Zepp, a LINE-like retrotransposon accumulated in the *Chlorella* telomeric region

Takanobu Higashiyama,
Yoshiteru Noutoshi, Makoto Fujie and
Takashi Yamada¹

Faculty of Engineering, Hiroshima University, 1-4-1 Kagamiyama,
Higashi-Hiroshima 739, Japan

¹Corresponding author

T.Higashiyama and Y.Noutoshi contributed equally to this paper

Six copies of insertion elements accumulate in the subtelomeric region immediately proximal to the telomeric repeats on *Chlorella* chromosome I. The elements, designated Zepps, bear the characteristic features of non-viral (LINE-like) retrotransposons, including a poly(A) tail, 5'-truncations, a retroviral reverse transcriptase-like ORF and flanking target duplications. Detailed sequence analysis of the *Chlorella* subtelomeric region revealed a novel mechanism of Zepp transposition; successive insertions of each Zepp element into another Zepp as a target, leaving a tandem array of their 3'-regions with poly(A) tracts facing toward the centromere. Only the most distal Zepp copy was inverted to connect its poly(A) tail with the telomeric repeats. A similar Zepp cluster but without the telomeric repeats was also found at the terminus of another *Chlorella* chromosome. These structures contrast with that proposed for the addition of HeT-A and TART elements to *Drosophila* telomeres. Expression of Zepp elements is induced by heat shock treatment. Possible roles of the subtelomeric retrotransposons in formation and maintenance of telomeres are discussed.

Keywords: algal chromosome/LINE-like retrotransposon/nested integration/telomeric repeats

Introduction

Telomeres are the protein–DNA structures at the ends of eukaryotic chromosomes. They serve two major functions: to distinguish intact from broken chromosomes and to protect chromosomes from degradation and incomplete replication. In most organisms, telomeric DNA consists of a tandem array of simple sequence elements (Zakian, 1989, 1995; Blackburn, 1991). For example, the telomeric repeat 5'-TTTAGGG (as read toward the chromosome terminus) is found to be common to several higher plants, including *Arabidopsis thaliana* (Richards and Ausubel, 1988), maize (Broun *et al.*, 1992) and tomato (Ganel *et al.*, 1991), and is also highly conserved in the unicellular green alga *Chlorella vulgaris* (Higashiyama *et al.*, 1995). In general, the telomeric repeats are synthesized by a specialized reverse transcriptase called telomerase (Blackburn, 1992). However, there are alternative mechanisms

for extending telomeric sequences. In *Saccharomyces cerevisiae*, some recombination mechanisms can restore deficiencies in telomere elongation (Louis and Harber, 1990; Blackburn, 1992). The *Drosophila* telomere is composed of one or more LINE (long interspersed DNA element)-like retrotransposable elements designated HeT-A and TART (Young *et al.*, 1983; Biessmann *et al.*, 1992; Levis *et al.*, 1993). Instead of elongation by telomerase, incomplete DNA replication at the chromosome termini is counter-balanced in *Drosophila* by frequent transposition of these elements (Biessmann *et al.*, 1992; Levis *et al.*, 1993; Sheen and Levis, 1994; Mason and Biessmann, 1995).

In some organisms, the subtelomeric regions immediately proximal to the telomeric repeats consist of moderately repetitive sequences, which bear a superficial similarity to the *Drosophila* telomeric retrotransposons. Organization and possible functions of subtelomeric regions are largely unknown, but they may play important roles in maintaining the three-dimensional organization in the nucleus (Biessmann and Mason, 1992; Gilson *et al.*, 1993).

We have examined the molecular organization of chromosomes of the unicellular green alga *Chlorella vulgaris*. The genome of this organism consists of 16 small chromosomes that can be easily separated by pulsed-field gel electrophoresis (Higashiyama and Yamada, 1991). In our previous report we described detection of a telomeric repeat-associated sequence element that has a poly(A) tract immediately adjacent to the telomeric repeat (5'-TTTAGGG) on the smallest chromosome, chromosome I (Higashiyama *et al.*, 1995). We designated this poly(A)-bearing element Zepp.

In this study we have examined the detailed structure of the *Chlorella* chromosome I subtelomeric region, which contains several copies of the Zepp element. It was found that Zepp is a novel family of non-viral (LINE-like) retrotransposons (Rogers, 1985; Finnegan, 1989; Hutchinson *et al.*, 1989) and that Zepp integrates preferentially into itself, another Zepp sequence, as the target site.

Results

Detection of Zepp-related sequences in the *Chlorella* genome

We previously reported that the telomeres of *C.vulgaris* chromosome I are very short (~500 bp), consisting of 70 repetitions of a 5'-TTTAGGG sequence motif. Immediately proximal to the telomeric repeats on one arm, a poly(A)-bearing sequence element (Zepp) is repeated several times (Higashiyama *et al.*, 1995). Southern blot hybridization using a structurally conserved 3'-portion of the Zepp element (probe A, a 800 bp *SphI*–*HindIII* fragment close to the poly(A) tail; Figure 1B) as probe

