A LINE-like transposable element in *Drosophila*, the *I* factor, encodes a protein with properties similar to those of retroviral nucleocapsids

A.Dawson, E.Hartswood, T.Paterson and D.J.Finnegan¹

Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, UK

¹Corresponding author

I factors are members of the LINE-like family of transposable elements and move by reverse transcription of an RNA intermediate. Complete *I* factors contain two open reading frames. The amino acid sequence encoded by the first of these, ORF1, includes the motif CX₂CX₄HX₄C that is characteristic of the nucleocapsid domain of retroviral gag polypeptides followed by a copy of the slightly different sequences CX₂CX₄HX₆C and CX₂CX₉HX₆C. The function of this protein is unknown. We have expressed this protein in *Escherichia coli* and *Spodoptera frugiperda* cells and have shown that it binds both DNA and RNA but without any evidence for sequence specificity. The properties of deletion derivatives of the protein indicate that more than one region is responsible for DNA binding and that the CCHC motif is not essential for this. The ORF1 protein expressed in either *E.coli* or *Spodoptera* cells forms high molecular weight structures that require the region of the protein including the CCHC motif for their formation. This protein can also accelerate the annealing of complementary single-stranded oligonucleotides. These results suggest that this protein may associate with the RNA transposition intermediates of the *I* factor to form particles that enter the nucleus during transposition and that it may stimulate both the priming of reverse transcription and integration. This may be generally true for the product of the first open reading frame of LINE-like elements.

Keywords: DNA binding protein/*Drosophila*/nucleocapsid/retrotransposon/transposable element

Introduction

Retrotransposons are eukaryotic transposable elements that transpose by reverse transcription of an RNA intermediate (Boeke and Corces, 1989) and are of two types. The first, the retroviral-like or LTR elements, have long direct repeats at their termini and contain open reading frames homologous to the *gag* and *pol* genes of retroviruses with some having a third open reading frame equivalent to an *env* gene (Pelisson *et al.*, 1994; Song *et al.*, 1994). Retrotransposons of the second type have an adenine-rich sequence at the 3' end of the coding strand but no terminal repeats. They are usually referred to as non-LTR retrotransposons or LINE-like elements after the Long

Interspersed Nuclear Elements that were first identified as repeat sequences in mammalian genomes (Singer and Skowronski, 1985; Hutchison *et al.*, 1989). Most LINE-like elements have two open reading frames. In each case the second of these appears to encode a reverse transcriptase (Gabriel and Boeke, 1991; Mathias *et al.*, 1991). The function of the products of the first open reading frames of LINE-like elements is less clear and there is little similarity between these sequences themselves or between them and other known proteins.

The *I* factor is a LINE-like element of *Drosophila melanogaster* (Bucheton *et al.*, 1984; Fawcett *et al.*, 1986) that was identified because it is responsible for a form of hybrid dysgenesis (Finnegan, 1989; Bucheton, 1990). Strains of *D.melanogaster* can be classified as being either inducer or reactive depending on whether or not they contain active *I* factors in their chromosomes. The female progeny from crosses between inducer males and reactive females are abnormal, showing reduced fertility and increased frequencies of germ-line mutations, a phenomenon known as I-R hybrid dysgenesis. This is associated with a greatly increased frequency of *I* factor transposition. The male progeny from such a cross are normal, as are the progeny of crosses between reactive males and inducer females (Finnegan, 1989; Bucheton, 1990).

The open reading frames of the *I* factor are 1287 bp (ORF1) and 3675 bp (ORF2) long and are separated by 53 bp. The amino acid sequence encoded by ORF2 contains reverse transcriptase and RNaseH domains similar to those found in retroviral *pol* genes (Fawcett *et al.*, 1986) and other LINE-like transposable elements, while that of ORF1 has a copy of the motif CX₂CX₄HX₄C (CCHC) similar to those in the basic nucleocapsid portion of retroviral gag polypeptides (Berg, 1990) and the coat protein of cauliflower mosaic virus (Franck *et al.*, 1980). This motif is also found in the amino acid sequences encoded by the first open reading frames of many other LINE-like elements including most of those identified in *D.melanogaster*, *BS* (Udomkit *et al.*, 1995), *Doc* (O'Hare *et al.*, 1991), *F* (Di Nocera and Casari, 1987), *G* (Di Nocera, 1988), *jockey* (Primagii *et al.*, 1988) and *R1Dm* (Jakubczak *et al.*, 1990), as well elements from plants (Schwartz-Sommer *et al.*, 1987; Leeton and Smyth, 1993; Wright *et al.*, 1996) and fungi (Cambareri *et al.*, 1994). In the *I* factor this motif is followed by a copy of the slightly different sequences CX₂CX₄HX₆C and CX₂CX₉HX₆C (Fawcett *et al.*, 1986). The *R2Dm* element of *D.melanogaster* (Jakubczak *et al.*, 1990) and *cre* (Gabriel *et al.*, 1990) and *ingi* (Aksoy *et al.*, 1990) elements in trypanosomes have more conventional 'zinc finger' motifs within their putative ORF1 proteins, but nothing similar is seen in the amino acid sequences encoded by first open reading frames of mammalian LINE elements (D'Ambrosio *et al.*, 1986; Loeb *et al.*, 1986; Demers *et al.*, 1989; