in **A** (tetrad no. 3) and pR2 in **B** (tetrad no. 2). The products of individual meiotic events are marked a, b, c and d. In **A**, digests were fractionated by electrophoresis in a 0.8% agarose gel for 60 h at 1 V/cm. In **B**, digests were fractionated overnight on a 1.1% agarose gel at 1.5 V/cm. Arrows indicate strongly hybridizing bands in digests of progeny DNA that are either reduced in intensity (marked by dots) or absent in digests of the parental DNAs. The positions and sizes of marker restriction fragments are indicated on the left of each panel.

0.272-kbp *Hinf*I fragments present in all the total DNA digests in Figure 4A are most probably derived from both conserved solo LTRs and conserved TOC1 elements (conserved relative to TOC1.1 and R2). The general correspondence between the copy numbers of the 0.228- and 0.272-kbp LTR-derived *Hinf*I fragments (Figure 4A), and the copy number of the TOC1-specific 1.7-kbp *Mlu*I fragment (Figure 4B), leads us to believe that TOC1-associated LTRs make a major contribution towards the signal intensity of the 0.228- and 0.272-kbp *Hinf*I bands. Note also the similar hybridization intensities of the TOC1-specific 0.721-kbp *Hinf*I fragment and the LTR-derived 0.272- and 0.228-kbp *Hinf*I fragments in Figure 4A. A comparison of the hybridization intensities of pTOC1.1R1 to the 1.7-kbp *Mlu*I fragment in total DNA digests with the hybridization intensity of the 1.7-kbp *Mlu*I in dilutions of *Mlu*I-digested pTOC1.1R yields copy number estimates that are generally  $\sim$ 2-fold lower than copy number estimates based on number of bands (see above). None the less, with the exception of S1-D2, the multiplet nature of the 2.7- and 1.7-kbp *Mlu*I bands, and 0.721-, 0.272- and 0.228-kbp *Hinf*I bands in digests of DNA from a number of strains indicates that a large proportion of TOC1 elements within a variety of strains contain the odd split LTR arrangement of the cloned elements. This conclusion is supported by sequence analysis of TOC1.2 (see Figure 9 in Materials and methods). The [217-bp]/-[237-bp][217-bp][123-bp][237-bp] LTR organization of the TOC1.1 insertion in OEE1 is also present in TOC1.2. Comparison of the regions of TOC1.2 sequenced with the TOC1.1 sequence revealed no differences. This sequence identity is indicative of a high degree of gene conversion or a recent amplification of one particular element in FUD44.

Divergent LTRs and divergent TOC1 elements (Figure 4A and B respectively) give rise to the large number of size

classes of restriction fragment in total DNA digests. The non-TOC1.1 size classes of *Hinf*I fragments do not represent single copy junction fragments of conserved LTRs: the right junction fragment only contains 9 bp of the LTR; the single copy 1.5-kbp *Hinf*I fragment of pR2 represents a left junction fragment with 68 bp of LTR sequences and 12 bp of OEE1 complementary to the probe and is barely detectable in Figure 4A. It is also possible that divergent *Hinf*I and *Mlu*I bands are the result of methylation (McClelland and Nelson, 1985). However, we think this is unlikely since *C. reinhardtii* nuclear DNA is not extensively methylated. Analyses of *Hpa*II (does not cleave <sup>me</sup>CCGG or C<sup>me</sup>CCG) and *Msp*I (cleaves C<sup>me</sup>CCG but not <sup>me</sup>CCGG) digests indicate that the bulk of *C. reinhardtii* nuclear DNA (ethidium bromide stain), and all the internal (pTOC1.1R1 probe) and LTR sequences of TOC1 (pR2 probe) are relatively free of the methylated dinucleotide <sup>me</sup>CG (not shown).

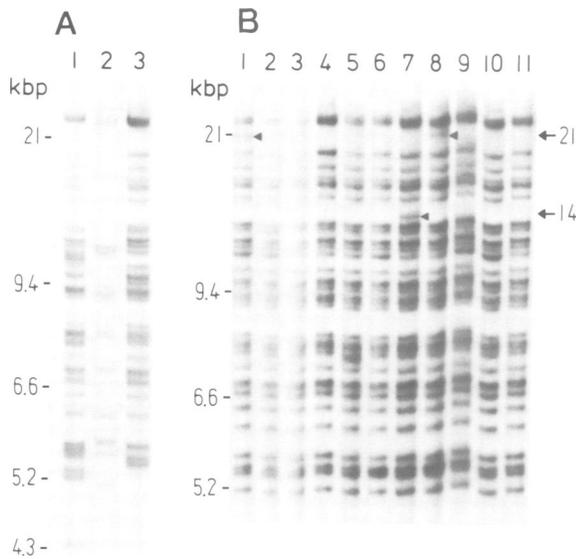
S1-D2 is conspicuous because it only contains TOC1-related sequences that lack the characteristic *Mlu*I sites of TOC1.1 (Figure 4B). Bands that are present in a number of different strains, e.g. the 3.5-kbp *Mlu*I fragment in Figure 4B, are presumably the result of events that had occurred in a common ancestor of the strains examined. Mechanisms that would disrupt the TOC1 structure include recombination events between the 237-bp repeats of the right [237-bp][217-bp][123-bp][237-bp] unit. Such a recombination event would remove the [217-bp][123-bp][237-bp] unit and the *Mlu*I site therein. Over a period of time these events may be expected to lead to a stoichiometric excess of the 2.7-kbp *Mlu*I fragment.