A **B**

Fig. 1. Detection of the Zepp-related sequences in the *C. vulgaris* genome. (A) Before Southern blot analysis, total genomic DNA and isolated chromosome I DNA were digested with *Bal31* exonuclease for various periods and then with *HincII*. Size-fractionated bands were hybridized with a 800 bp *SphI*–*HindIII* fragment of the H1.3 clone (B) as Zepp probe (probe A) or pCHt-1 as probe for the telomeric repeats (Higashiyama *et al.*, 1995). Two *HincII* bands indicated by arrowheads were gradually degraded by extended *Bal31* digestion, indicating that they were derived from chromosomal termini. The same procedure with isolated chromosome I DNA produced a degraded band that corresponded to the larger of the two above-mentioned bands, so that the smaller band must have been from another chromosome. Note that the smaller band did not hybridize with the telomeric repeats as probe. (B) Partial restriction map of the left terminus of *Chlorella* chromosome I. λ clones I-3 and I-8 spanning this region were screened from a λ DASHII/chromosome I library. The clone H1.3 contains the most distal sequences mainly consisting of the telomeric repeats 5'-TTTAGGG (Higashiyama *et al.*, 1995). pYM25 is a clone containing the gap region between H1.3 and I-8, which was amplified by PCR using a pair of primers (small arrows). Probe A for Southern blot analysis is indicated by a shaded box. The orientation of this map is reversed from the previous one (Higashiyama *et al.*, 1995) for further comparison with other systems (Levis *et al.*, 1993; Mason and Biessmann, 1995). Symbols for restriction enzymes: H3, *HindIII*; N, *NotI*.

revealed many copies on restriction fragments of total *Chlorella* chromosomes (data not shown). Our previous Southern blot analysis of *Chlorella* chromosomes separated by pulsed-field gel electrophoresis indicated that Zepp elements were almost evenly distributed over all 16 chromosomes (Higashiyama *et al.*, 1995).

To see if Zepp elements are also located close to the termini of other chromosomes, *Chlorella* chromosomal DNA was treated with *Bal31* exonuclease for various

periods before *HincII* digestion and subjected to Southern blot analysis with the same probe (probe A). The results shown in Figure 1A reveal that at least two bands (8 and 2.3 kbp) of total chromosomal DNA fragments were gradually digested with *Bal31*, implying that these two bands were derived from chromosomal termini. Since the 8 kbp band came from the one terminus of chromosome I (Figure 1A), the 2.3 kbp fragment should originate from a similar terminal structure of another chromosome.

Fig. 2. Composite structures of Zepp elements on the left arm of *Chlorella* chromosome I. The boxes represent individual Zepp copies with a poly(A) tail at the 3'-end (bold vertical lines). Arrow indicates the length and orientation (5'→3') of each element. Nucleotide sequences of Zepp elements are referenced to a larger clone ZA-1 (Figure 4): Z-2B was found to be contiguous with Z-2A; Z-2B and Z-2A combined to form element Z-2. In the same way, Z-3B and Z-3A were joined in the Z-3 element and Z-4B and Z-4A made up element Z-4. Thus, there were six copies of Zepp elements in this region with their poly(A) tracts facing toward the centromere, except for the most distal copy Z-1R, which was inverted. The region of telomeric repeats is indicated by a striped box. The stippled box on the other side shows a region of unique subtelomeric sequence. The bar under Z-3A indicates the position of the probe fragment.

Interestingly, this 2.3 kbp *HincII* fragment did not hybridize with a probe of telomeric repeats (pCHT-1; Higashiyama *et al.*, 1995), as shown in Figure 1A. Similar results were also obtained with other restriction enzymes, for example *SacI*, which gave two *Bal31*-susceptible and Zepp-hybridizing bands (1.5 and 0.8 kbp in size; data not shown).

Cloning of the left arm terminal region of *Chlorella* chromosome I that contains a cluster of Zepp elements

In our previous work, a 1.3 kbp *HindIII* fragment containing the left arm (the orientation of the chromosome molecule is reversed in this work in comparison with other systems) telomeric repeats (500 bp) and a 5'-truncated Zepp element of chromosome I was cloned and sequenced (DDBJ/EMBL/GenBank databases accession no. D26375). Because Zepp elements extend further into the proximal region (several kbp) on chromosome I, the DNA fragments covering this region were screened from a λ DASHII/*Chlorella* chromosome I library by plaque hybridization with the same probe as above (probe A). The maps of two clones thus obtained (I-3 and I-8) are shown in Figure 1C. As seen here, there was a gap between the 1.3 kbp *HindIII* fragment and the I-8 clone. A fragment of ~1.4 kbp encompassing this gap was obtained by PCR and cloned (pYM25) as shown in Figure 1B. The clone pYM25 gave a restriction map exactly consistent with that of chromosome I. The nucleotide sequence determined at the rightmost end of clone I-3 was completely different from Zepp sequences. Thus, these clones together covered a region of 16 kbp containing all copies of the Zepp elements of chromosome I.

Zepp elements are in a tandem array with their poly(A) tails facing toward the centromere

An ~15 kbp region of the left arm of *Chlorella* chromosome I containing Zepp elements was completely sequenced (DDBJ/EMBL/GenBank accession no. D82879). The results are schematically summarized in Figure 2. There

Z-3A: CTGACAGGTTTCCCAATATCCTATTGCATGCGTGCTGG
 Z-4A:
 Z-6 :
 Z-1R:
 Z-5 :
 Z-2A: T

Z-3A: GAAGCGTGCATTGGCACCTGTGTAATAGCGGGG-CTCCC
 Z-4A: C G T
 Z-6 : A G
 Z-1R: C G T
 Z-5 : CA G
 Z-2A: - C - G T

Z-3A: CCTTT----- (A) 11T (A) 17 taattgctgcccagtg
 Z-4A: TTTTTTTT (A) 16 gcgtgcacgaggactc
 Z-6 : ----- (A) 27 tcattgtaaggaggattt
 Z-1R: T----- (A) 15 gggtttagggtttag
 Z-5 : ----- (A) 22 acctttgcagcgaagc
 Z-2A: T----- (A) 20 ggaatcgacacgcgcg