#### Genetic linkage of TOC1 elements

Southern blots bearing *Bam*HI digests of DNA from 25 meiotic (four complete and three incomplete tetrads) segregants of a cross between high and low copy TOC1 containing strains (137c  $\times$  cc1373) were hybridized with pTOC1.1R1 and pR2. The results obtained with these hybridization probes against *Bam*HI digests of complete tetrads are illustrated by tetrads nos. 3 and 2 in Figure 5A and B. The pattern of segregation of parental bands between the progeny is largely independent. Lack of linkage is consistent with the view that TOC1 elements are scattered over the 17 or so linkage groups of *C. reinhardtii*.

New bands, e.g. 7.3 kbp (Figure 5A) and 2.8 and 1.3 kbp (Figure 5B), that are not found in digests of either of the parental DNAs (137c or cc1373) can be detected in digests of the progeny DNA. They always segregate 2:2 in complete tetrads. Some bands that hybridize weakly to pTOC1.1R1 in *Bam*HI digests of 137c, e.g. 4.5 and 14 kbp (Figure 5A, marked by dots), increase in hybridization intensity in the progeny of some complete tetrads, e.g. tetrad no. 3 (Figure 5A, marked by arrows), and are not present in any of the progeny of other complete tetrads (tetrads nos. 1, 2 and 4, data not shown).

The simplest interpretation of the results is that these changes had occurred in the chromosomes of one or both of the parents prior to the DNA replication step that takes place before meiotic division of the young zygotes. This raises the possibility that one or both of the parents was composed of a heterogeneous population of cells. The *Bam*HI digestion pattern of TOC1 elements in mitotically grown cc1373 was stable over a period of 8 months. The 137c parent is more likely to have been composed of a mixed

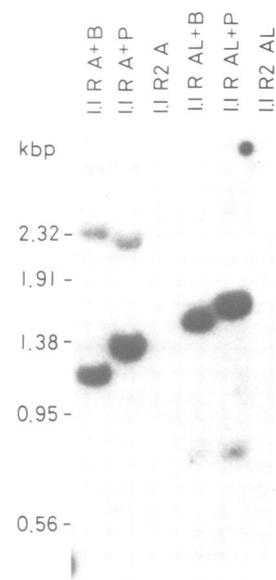


**Fig. 6.** Mitotic instability of TOC1 in strain 137c. Southern blots bearing *Bam*HI digests of DNA from cultures grown from single colonies hybridized with pTOC1.1R1. Colonies were separated by at least 14 months of growth from a single colony in **A** and by an average of 29 mitotic divisions in **B**. For **B**, a single colony was resuspended in 65 ml of TAP medium and grown in dim light (mixotrophic growth) at 25°C to a density of  $1.0 \times 10^7$  cells/ml before plating at low density on solid TAP medium. New bands appearing after mitotic growth are indicated by arrows. The positions and sizes of marker restriction fragments are indicated on the left of each panel.

population of cells for two reasons: first, weakly hybridizing *Bam*HI bands in 137c either increase in intensity or are lost in the progeny of the cross; and second, new TOC1 bands appear after normal mitotic growth of 137c (see below).

**TOC1 instability during normal mitotic growth of strain 137c**

A pTOC1.1R1 hybridization probe reveals large differences between *Bam*HI digests of DNA from colonies of 137c mt-propagated independently on plates (Figures 6A, lanes 1–3) for at least 14 months by two of the authors (M.K. and A.D.). Since under optimal growth conditions, *C.reinhardtii* divides every 6–8 h, 14 months of growth represents in the order of 1000 mitotic divisions. The extent of the TOC1 instability in 137c mt– was unexpected. Under normal growth conditions, new TOC1 bands of 14 and 21 kbp were found in digests of DNA from three out of 11 colonies tested that were grown up from cells plated after only 29 mitotic divisions (average value) of a single progenitor cell (see Figure 6B, note also that all colonies tested in Figure 6B give rise to a 5.2-kbp band that is not produced by colony 3 in Figure 6A which is a relatively recent ancestral colony). Colonies 1 and 8 of Figure 6B are probably derived from the same progenitor cell since they both produce a new TOC1 band of the same size. It seems unlikely that these changes in *Bam*HI digestion pattern are due to single base changes or gross chromosomal rearrangements since a hybridization probe for a short dispersed repetitive (> 500 copies/haploid genome) sequence detects no differences between *Bam*HI digests of these DNAs (not shown). A change in the chromosomal distribution of TOC1, mediated by gene con-



**Fig. 7.** Retrotransposon-like features of TOC1. Southern blot bearing digests of pTOC1.1R and pTOC1.1R2 hybridized with a mixture of four 5' <sup>32</sup>P-labelled 14-mer oligonucleotides complementary to bases encoding most of a pentapeptide conserved between reverse transcriptases. Hybridization conditions: 16 h at 25°C in 5 × SSC, 30% (v/v) HCONH<sub>2</sub>, 3% (w/v) dextran sulphate (mol. wt. 500 000), 8 × Denhardt's solution (Denhardt, 1966), 0.15% (w/v) Na<sub>2</sub>P<sub>2</sub>O<sub>7</sub>, 0.1% SDS, 100 µg/ml salmon sperm DNA, M13 sequencing primer (0.2 µg/ml). Wash conditions: 1 × SSC, 0.1% SDS (w/v) at 20–24°C. A = *Ava*II, AL = *Alu*I, B = *Bam*HI, P = *Pst*I. The positions and sizes of marker restriction fragments are indicated.

version or transposition, is the most plausible explanation for these changes. In *S.cerevisiae*, movement of Ty by gene conversion is 10-fold higher than Ty transposition (Roeder and Fink, 1982). Large differences in restriction fragment lengths will result from gene conversion events between solo LTRs and intact TOC1 elements.

**Is TOC1 a retrotransposon?**

If TOC1 is a retrotransposon it may be expected to encode reverse transcriptase. We therefore synthesized a mixture of four 14mer oligonucleotides complementary to the oligonucleotides that encode Tyr-Met-Asp-Asp-Leu and Tyr-Met-Asp-Asp-Ile using the biased codon usage of *C.reinhardtii* nuclear genes. These pentapeptides correspond to a region (Box III) that is conserved between reverse transcriptases (Toh *et al.*, 1983; Patarca and Haseltine, 1984). The tyrosine and first aspartic acid residue are essential for reverse transcriptase activity (Larder *et al.*, 1987). These oligonucleotides hybridize strongly to pTOC1.1R and do not hybridize at all to pTOC1.1R2 (see Figures 7 and 1E). The region of oligonucleotide complementarity maps to an internal region of TOC1 that lies within a 500-bp stretch right of the unique *Bam*HI site (see Figure 2A).

**Discussion**

Insertion of a repeated sequence into the second intron of the OEE1 gene blocks oxygen evolution in the *C.reinhardtii* strain FUD44. We have named this element Transposon of *Chlamydomonas* 1 or TOC1 based on its structure and lack of sequence similarity with its integration site in OEE1. Properties shared by TOC1 and viral-like retrotransposons

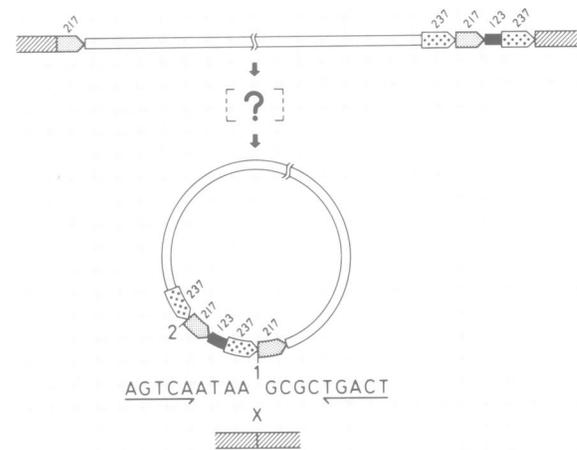
include: (i) long-terminal direct repeats, and (ii) an oligonucleotide stretch corresponding to a conserved pentapeptide stretch of reverse transcriptases. In addition, both copies of the 237- and 217-bp repeats are identical suggesting that, like retroviral LTRs, they are derived from the same template (Gilboa *et al.*, 1979; Boeke *et al.*, 1985).

### Retrotransposition of TOC1

The unusual split LTR arrangement of TOC1 (see Figure 2), in which a unique 123-bp sequence is located between the 217- and 237-bp repeats at its right end, means that if TOC1 moves by retrotransposition, the retrovirus replication model (Gilboa *et al.*, 1979; Varmus, 1982) that also provides a model for Ty transposition (Boeke *et al.*, 1985) needs to be substantially modified for TOC1. Although unusual LTR arrangements in retroviral proviruses have been reported (for example, see Sonigo *et al.*, 1986), these cases are rare and reflect aberrant recombination events. In contrast, the unusual LTR arrangement of TOC1 is shared by a large fraction of the total TOC1 elements in a given strain, which suggests that its split LTR structure is a normal feature of its transposition cycle.

We have detected an unusually large (11–12 kb) TOC1 transcript whose 5' to 3' orientation is left to right with respect to Figure 2A (A.Day and J.-D.Rochaix, unpublished results). The 5' upstream sequences necessary for transcription initiation in *C.reinhardtii* are poorly defined and cannot be used to locate the putative promoter region of TOC1. TGTA, a likely component of the *C.reinhardtii* polyadenylation signal (Silflow *et al.*, 1985; Goldschmidt-Clermont and Rahire, 1986), is found at base 206 of the 237-bp repeat. It is interesting to note that the solo LTR sequence in the OEE1 gene of FUD44-R2, which presumably contains transcription initiation and polyadenylation signals, reduces but does not prevent the accumulation of OEE1 mRNA (Mayfield *et al.*, 1987). The absence of a TGG beside the left or right 217-bp TOC1 repeats and lack of a suitable oligopurine stretch adjacent to either of the two 237-bp repeats (see Figure 2A and B), indicates that the priming mechanisms used for first and second strand synthesis by TOC1 are different from those used in retroviral replication. It is already known that retrotransposons need not adhere strictly to the retroviral replication model: *copla* extends first strand synthesis not from the 3' CCA but from an internal site of a tRNA (Kikuchi *et al.*, 1986).

In our model, we suggest that a retrotransposition mechanism, the details of which remain to be determined, is responsible for producing the substrate for integration, the hypothetical circular double-stranded DNA molecule shown in Figure 8. This circular molecule contains two putative integration sites of which one site (site 1) preferentially directs integration. Integration at site 2 would generate a 577-bp *MluI* fragment (the 217-bp repeat contains a *MluI* site) that should hybridize to pR2. This is clearly not observed (Figure 1D, lane marked S+M). We suggest that either the flanking unique region of TOC1 produces this asymmetry in recombination sites or that steps in TOC1 retrotransposition produce a linear integration substrate which is already open at site 1. It is still unclear whether the immediate precursors of the integrated proviral DNA are linear or circular DNA molecules (Brown *et al.*, 1987; Colicelli and Goff, 1988).