Fig. 3. Nucleotide sequences of the 3'-region of Zepp elements. Only nucleotide differences compared with a representative element Z-3A are indicated. All the elements terminate in a poly(A) tract of different length, followed by different flanking sequences (small letters). The flanking sequence of the Z-1R element is the telomeric repeats (italic letters). The position of a putative polyadenylation signal is underlined. For the location of each element on chromosome I, see Figure 2.

were six poly(A) tracts of various sizes (ranging from 15 to 27 adenine residues) facing toward the centromere, except for the most terminal one, which was in the reverse orientation. Matrix comparison analysis of the nucleotide sequence of this entire region revealed that the sequences immediately distal to the poly(A) tracts were highly conserved, i.e. the 380 bp sequence of Z-2A flanked by two poly(A) tracts was almost perfectly co-linear with the corresponding 380 bp regions of Z-1R (reversed), Z-3A, Z-4A, Z-5 and Z-6. In the same way, the nucleotide sequence (1166 bp) of Z-5 was highly homologous with that of the corresponding regions of Z-1R (reversed), Z-3A, Z-4A and Z-6, and Z-4A (3291 bp) coincided with the 3291 bp region of Z-3A.

Nucleotide sequences of the 3'-end of these elements (~120 bp) containing the poly(A) tails and their 3'-flanking sequences (16 bp) are compared in Figure 3. There was a putative polyadenylation signal, 5'-AATAT, ~60 bp upstream of the poly(A) tract in each element. The 3'-flanking sequence of Z-1R was made up of telomeric repeats. It is noteworthy that there was no discernible homology among these 3'-flanking sequences, indicating no specific target sequence for these Zepp elements.

At first glance, this tandem array structure of poly(A)-bearing elements of various sizes remarkably resembled multiple HeT-A and TART retrotransposable elements attached to the *Drosophila* chromosomal termini (Biessmann *et al.*, 1992; Levis *et al.*, 1993; Mason and Biessmann, 1995). The transposition of HeT-A and TART are thought to be a direct addition to the ends of the chromosomes. However, a very different picture emerged for Zepp elements after detailed characterization of the most distal part of the array, where genuine 5'-portions of individual Zepp copies (Z-2B, Z-3B and Z-4B) were found to have accumulated.

Fig. 4. Schematic representation of the structures of Zepp members. The Zepp members Z-1R, Z-2, Z-3, Z-4, Z-5 and Z-6, all of which accumulated in the left arm of *Chlorella* chromosome I, are 5'-truncated copies of a larger element, for example ZA-1. ZA-1 was obtained from a *Chlorella* genomic library and contained an extended 5'-region (dotted line), where a reverse transcriptase-like ORF was located. The elements shared extensive homology (>90%) except for a 1290 bp 5'-portion of Z-6 (shaded box), whose homology to the corresponding region of ZA-1 is 75%. A poly(A) tail at the 3'-end of each element is indicated with a reiteration number. The arrowhead represents the position where another Zepp element is inserted. Symbols for restriction enzymes are: B, *Bam*HI; E, *Eco*RI; H2, *Hinc*II; H3, *Hind*III; P, *Pst*I; Sp, *Sph*HI; X, *Xho*I.

Zepp elements integrated into the Zepp sequence itself as a target

Combining overlapping DNA sequences derived from six independent copies (Z-1–Z-6) revealed that these were all 5'-truncated derivatives and there should be a larger 'master' Zepp element somewhere on the genome. To obtain larger elements, a cosmid/*Chlorella* genomic library was screened by colony hybridization using as probe a 549 bp *Hind*III–*Sph*I fragment from the most 5'-region of Z-3A (Figure 2) that contained an ORF with homology to reverse transcriptase sequences (described below). The resultant clone (ZA-1) contained a 33 kbp insert, on which a reverse transcriptase-like ORF was confirmed. Using this copy of Zepp as a guide, the structure of the subtelomeric copies was deduced. A compilation of the Zepp copies (Z-1–Z-6) based on the determined sequence of ZA-1 is shown in Figure 4. Z-2 was made from Z-2B (612 bp) connected to Z-2A (380 bp), Z-3 consisted of Z-3B (679 bp) and Z-3A (4111 bp) and Z-4 contained Z-4B (314 bp) combined with Z-4A (3291 bp). These partitioned elements resulted from insertion of one element into the other, i.e. Z-3 inserted into Z-2, Z-4 into Z-3 and Z-5 into Z-4 (arrows in Figure 4). Evidence for integration of these elements was supported by the presence of the small direct repeats flanking them; for example, 5'-TAAT for Z-3, 5'-GCGTGCACGCG for Z-4 and 5'-ACCTTTGC for Z-5 (Figure 5). Since the size and sequence of the flanking direct repeats differed among individual elements, Zepp elements do not seem to insert in a sequence-specific manner.

Zepp ZA-1 contains an ORF with homology to conserved domains of non-viral retroposon reverse transcriptase

The clone ZA-1 contained an ORF of 1317 amino acid residues in length (DDBJ/EMBL/GenBank accession no.

Fig. 5. Target site duplications flanking individual Zepp elements. Boxes represent inserted Zepp elements in which five nucleotides at the 5'-end and the poly(A) tail at the 3'-end are indicated. The flanking short duplications are underlined. There were no flanking duplications for the Z-2 element.

D83919). The corresponding ORF of ~580 amino acid residues could also be identified at the 5'-terminal region of Z-3 (Figure 4), but occasional point mutations introduced stop codons and disrupted the reading frame in this clone. Therefore, the Z-3 ORF may be a pseudogene. Database searches, using the FASTA program (Pearson and Lipman, 1988), revealed a significant relatedness (FASTA score 460) of the ZA-1 ORF to the reverse transcriptase of *Xenopus laevis* Tx1 (Garrett *et al.*, 1989). Seven amino acid domains totalling 178 residues that are highly conserved in all reverse transcriptases so far characterized (Xiong and Eickbush, 1990) were identified in this ORF. The amino acid sequences of these seven domains are compared among several non-viral retrotransposons in Figure 6. There is a 36% sequence identity between the ZA-1 and Tx1 ORFs. Zepp, together with

Fig. 6. Amino acid sequence homology within the conserved domains of known and putative reverse transcriptases. The amino acid sequence of the conserved domains within the Zepp ZA-1 ORF is compared with those of the reverse transcriptases of *X.laevis* retrotransposon Tx1 (Garrett *et al.*, 1989), human LINE (L1Hs; Hattori *et al.*, 1986), mouse LINE (L1Md; Xiong and Eickbush, 1988) and *Z.mays* retrotransposon Cin4 (Schwarz-Sommer *et al.*, 1987). The invariant amino acid residues are indicated by dark boxes. The residues conserved in more than three of the five retrotransposons are shown by shaded boxes.

Cin4 of *Zea mays* (Schwarz-Sommer *et al.*, 1987), Tx1 of *X.laevis* (Garrett *et al.*, 1989), L1Hs of human (Hattori *et al.*, 1986) and L1Md of mouse (Xiong and Eickbush, 1988), are located on the major non-LTR retrotransposon branch of the phylogenetic tree (Xiong and Eickbush, 1990). The ZA-1 clone still has a truncated 5'-region, therefore, it is not clear whether there is another ORF located further upstream in a 'full-size' copy. Identification of the reverse transcriptase-like ORF in Zepp elements, a characteristic feature of this family of retrotransposons, revealed an extended 3'-non-coding region (2.4 kbp). The *Drosophila* telomeric retrotransposons HeT-A and TART also have long 3'-non-coding regions (Levis *et al.*, 1993; Sheen and Levis, 1994), whose structures may be important for telomere-specific recognition and for the unusual transposition of these elements. Alternatively, they may serve only to increase the amount of DNA added to the chromosome end with each transposition (Mason and Biessmann, 1995).

A Zepp cluster also resides at the telomeric region of another *Chlorella* chromosome

The results of Figure 1A suggest that at least one more chromosomal terminus, besides the chromosome I end, contains Zepp elements. To see whether the Zepp elements may be arranged at that position in a similar array as seen in the chromosome I terminus, the 2.3 kbp *HincII* fragment gradually digested with *Bal31* and hybridized with the Zepp probe (Figure 1A) was cloned and characterized. After filling the most terminal end using Klenow enzyme, the 2.3 kbp band was digested with *SacI* to ligate with the *SacI* and *HincII* sites of the vector (pBluescript II). *SacI* was used because it gave smaller *Bal31*-susceptible and Zepp-hybridizing bands as mentioned above. Several clones were obtained by colony hybridization with probe A (Figure 1B). Nucleotide sequences determined for two of the clones (DDBJ/EMBL/GenBank accession no. D89938) showed essentially the same results. In one case, the 759 bp insert contained three Zepp copies in a tandem

array with their poly(A) tails facing towards the centromere (a 30 bp 3'-end with A₁₈, a 388 bp 3'-end with A₂₃ and a 341 bp middle region of Zepp). In the other case, the Zepp arrangement was the same but the most distal copy was a 256 bp 3'-end of Zepp with A₁₀. The blunt ends of each clone were not *HincII* sites and were different from each other, indicating that these clones were not contaminants from other chromosomal regions produced by *HincII* digestion but were derived from a chromosomal end. Most interestingly, there were no canonical telomeric repeats attached to the Zepp sequences, suggesting that Zepps may take the place of the telomeric repeats at this chromosome terminus. Which chromosomal end these clones were derived from is currently unidentified.

Expression of Zepp-related sequences

To detect the expression of Zepp elements, Northern blot hybridization of total RNA isolated from *Chlorella* cells was carried out using probe A (Figure 1B). The results shown in Figure 7 indicate that full-sized Zepp elements >6.0 kb in size were not expressed in the normal growth cycle. However, smaller smeared RNA bands of ~1.5–2.0 kb in size were detected by hybridization. The broad banding pattern of these RNAs suggested heterogeneously sized transcripts which might be initiated at internal promoters or might be degradation products of the full-length transcript. We examined the effects of various growth conditions on the expression of Zepp elements and found that a heat shock treatment where the growth temperature was shifted from 25 to 37°C for 5 min could efficiently induce transcription of very large (~10 kb) RNA that might correspond to a full-length Zepp (Figure 7). Stress-induced transcripts of retrotransposons are well documented for various systems (Hirochika, 1993; Pouteau *et al.*, 1994). Expression of Zepp elements may depend on non-telomeric Zepp copies that serve as templates for transcription. Telomeric copies that become progressively truncated at their 5'-regions may be incapable of expression.