element	species	size (kbp)	terminal bases and internal size of LTR (kbp)	5'	3'	references
<i>copla</i>	<i>D.melanogaster</i>	5	TGTTGG-(0.264)	-acaACA	Levis <i>et al.</i> 1980.	
<i>B104</i>	<i>D.melanogaster</i>	8.7	TGTTca-(0.417)	-tttACA	Scherer <i>et al.</i> 1982.	
SNV	<i>G.gallus</i>	8.3	TGTCGG-(0.557)	-ACAACA	Shimotohno <i>et al.</i> 1980.	
IAP	<i>M.musculus</i>	7.2	TGTTGG-(0.326) <sup>1</sup>	-ataACA	Kuff <i>et al.</i> 1983.	
<i>Hpa II</i>	<i>P.poliocephalum</i>	8.6	TGTTGG-(0.265) <sup>1</sup>	-ctaACA	Pearston <i>et al.</i> 1985.	
Ty 1	<i>S.cerevisiae</i>	5.9	TGTTGG-(0.322)	-ttctCA	Clare and Farabaugh, 1985.	
Wis-2	<i>T.aestivum</i>	8	TGTTGG-(>0.50) <sup>1</sup>	-cctACA	Harberd <i>et al.</i> 1987.	
Bs 1	<i>Z.mays</i>	3.3	TGTTAg-(0.292)	-tctgac	Johns <i>et al.</i> 1985.	
Cin 1	<i>Z.mays</i>	0.7	TGTTGG-(0.679) <sup>2</sup>	-ccaACA	Shepherd <i>et al.</i> , 1984.	
THE 1	<i>H.sapiens</i>	2.3	TGgTat-(0.338) <sup>1</sup>	-aatACA	Paulson <i>et al.</i> 1985.	
MSV	<i>M.musculus</i>	5.2	TGaaaG-(0.576)	-cttCA	Dhar <i>et al.</i> 1980.	
17.6	<i>D.melanogaster</i>	7.4	agTGacat-(0.495) <sup>1</sup>	-tttgCAatt	Saigo <i>et al.</i> 1984.	
TOC1	<i>C.reinhardtii</i>	5.7	gcgcTGactt-(0.434) <sup>2</sup>	-gaGTCAataa	this paper.	

**Fig. 8.** A model for the retrotransposition cycle of TOC1 and a comparison of the terminal bases of two integrated retroviral proviruses (SNV and MSV) and various retrotransposon and retrotransposon-like elements. For further details see text. Copy number estimations from the reference shown above: *copla* (20–45)<sup>4</sup>, *B104* (strain 1, 80; strain 2, 95)<sup>5</sup>, IAP (~1000)<sup>5</sup>, *HpaII* (4000 ± 1000)<sup>4</sup>, Ty1 (~30)<sup>5</sup>, Wis-2 (>100)<sup>4</sup>, Bs-1 (1–5)<sup>4</sup>, Cin-1 (~1000)<sup>4</sup>, THE 1 (~10 000)<sup>5</sup>, TOC1 (2–40)<sup>5</sup>. (1) Imperfect long terminal repeats. (2) Cin 1 is a dispersed repeat of maize that resembles a solo LTR. (3) The size of the TOC1 LTR is taken as the sum of the 217- and 237-bp repeats. (4) Copy number estimates/genome. (5) Copy number estimates/haploid genome. Arrows indicate short inverted repeat sequences.

In the putative circular integration intermediate of retroviruses (2-LTR circle, Panganiban and Temin, 1984; Varmus, 1985) the 3' CA and 5' TG of adjacent tandem LTRs are separated by four bases and make up part of a short palindrome whose rotational symmetry lies at the LTR junction (Varmus, 1985). The four bases that separate the 5' TG and 3' CA are absent in the integrated proviral DNA (Hughes *et al.*, 1978). Two features of the proposed integration pathway for TOC1 are novel: first, the site of integration is not at the centre of a palindrome but at the centre of eight bases that separate the dinucleotides CA and TG which are part of a five-base inverted repeat (see Figure 8); and, second, the eight bases that separate the five-base inverted repeat are not lost upon integration. We suggest that these points reflect the properties of the TOC1 integrase and are responsible for the observation that although TOC1 does not begin with a 5' TG and end with a 3' CA, it does contain a 5-bp inverted repeat located 4 bp from the ends that follows the 5' TG 3' CA rule of most retro-

transposons (see Figure 8). Recent experiments with Moloney murine leukemia virus (Colicelli and Goff, 1988) support our TOC1 integration model. Colicelli and Goff show that disruption of the palindrome of the 2-LTR circle by insertion of bases between the 3' CA and 5' TG containing inverted repeats of adjacent tandem LTRs does not prevent retroviral integration. The identities of the bases separating the inverted repeat are relatively unimportant and, in contrast to our model of TOC1 integration, they are usually lost upon integration (Colicelli and Goff, 1988).

Figure 8 shows that the 5' terminal bases of the proviral DNA of chicken spleen necrosis virus (SNV) and a number of putative retrotransposons conform closely to the consensus sequence TGTTGG which is present in three established retrotransposons, Ty, copia and integrated intracisternal A-particle elements (IAP, see Baltimore, 1985). The termini of Moloney murine sarcoma proviral DNA (MSV) and the putative human retrotransposon THE 1 show that the 5' TGTTGG consensus is not universal amongst elements observing the 5' TG 3' CA rule. The putative retrotransposons 17.6 (Figure 8) and 297 (not shown, Ikenaga and Saigo, 1982) from *D.melanogaster* join TOC1 in breaking the 5' TG 3' CA rule. Like TOC1, 17.6 and 297 contain a TG close to their 5' end and CA close to their 3' end, but unlike TOC1 these dinucleotides are not part of an inverted repeat. Perhaps DIRS-1 from *Dictyostelium discoideum* is the most bizarre of the retrotransposon candidates. In DIRS-1 the LTRs are inverted rather than direct with non-complementary extensions at their ends (Rosen *et al.*, 1983; Capello *et al.*, 1985). Some of the elements listed in Figure 8 are bounded by perfect LTRs which suggests that they originate from relatively recent retrotransposition events.

TOC1.1 does not produce a target site duplication at its integration site in OEE1. Transposable elements that do not produce target site duplications include Tn554 of *E.coli* (Murphy and Lofdahl, 1984), DIRS-1 (Zuker *et al.*, 1984) and perhaps Tc1 of *Caenorhabditis elegans* (Herman and Shaw, 1987). TOC1.2 is flanked by a 4-bp repeat. Whether these 4 bp represent the target site duplication or whether they were present prior to TOC1.2 integration is not known. More regions flanking TOC1 integration sites will need to be characterized before deciding whether TOC1, like Tam 3 from *Antirrhinum majus* (Coen *et al.*, 1986) and IS4 of *E.coli* (Haberman *et al.*, 1979), produces a variable length of target site duplication.