Fig. 7. Northern blot hybridization of total RNA isolated from *Chlorella* cells with the Zepp sequence as probe. Lanes 1–3 contained RNAs (20 µg) isolated from cells grown at 25°C (control), from cells treated at 30°C for 5 min and from cells treated at 37°C for 5 min respectively. The positions of 26S and 18S rRNAs are indicated on the left side. The arrow on the right side indicates a Zepp transcript of approximately full size (6–7 kb).

Discussion

The Zepp elements in Chlorella are LINE-like retrotransposons

Zepp is a family of repetitive sequences which can serve as subtelomeric or telomeric elements in *Chlorella* chromosomes. Sequence analysis of the elements and their flanking sequences indicated that the elements arrived at their locations by successive transpositions. Several lines of evidence strongly suggest that Zepp elements are non-viral (LINE-like) retrotransposons and that insertion of them into the subtelomeric region of *Chlorella* chromosome I occurred after reverse transcription of RNA intermediates. (i) The *Chlorella* genome contains many Zepp copies whose sequences are closely related to one another, especially at their 3'-ends. Most Zepp copies are heterogeneously truncated at their 5'-ends, apparently as a result of reverse transcription ending before the end of the template had been reached (Hutchinson *et al.*, 1989). (ii) The Zepp elements terminate in poly(A) tracts, as do other LINE-like retrotransposons (Weiner *et al.*, 1986; Schwarz-Sommer *et al.*, 1987; Boeke and Corces, 1989). A putative polyadenylation signal is located ~60 bp upstream of the poly(A) tracts. The poly(A) tract at the 3'-end supports the idea that the element arises by reverse transcription of poly(A) RNA. (iii) The Zepp elements lacked long terminal repeats other than the target site duplications, which distinguishes them from viral transposons. (iv) Each element, except for a few cases at the chromosome terminus, is flanked by a direct duplication of nucleotides, but both the sequence and number of duplicated bases differs at each independent location of the element. The duplication presumably arose as a consequence of the repair of staggered breaks at the insertion site. (v) Larger Zepp elements contain an ORF with strong homology to retrovirus reverse transcriptase. (vi) An almost full-sized RNA is transcribed from Zepp

elements after heat shock treatment. By these criteria, Zepp elements qualify as a family of non-viral retrotransposons.

In addition to the six copies described above on the chromosome I terminus and three copies on the unidentified chromosomal end, we also obtained and sequenced five other Zepp clones that were screened by plaque hybridization from a *Chlorella* genomic library (unpublished results). All of them were 5'-truncated derivatives of the ZA-1 clone and found inserted within each other, suggesting that the target site for Zepp elements is generally a Zepp sequence itself at other positions on *Chlorella* chromosomes. We have not yet been able to determine the extreme distal sequence of a complete element.

A model for the mechanism of Zepp integration into chromosome I

Among various retrotransposons so far characterized, Zepp is peculiar in that it sometimes accumulated in the chromosomal termini. In this respect, Zepp resembles HeT-A and TART of *D.melanogaster* (Biessmann *et al.*, 1992; Levis *et al.*, 1993; Mason and Biessmann, 1995) and TRAS1 of the silkworm *Bombyx mori* (Okazaki *et al.*, 1995). However, the manners in which these retrotransposons integrate apparently differ from each other: HeT-A and TART attach themselves by a poly(A) tail to broken chromosome ends. TRAS1 is inserted into the telomeric repeat sequences as a target site. On the other hand, Zepp integrated preferentially into itself as a target. Analysis of several Zepp copies on *Chlorella* chromosome I provided some insights into the process of retroposon integration: (i) the flanking direct duplication indicated that the elements integrate into staggered nicks; (ii) the variability of the flanking duplication both in size and nucleotide sequence suggested that the nicks are preformed randomly or formed by element-encoded functions (Sommer *et al.*, 1985); (iii) most of the integrated copies showed 5'-truncations, but there was no evidence for truncation at the 3'-end; (iv) since Zepp integrated into the Zepp sequence itself, the sequence of each target site is homologous with a part (5'-region) of the inserting element.

Based on these data, a model for the mechanism of Zepp integration can be envisaged, which is similar to that proposed for other elements (Schwarz-Sommer *et al.*, 1987; Pritchard *et al.*, 1988; Jensen and Heidmann, 1991). In the case of Zepp, RNA–DNA hybridization at the left side of the staggered nick, which blocks reverse transcription (Schwarz-Sommer *et al.*, 1987), would occur efficiently because Zepp integrates into itself.

Probably through this mechanism, Zepp elements were repeatedly inserted into the subtelomeric region of *Chlorella* chromosome I as shown in Figure 5. The structure of this region (Figure 2), however, is complicated and needs additional explanation. Questions to be answered are: (i) how the first Zepp copies (Z-6 and Z-2) reached there; (ii) how the most distal copy Z-1R was reversed; (iii) how the small telomeric repeats attached to the poly(A) tail of Z-1R (reversed copy). The answers must be compatible with the following observations: (i) there are no flanking duplications for the Z-6 and Z-2 copies (Figure 5); (ii) there are no subtelomeric repeat sequences except for Zepp elements on this arm, whereas the other arm contained an array of characteristic elements (14–

Fig. 8. A model for the mechanism of repeated integrations of Zepp elements into the subtelomeric region of *Chlorella* chromosome I. The telomeric repeats and subtelomeric regions are shown by striped and dark boxes respectively. Zepp copies (open boxes) with poly(A) tails (A/T) were integrated successively via an RNA intermediate (winding line). The most distal copy Z-1R was inverted by unknown mechanisms (recombination?) and became accessible to telomerase-telomeric repeats.