### Excision of TOC1

Imprecise excision of transposable elements has been noted in bacteria (Nag and Berg, 1987), the P element system of *D.melanogaster* (Daniels *et al.*, 1985) and is well documented in higher plants. In higher plants, excision of most DNA-mediated transposable elements leaves a 'foot-print' which is derived from the target site duplication (Peacock *et al.*, 1984; Schwarz-Sommer *et al.*, 1985). Excision of TOC1 from the OEE1 locus resembles Ty excision events. Ty excision generally leaves behind a solo LTR (Farabaugh and Fink, 1980). In both the *S.cerevisiae*/Ty and *C.reinhardtii*/TOC1 systems, solo LTRs form a separate family of repeated sequences. This is in contrast to the situation found in the *D.melanogaster*/copia system where solo LTRs have not been found (Levis *et al.*, 1980). FUD44-R2, a representative of the major class of

revertants (38/39 revertants), probably resulted from a recombination event between the directly repeated terminal 217-bp sequences of TOC1.1 since the [217-bp][123-bp] [237-bp] LTR unit in the OEE1 gene of FUD44-R2 is identical in sequence to the corresponding [217-bp][123-bp] [237-bp] LTR unit of TOC1.1. The deleted solo LTR present in the OEE1 gene of FUD44-R3 contains an adenosine residue in place of bases 122–508 of the intact solo LTR. Two cross-over events are required to produce this structure and we prefer a gene conversion event with a deleted solo LTR to account for the excision of TOC1.1 from the OEE1 locus of FUD44-R3.

Could TOC1 excision products give rise to transposition, gene conversion or recombination intermediates that are able to reintegrate into the genome? It will only be possible to attempt to answer this question with a marked TOC1 element. However, it is interesting to note that reversion of FUD44 is associated with the loss of the TOC1 element at the OEE1 locus and the appearance of TOC1 elements elsewhere in the genome.

The isolation of TOC1 provides the basis for developing a transposon tagging system (Bingham *et al.*, 1981) in *C.reinhardtii*. The ability to reduce the TOC1 copy number of a given strain by crossing it with a low copy TOC1 containing strain should facilitate the isolation of tagged genes.

## Materials and methods

### Strains and media

*Escherichia coli* strains with relevant genotype. C600, NM539 (P2cox3), BHB 2688 (N205 recA- [clts.Eam4/lambda]), BHB 2690 (N205 recA-[clts, Dam15/lambda]), ED8767 (recA56), DH5 (recA1).

*Chlamydomonas reinhardtii* strains. The wild-type cc407, cc410, cc1374, cc1418 and *C.smithii* cc1373 strains were obtained from E.H.Harris at the *Chlamydomonas* Culture Collection, Duke University, Durham, NC. S1-D2 was obtained from P.Lefebvre (University of Minnesota). The mutant strain FUD44, which is deficient in the OEE1 protein, is photosynthetically inactive and therefore unable to fix CO<sub>2</sub>. FUD44 will only grow on a medium containing a reduced carbon source such as acetate. Revertants of FUD44 that are able to fix CO<sub>2</sub> and as a result grow on minimal medium have been isolated at a frequency of ~10<sup>-5</sup>/cell. The isolation of FUD44 has been described (Mayfield *et al.*, 1987). yc4 and the two mating types of the cell wall-less mutant cw15 (Davies and Plaskitt, 1971) were obtained from P.Bennoun (Institut de Biologie Physico-chimique, Paris). Our 137c mt- and mt+ strains were isolated as single meiotic segregants after repeated crosses between our mt- and mt- 137c strains in the spring of 1985.

Bacteria were grown on LB and NZY media made up according to Maniatis *et al.* (1982). Ampicillin was added to a final concentration of 25 µg/ml. *Chlamydomonas* cells were grown in an acetate medium (TAP) or in the same medium that lacks acetate (min, Gorman and Levine, 1965).

### UV mutagenesis

FUD44 cells plated out on minimal and complete TAP medium were exposed for 3 min to a Westinghouse Sterilamp (782-L-30) emitting primarily at 254 nm from a distance of 0.88 m. Ten per cent of the plated cells survived irradiation. The frequency of revertants obtained by UV treatment was similar to the spontaneous reversion frequency.

### Genetic crosses

137c mt- and cc1373 mt+ crosses were performed as described in Levine and Ebersold (1960). The fertility of some of the resulting progeny was tested. The zygotes from 1D × cc1373, 1B × 137c and 3C × 137c crosses gave rise to viable meiotic products.

### Cloning experiments

*Phage cloning.* Phage packaging extracts (5 × 10<sup>7</sup> plaques/µg of exogenous λ papa DNA) were prepared according to Hohn (1979). The phage cloning procedures of Kaiser and Murray (1985) were followed. To maximize

the percentage of recombinants, *Bam*HI arms of  $\lambda$  EMBL4 were purified on NaCl gradients and ligated to 12- to 20-kb size fractions (NaCl gradients) of FUD44 and FUD44-R2 DNA digested partially with *Sau*3a. Ligation conditions were chosen so that [i] > [j] for a 12-kbp restriction fragment (Maniatis et al., 1982). Ligation mixes were packaged and then plated on *E. coli* NM539. This procedure contains both physical and genetic selection steps against the central stuffer fragment of  $\lambda$  EMBL4. Libraries comprising > 50 000 plaques were obtained for FUD44 and FUD44-R2 DNAs. These were screened as described in Maniatis et al. (1982). Using an OEE1 cDNA probe six phages were isolated from the FUD44 library. No positively hybridizing phages (OEE1 cDNA) were found in the FUD44-R2 library even though this library contained multiple copies of a different single copy gene.

#### Plasmid cloning

Total *C. reinhardtii* DNA was digested with two restriction enzymes, electrophoresed through 20 × 20 × 0.3 cm vertical 0.8% agarose gels and fractions (0.4-cm slices) corresponding to the size of the desired fragment eluted by the freeze-squeeze method of Thuring et al. (1975). Southern blots of the fractions were prepared to locate the fraction containing the desired fragment. This fraction was ligated to a fragment preparation of doubly digested pAT153 (Twigg and Sheratt, 1980) under conditions where [j] > [i] for the desired size of linear recombinant fragment (Maniatis et al., 1982) and pAT153 ends were in excess. The ligation mix was used to transform *E. coli*. Competent ED8767 (1–5 × 10<sup>6</sup> transformants/ $\mu$ g pUC18) and competent DH5 cells (5–10 × 10<sup>6</sup> transformants/ $\mu$ g pUC18) were prepared and used according to the standard transformation protocol of Hanahan (1985). Colony screening was by the method of Hanahan and Meselson (1983). The *Hind*III–*Sal*I OEE1 fragments of FUD44-R2 (2.0 kbp) and FUD44-R3 (1.6 kbp), and the *Sal*I–*Bam*HI OEE1 fragment from FUD44 (1.2 kbp) were present in ~1 in every 500 transformants screened. The 7.5-kbp OEE1 *Hind*III–*Sal*I fragment of FUD44 was present in ~1 of every 2500 transformants screened. The inserts of the recombinant pAT153 plasmids bearing *Hind*III–*Sal*I inserts were excised, digested with *Pst*I and subcloned into pUC18 (Yanisch-Perron et al., 1985). TOC1.2 and 1.3 were cloned from a *Sal*I–*Hind*III partial digest of FUD44 DNA; the inserts in TOC1.2 and 1.3 did not contain internal *Hind*III sites. The right junction region of TOC1.2 and 1.3 were subcloned as *Sau*3A fragments in pUC18. Super-coiled plasmids purified from bacterial cultures (chloramphenicol enriched for pAT153) by the method of Birnboim and Doly (1979) were further purified on preparative CsCl/ethidium bromide density gradients.

#### DNA manipulations

Total *C. reinhardtii* DNA was extracted essentially according to Cryer et al. (1975); the exact protocol is given in Rochaix et al. (1988). Nuclear DNA was purified from total DNA on CsCl/Hoechst 33258 gradients (Manuelidis, 1977). All glassware was acid washed before use. DNA was digested in a medium salt buffer (Mirkovitch et al., 1984) using restriction enzymes

obtained from a variety of suppliers. Manufacturers' buffer recommendations were followed for low salt requiring enzymes. DNA electrophoresis, blotting and sizing methods have been described (Southern, 1979a,b). Restriction fragments obtained by digesting pAT153 with *Hin*FI and  $\lambda$  papa DNA with *Hind*III, and *Hind*III plus *Eco*RI were <sup>32</sup>P-labelled with Klenow enzyme and used as radioactive markers. Small restriction fragments were blotted in 0.2 M NaOH to Gene-screen plus nylon membranes (New England Nuclear; Khandjian, 1987). Copy number estimates of TOC1 elements (probe pTOC1.1R1) were based on a linear curve obtained by plotting number of bands in *Bam*HI, *Hind*III and *Sal*I digests of various DNAs against relative total hybridization intensity. Hybridization intensity was determined by Cerenkov counting of filter-bound dots containing equal amounts of total DNA that had been probed with pTOC1.1R1.

Inserts purified away from plasmid sequences on agarose gels were used as hybridization probes. Large and small (< 1 kbp) probes were [ $\alpha$ -<sup>32</sup>P]-dATP labelled by DNA polymerase I (nick-translation) and T4 DNA polymerase (replacement synthesis) respectively (Maniatis et al., 1982). [ $\alpha$ -<sup>32</sup>P]dATP (3000 Ci/mmol) and [ $\alpha$ -<sup>32</sup>P]UTP (400 Ci/mmol) were from Amersham International, UK. Hybridization of DNA probes was in 4 × SSC (1 × SSC = 0.15 M sodium chloride, 0.015 M sodium citrate), 50% (v/v) HCONH<sub>2</sub>, 8 × Denhardt's solution (Denhardt, 1966), 0.1% (w/v) SDS, 0.1% (w/v) sodium pyrophosphate and 100  $\mu$ g/ml salmon sperm DNA at 42°C. After hybridization Southern blots were washed in 0.1 × SSC, 0.1% (w/v) SDS at 65°C. Kodak X-OMAT films were exposed to hybridized blots at room temperature (Figures 1B, 1C, 3A, 3B, 3C, 3E, 4A, 4B, 5A and 6A) or at –60 to –70°C with intensifying screens (Ilford, fast tungstate).

#### Sequencing

5' and 3' <sup>32</sup>P-labelled restriction fragments were prepared and sequenced according to Maxam and Gilbert (1980). *Hae*II ends were labelled with T4 DNA polymerase (Maniatis et al., 1982). The sequencing strategy is shown in Figure 9.