20 bp in size; Higashiyama *et al.*, 1995) of ~1.7 kbp; (iii) there are no telomeric repeat sequences in the region proximal to the Zepp array. The model we propose for the subtelomere restoration on *Chlorella* chromosome I is depicted in Figure 8. Here, after the first and second Zepp element insertions (Z-6 and Z-2) at the subtelomeric region by the Zepp mechanism, a portion at the chromosome end containing the telomeric repeats and 5'-regions of Z-6 and Z-2 might be lost by some mechanism. The other copies, Z-3, Z-4 and Z-5, were successively integrated in this order by the mechanism outlined in Figure 8. After addition of the inverted copy of Z-1R, which may have occurred by recombination, its poly(A) tail could serve as a substrate for telomerase. The Zepp structure found in the 2.3 kbp *HincII* fragment described above is consistent with the intermediates of this model. Otherwise, this structure may be in the process of formation of a 'normal' chromosome with telomeric repeats via this mechanism.

Possible functions of Zepp elements residing in the telomeric and subtelomeric regions

Since most of the Zepp elements resided at non-telomeric locations and only a few of the *Chlorella* chromosome

termini (at least two of 32) contained Zepp elements, telomere or subtelomere maintenance does not seem to be a primary function of the Zepp elements. Instead, the Zepp elements may serve a telomere-stabilizing role in certain circumstances or jump into subtelomeric regions and accumulate, resulting in a structural buffer for telomere protection. It is likely that Zepp elements play an important role in backing up the telomerase system that maintains telomeric repeats, as demonstrated in the Y' repeat of yeast (Lundblad and Blackburn, 1993). Similar to Zepp elements, Y' repeats are immediately proximal to the yeast telomeric repeats. The amplification and acquisition of Y' repeats could suppress the lethal *estI* mutation, a deficiency in telomeric elongation (Lundblad and Blackburn, 1993). The accumulation of Y' repeats may serve as a buffer to protect the chromosomal termini from progressive loss. This can serve as an alternative mechanism that restores telomeric function if the telomerase-telomeric repeat system is deficient or weakened.

Moreover, it is remarkable that telomeric repeats are actually replaced by retrotransposons in at least one *Chlorella* chromosomal end. In this respect, the *Chlorella* Zepp-chromosome system may be considered as a hybrid of the canonical eukaryotic telomere and the *Drosophila* telomere strategies. *Chlorella* Zepp elements would provide a good system to study the dynamic mechanisms involved in telomere formation and maintenance.

Materials and methods

Strains and growth conditions

Chlorella vulgaris C-169 was obtained from the culture collection of the Institute of Molecular and Cellular Biosciences, The University of Tokyo. Cells were cultured photosynthetically in modified Bristol medium (MBM) as described previously (Higashiyama and Yamada, 1991).

DNA and RNA preparation and techniques

Algal DNAs (nuclear, chloroplast and mitochondrial) were isolated and fractionated by ultracentrifugation in a CsCl/Hoechst 33258 gradient as described (Yamada, 1982). DNA was digested with *Bal31* and restriction enzymes and analyzed by Southern blot hybridization under standard conditions (Sambrook *et al.*, 1989). To detect gradual degradation of DNA fragments derived from the chromosomal termini by treatment with *Bal31*, total *Chlorella* chromosomal DNA (10 µg) or isolated chromosome I DNA (1 µg) in 500 µl 1× *Bal31* buffer (20 mM Tris-HCl, pH 7.2, 0.6 M NaCl, 12.5 mM MgCl₂, 12.5 mM CaCl₂, 1 mM EDTA) was treated with 5 U *Bal31* (Takara Shuzo) at 28°C for various periods. The reaction was terminated by adding EGTA to a final concentration of 50 mM. DNA samples were phenol extracted and precipitated with ethanol before digestion with *HincII*.

Chromosomal DNA molecules of *Chlorella* cells were separated by pulsed-field gel electrophoresis (CHEF) as described previously (Higashiyama and Yamada, 1991). To isolate chromosome I, CHEF (contour-clamped homogeneous electric field) gel electrophoresis was performed in a 1% low melting point agarose (InCert agarose; Takara) gel in 0.5× TBE (45 mM Tris, 45 mM boric acid, 1 mM EDTA, pH 8.3) with a 3 min switching interval at 5 V/cm for 24 h.

For hybridization using chromosome I DNA and its cloned fragments as probes, non-radioactive digoxigenin-dUTP labeling of the probe was carried out with a Boehringer kit according to the manufacturer's manual. Hybridization was performed in a mixture containing 50% formamide, 5× SSC (1× SSC is 0.15 M NaCl, 15 mM sodium citrate), 5% blocking reagent (Boehringer Mannheim), 0.1% Sarkosyl and 0.02% SDS for 20 h at 42°C. For detection of the Zepp sequences, a 800 bp *SphI*-*HindIII* fragment close to the poly(A) tail in the most distal Zepp copy (Z-1R) was used as hybridization probe (probe A, Figure 1).

Total RNA for Northern blot analysis was isolated from *Chlorella* cells according to Rochaix *et al.* (1988), blotted onto nylon filters and hybridized with ³²P-labeled probes. Washed filters were autoradiographed for 24–120 h. In some cases cells were subjected to heat shock treatment

at 30 or 37°C for 5 min. After standing for 6 h at room temperature, RNA was isolated as above.