#### Oligonucleotide synthesis

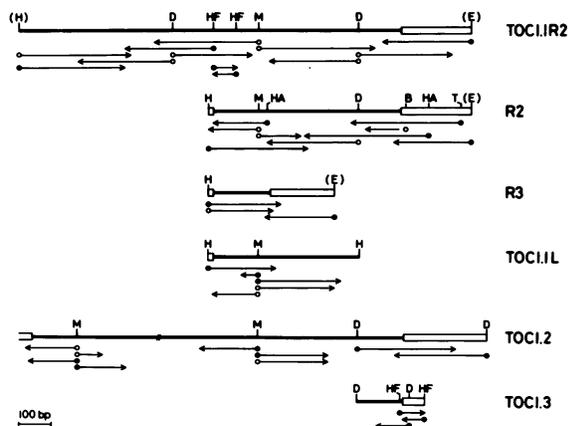
A mixture of oligonucleotides 5' A<sub>6</sub>GTCGTCCTCA<sub>6</sub>GTA were synthesized on an Applied Biosystems 381A DNA synthesizer.

## Acknowledgements

We thank C.Meric and R.Stick for general help and support, C.Mueller for sequencing tips, E.W.Khandjian for detailed advice on blotting methods, F.Veuey for help with computing, M.Goldschmidt-Clermont for the name TOC, E.H.Blackburn and P.Lefebvre for their gifts of *C. reinhardtii* strains, R.Debuchy, L.-G.Franzén and M.Goldschmidt-Clermont for useful comments on the manuscript and finally O.Jenni and F.Ebener for preparing the figures. This work was supported by grants 3.587.084 and 3.328.086 from the Swiss National Foundation to J.-D.R. and an EMBO long-term fellowship ALTF-1984 to A.D.

## References

- Baltimore,D. (1985) *Cell*, **40**, 481–482.  
 Boeke,J.D., Garfinkel,D.J., Styles,C.A. and Fink,J.R. (1985) *Cell*, **40**, 491–500.  
 Bingham,P.M., Levis,R. and Rubin,G.M. (1981) *Cell*, **25**, 693–704.  
 Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513–1523.  
 Brown,P.O., Bowerman,B., Varmus,H.E. and Bishop,J.M. (1987) *Cell*, **49**, 347–356.  
 Cappello,J., Handelsman,K. and Lodish,H.F. (1985) *Cell*, **43**, 105–115.  
 Choquet,Y., Goldschmidt-Clermont,M., Girard-Bascou,J., Kück,U., Bannoun,P. and Rochaix,J.-D. (1988) *Cell*, **52**, 903–913.  
 Clare,J. and Farabaugh,P. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 2829–2833.  
 Coen,E.S., Carpenter,R. and Martin,C. (1986) *Cell*, **47**, 285–296.  
 Colicelli,J. and Goff,S.P. (1988) *J. Mol. Biol.*, **199**, 47–59.  
 Cryer,D.R., Eccleshall,R. and Marmur,J. (1975) In Prescott,D.M. (ed.), *Methods in Cell Biology—Yeast Cells*. Academic Press, New York, Vol. 12, pp. 39–44.  
 Daniels,S.B., McCarron,M., Love,C. and Chovnick,A. (1985) *Genetics*, **109**, 95–117.  
 Davies,D.R. and Plaskitt,A. (1971) *Genet. Res.*, **17**, 33–43.  
 Denhardt,D.T. (1966) *Biochem. Biophys. Res. Commun.*, **23**, 641–646.



**Fig. 9.** Sequencing strategy. Regions flanking TOC1 or LTR sequences are drawn as boxes. B = *Bst*NI, D = *Dde*I, E = *Eco*RI, H = *Hind*III, HA = *Hae*III, HF = *Hin*FI, M = *Mlu*I, T = *Taq*I. 5' and 3' labelled ends are indicated by open and closed circles, respectively. Restriction sites within brackets lie in the polylinker region of pUC18. Only the 450-bp *Hind*III fragment that lies within the 1.2-kbp *Sal*I–*Bam*HI fragment of TOC1.1L was sequenced.