Construction of genomic libraries

A library of *C.vulgaris* C-169 chromosome I was constructed as follows. Chromosome I DNA separated by CHEF gel electrophoresis was partially digested with *Sau3AI* (fragments of 9–23 kbp) and ligated to *Bam*HI-digested λ DASHII arms (Stratagene). After packaging with GigapackII Gold packaging extract (Stratagene), the phages were grown in *Escherichia coli* SRB(P2). The preparation yielded $\sim 2 \times 10^7$ p.f.u./ μ g DNA (Yamada *et al.*, 1993). A cosmid library of the whole *Chlorella* genome was prepared using cosmid pWE15 arms; *Sau3AI*-partial digests of *Chlorella* chromosomal DNA (30–42 kbp in size) were ligated to the *Bam*HI-digested cosmid. After packaging as above, the phages were grown in *E.coli* XL1-Blue MRF'. An efficiency of 1.3×10^5 c.f.u./ μ g DNA was obtained.

PCR amplifications

PCR was used to obtain a DNA clone encompassing the gap between the H1.3 and I-8 clones on the left arm of chromosome I (Figure 1B). PCR amplification was performed for 30 rounds with purified chromosome I DNA as template according to the method of Saiki *et al.* (1988) using a LA PCR kit (Takara). The forward primer was the 25mer 5'-CCCTAAACCCTAAACCCTTTTTTTT-3', corresponding to the boundary sequence between the telomeric repeats and the poly(A) tail of Z-1R (Figure 2), and the reverse primer was the 20mer 5'-GCCGAGGGATTGTGTGAAC-3', corresponding to the sequence between Z-1R and Z-2B (Figure 2). DNA fragments amplified in PCR experiments were purified on agarose gels and cloned into the pT7Blue T-vector (Novagen).

Isolation of the 2.3 kbp *HincII* fragment containing Zepp sequences and derived from a chromosomal terminus

The 2.3 kbp *HincII* band gradually digested with *Bal31*, which did not hybridize with a probe of telomeric repeats but did with a Zepp probe (Figure 1), was isolated and cloned as follows. After electrophoretic separation of total *Chlorella* chromosomal DNA digested with *HincII*, a portion containing DNA fragments corresponding to 2.2–2.4 kbp was cut from the agarose gel. Purified with a GeneClean kit II (Bio101), the DNA (1 μ g) was treated with Klenow enzyme (Toyobo) in a reaction mixture (25 μ l) containing 67 mM phosphate buffer, pH 7.4, 6.7 mM $MgCl_2$, 1 mM 2-mercaptoethanol, 0.1 mM dNTPs and 1 U enzyme at 37°C for 2 h. Following digestion with *SacI*, the DNA was ligated to the *SacI* and *HincII* sites of pBluescript II (Stratagene) and transformed into *E.coli* XL1-Blue MRF'. The corresponding DNA clones were selected by colony hybridization with the same probe (A) as shown in Figure 1B.

DNA sequencing and analysis

Restriction fragments containing the Zepp sequences were cloned into M13 mp18 and mp19. Single-stranded DNA was sequenced by the chain termination procedure with a kit (Auto Read Sequencing Kit; Pharmacia) using an Automated Laser Fluorescence (ALF) DNA Sequencer (Pharmacia).

Acknowledgement

We thank Shinya Maki for RNA preparation and Northern hybridization.

References

Biessmann, H. and Mason, J.M. (1992) Genetics and molecular biology of telomeres. *Adv. Genet.*, **30**, 185–249.
 Biessmann, H., Vargeisdottir, K., Lofsky, A., Chin, C., Ginther, B., Levis, R.W. and Pardue, M.-L. (1992) Het-A, a transposable element specifically involved in 'healing' broken chromosome ends in *Drosophila melanogaster*. *Mol. Cell. Biol.*, **12**, 3910–3918.
 Blackburn, E.H. (1991) Structure and function of telomeres. *Nature*, **350**, 569–573.
 Blackburn, E.H. (1992) Telomerase. *Annu. Rev. Biochem.*, **61**, 13–129.
 Boeke, J.D. and Corces, V. (1989) Transcription and reverse transcription of retrotransposons. *Annu. Rev. Microbiol.*, **43**, 403–434.
 Broun, P., Ganel, M.W. and Tanksley, S.D. (1992) Telomeric arrays display high levels of heritable polymorphism among closely related plant varieties. *Proc. Natl Acad. Sci. USA*, **89**, 1354–1357.