- Dhar,R., McClements,W.L., Enquist,L.W. and Van de Woude,G.F. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 3937–3941.
- Farabaugh,P.J. and Fink,G.R. (1980) *Nature*, **286**, 352–356.
- Finnegan,D.J. (1983) *Nature*, **302**, 105–106.
- Garfinkel,D.J., Boeke,J.D. and Fink,G.R. (1985) *Cell*, **42**, 507–517.
- Gilboa,E., Mitra,S.W., Goff,S. and Baltimore,D. (1979) *Cell*, **18**, 93–100.
- Goldschmidt-Clermont,M. and Rahire,M. (1986) *J. Mol. Biol.*, **191**, 421–432.
- Gorman,D.S. and Levine,R.P. (1965) *Proc. Natl. Acad. Sci. USA*, **54**, 1665–1669.
- Habermann,P., Klaer,R., Kühn,S. and Starlinger,P. (1979) *Mol. Gen. Genet.*, **175**, 369–373.
- Hanahan,D. and Meselson,M. (1983) In Wu,R., Grossman,L. and Moldave,K. (eds), *Recombinant DNA Part B—Methods in Enzymology*. Academic Press, New York, Vol. 100, pp. 333–341.
- Hanahan,H. (1985) In Glover,D.M. (ed.), *DNA Cloning—A Practical Approach*. IRL Press, Oxford, Vol. 1, pp. 109–135.
- Harberd,N.P., Flavell,R.B. and Thompson,R.D. (1987) *Mol. Gen. Genet.*, **209**, 326–332.
- Herman,R.K. and Shaw,J.E. (1987) *Trends Genet.*, **3**, 222–225.
- Hohn,B. (1979) In Wu,R. (ed.), *Recombinant DNA—Methods in Enzymology*. Academic Press, New York, Vol. 68, pp. 299–308.
- Hoshaw,R.W. and Ettl,H. (1966) *J. Phycol.*, **2**, 93–96.
- Hughes,S.H., Shank,P.R., Spector,D.H., Kung,H.-J., Bishop,J.M., Varmus,H.E., Vogt,P.K. and Breitman,M.L. (1978) *Cell*, **15**, 1397–1410.
- Ikenaga,H. and Saigo,K. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4143–4147.
- Johns,M.A., Mottinger,J. and Freeling,M. (1985) *EMBO J.*, **4**, 1093–1102.
- Kaiser,K. and Murray,N.E. (1985) In Glover,D.M. (ed.), *DNA Cloning—A Practical Approach*. IRL Press, Oxford, Vol. 1, pp. 1–47.
- Khandjian,E.W. (1987) *Biotechnology*, **5**, 165–167.
- Kikuchi,Y., Ando,Y. and Shiba,T. (1986) *Nature*, **323**, 824–826.
- Kuchka,M.R., Mayfield,S.P. and Rochaix,J.-D. (1988) *EMBO J.*, **7**, 319–324.
- Kuff,E.L., Fenestra,A., Lueders,K., Smith,L., Hawley,R., Hozumi,N. and Shulman,M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1992–1996.
- Larder,B.A., Purifoy,D.J.M., Powell,K.L. and Darby,G. (1987) *Nature*, **327**, 716–717.
- Levine,R.P. and Ebersold,W.T. (1960) *Annu. Rev. Microbiol.*, **14**, 197–216.
- Levis,R., Dunsmuir,P. and Rubin,G.M. (1980) *Cell*, **21**, 581–588.
- Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning—A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Manuelidis,L. (1977) *Anal. Biochem.*, **78**, 561–568.
- Maxam,A.M. and Gilbert,W. (1980) In Grossman,L. and Moldave,K. (eds), *Nucleic Acids Part I—Methods in Enzymology*. Academic Press, New York, Vol. 65, pp. 499–559.
- Mayfield,S.P., Bennoun,P. and Rochaix,J.-D. (1987) *EMBO J.*, **6**, 313–318.
- McClelland,M. and Nelson,M. (1985) *Nucleic Acids Res.*, **13** (Suppl.), r201–r207.
- Mirkovitch,J., Mirault,M.-E. and Laemmli,U.K. (1984) *Cell*, **39**, 223–232.
- Murphy,E. and Lofdahl,S. (1984) *Nature*, **307**, 292–294.
- Nag,D.K. and Berg,D.E. (1987) *Mol. Gen. Genet.*, **207**, 395–401.
- Panganiban,A.T. and Temin,H.M. (1984) *Cell*, **36**, 673–679.
- Patarca,R. and Haseltine,W.A. (1984) *Nature*, **309**, 728.
- Paulson,E., Deka,N., Schmid,C.W., Misra,R., Schindler,C.W., Rush,M.G., Kadyk,L. and Leinwand,L. (1985) *Nature*, **316**, 359–361.
- Peacock,W.J., Dennis,E.S., Gerlach,W.L., Sachs,M.M. and Schwartz,D. (1984) *Cold Spring Harbor Symp. Quant. Biol.*, **49**, 347–354.
- Pearston,D.H., Gordon,M. and Hardman,N. (1985) *EMBO J.*, **4**, 3557–3562.
- Rochaix,J.-D., Mayfield,S.P.M., Erickson,J.M. and Goldschmidt-Clermont,M. (1988) In Shaw,C.H. (ed.), *Plant Molecular Biology—A Practical Approach*. IRL Press, Oxford, in press.
- Roeder,G.S. and Fink,G.R. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 5621–5625.
- Rosen,E., Sivertsen,A. and Firtel,R.A. (1983) *Cell*, **35**, 243–251.
- Saigo,K., Kugimiya,W., Matsuo,Y., Inouye,S., Yoshioka,K. and Yuki,S. (1984) *Nature*, **312**, 659–661.
- Scherer,G., Tschudi,C., Perera,J. and Delius,H. (1982) *J. Mol. Biol.*, **157**, 435–451.
- Schwarz-Sommer,Z., Gierl,A., Cuypers,H., Peterson,P.A. and Saedler,H. (1985) *EMBO J.*, **4**, 591–597.
- Shepherd,N.S., Schwarz-Sommer,Z., Blumberg vel Spalve,J.B., Gupta,M., Wienand,U. and Saedler,H. (1984) *Nature*, **307**, 185–187.
- Shiba,T. and Saigo,K. (1983) *Nature*, **302**, 119–124.
- Shimotohno,K., Mizutani,S. and Temin,H.M. (1980) *Nature*, **285**, 550–554.
- Silflow,C.D., Chisholm,R.L., Conner,T.W. and Ranum,L.P.W. (1985) *Mol. Cell. Biol.*, **5**, 2389–2398.
- Sonigo,P., Barker,C., Hunter,E. and Wain-Hobson,S. (1986) *Cell*, **45**, 375–385.
- Southern,E.M. (1979a) In Wu,R. (ed.), *Recombinant DNA—Methods in Enzymology*. Academic Press, New York, Vol. 68, pp. 152–175.
- Southern,E.M. (1979b) *Anal. Biochem.*, **100**, 319–323.
- Temin,H.M. (1980) *Cell*, **21**, 599–600.
- Thuring,R.W., Sanders,J.P.M. and Borst,P. (1975) *Anal. Biochem.*, **66**, 213–220.
- Togasaki,R.K. and Levine,R.P. (1970) *J. Cell Biol.*, **44**, 531–539.
- Toh,H., Hayashida,H. and Miyata,T. (1983) *Nature*, **305**, 827–829.
- Twigg,A.J. and Sherratt,D. (1980) *Nature*, **283**, 216–217.
- Varmus,H.E. (1982) *Science*, **216**, 812–820.
- Varmus,H.E. (1985) *Nature*, **314**, 583–584.
- Weiner,A.M., Deininger,P.L. and Efstratiadis,A. (1986) *Annu. Rev. Biochem.*, **55**, 631–661.
- Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.
- Zuker,C., Capello,J., Lodish,H.F., George,P. and Chung,S. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2660–2664.

Received on April 7, 1987; revised on April 26, 1988