Finnegan, D.J. (1989) The I factor and I-R hybrid dysgenesis in *Drosophila melanogaster*. In Berg, D.E. and Howe, M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington DC, pp. 503–517.
 Ganel, M.W., Lapitan, N.L.V. and Tanksley, S.D. (1991) Macrostructure of tomato telomeres. *Plant Cell*, **3**, 87–94.
 Garrett, J.E., Kuntzon, D.S. and Carroll, D. (1989) Composite transposable elements in the *Xenopus laevis* genome. *Mol. Cell. Biol.*, **9**, 3018–3027.
 Gilson, E., Laroche, T. and Gasser, S.M. (1993) Telomeres and the functional architecture of the nucleus. *Trends Cell Biol.*, **3**, 128–134.
 Hattori, M., Kuhara, S., Takanaka, O. and Sakaki, Y. (1986) L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase related protein. *Nature*, **321**, 625–627.
 Higashiyama, T. and Yamada, T. (1991) Electrophoretic karyotyping and chromosomal gene mapping of *Chlorella*. *Nucleic Acids Res.*, **19**, 6191–6195.
 Higashiyama, T., Maki, S. and Yamada, T. (1995) Molecular organization of *Chlorella vulgaris* chromosome I: presence of telomeric repeats that are conserved in higher plants. *Mol. Gen. Genet.*, **246**, 29–36.
 Hirochika, H. (1993) Activation of tobacco retrotransposons during tissue culture. *EMBO J.*, **12**, 2521–2528.
 Hutchinson, C.A., Hardies, S.C., Loeb, D.D., Shehee, W.R. and Bdgell, M.-H. (1989) Lines and related retrotransposons: long interspersed repeated sequences in the eukaryotic genome. In Berg, D.E. and Howe, M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington DC, pp. 594–617.
 Jensen, S. and Heidmann, T. (1991) An indicator gene for detection of germline retrotransposition in transgenic *Drosophila* demonstrates RNA-mediated transposition of the LINE I element. *EMBO J.*, **10**, 1927–1937.
 Levis, R.W., Ganesan, R., Houtchens, K., Tolar, L.A. and Sheen, F.-M. (1993) Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell*, **75**, 1083–1093.
 Louis, E.J. and Haber, J.E. (1990) The subtelomeric Y' repeat family in *Saccharomyces cerevisiae*: an experimental system for repeated sequence evolution. *Genetics*, **124**, 533–545.
 Lundblad, V. and Blackburn, E.H. (1993) An alternative pathway for yeast telomere maintenance rescue *estI* senescence. *Cell*, **73**, 347–360.
 Mason, J.M. and Biessmann, H. (1995) The unusual telomeres of *Drosophila*. *Trends Genet.*, **11**, 58–62.
 Okazaki, S., Ishikawa, H. and Fujiwara, H. (1995) Structural analysis of TRAS1, a novel family of telomeric repeat-associated retrotransposons in the silkworm, *Bombyx mori*. *Mol. Cell. Biol.*, **15**, 4545–4552.
 Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparisons. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.
 Pouteau, S., Grandbastien, M.A. and Boccara, M. (1994) Microbial elicitors of plant defence responses activate transposition of a retrotransposon. *Plant J.*, **5**, 535–542.
 Pritchard, M.A., Dura, J.-M., Pelisson, A., Bucheton, A. and Finnegan, D.J. (1988) A cloned I-factor is fully functional in *Drosophila melanogaster*. *Mol. Gen. Genet.*, **214**, 533–540.
 Richards, E.J. and Ausubel, F.M. (1988) Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell*, **53**, 127–136.
 Rochemaix, J.-D., Mayfield, S., Goldschmidt-Clermont, M. and Erickson, J. (1988) Molecular biology of *Chlamydomonas*. In Shaw, C.H. (ed.), *Plant Molecular Biology*. IRL Press, London, UK, pp. 253–275.
 Rogers, J.H. (1985) The origin and evolution of retrotransposons. *Int. Rev. Cytol.*, **93**, 188–280.
 Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487–491.
 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
 Schwartz-Sommer, Zs., Lecleq, L., Gobel, E. and Saedler, H. (1987) *Cin4*, an insert altering the structure of the *A1* gene in *Zea mays*, exhibits properties of nonviral retrotransposons. *EMBO J.*, **6**, 3873–3880.
 Sheen, F.-M. and Levis, R.W. (1994) Transposition of the LINE-like retrotransposon TART to *Drosophila* chromosome termini. *Proc. Natl Acad. Sci. USA*, **91**, 12510–12514.
 Sommer, H., Carpenter, R., Harrison, B.J. and Saedler, H. (1985) The transposable element Tam3 of *Antirrhinum majus* generates a novel type of sequence alterations upon excision. *Mol. Gen. Genet.*, **199**, 225–231.
 Weiner, A.M., Deininger, P.L. and Efstratiadis, A. (1986) Nonviral retrotransposons: genes, pseudogenes, and transposable elements generated

- by the reverse flow of genetic information. *Annu. Rev. Biochem.*, **55**, 631–661.
- Xiong, Y. and Eickbush, T.H. (1988) Similarity of reverse transcriptase-like sequences of viruses, transposable elements and mitochondrial introns. *Mol. Biol. Evol.*, **5**, 675–690.
- Xiong, Y. and Eickbush, T.H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.*, **9**, 3353–3362.
- Yamada, T. (1982) Isolation and characterization of chloroplast DNA from *Chlorella ellipsoidea*. *Plant Physiol.*, **70**, 92–96.
- Yamada, T., Maki, S. and Higashiyama, T. (1993) Nucleotide sequence of a *Chlorella vulgaris* alpha-tubulin gene. *Plant Physiol.*, **103**, 1467–1467.
- Young, B.S., Pession, A., Traverse, K.L., French, C. and Pardue, M.-L. (1983) Telomere regions in *Drosophila* share complex DNA sequences with pericentric heterochromatin. *Cell*, **34**, 85–94.
- Zakian, V.A. (1989) Structure and function of telomeres. *Annu. Rev. Genet.*, **23**, 579–604.
- Zakian, V.A. (1995) Telomeres: beginning to understand the end. *Science*, **270**, 1601–1607.

Received on January 6, 1997; revised on February 20, 1997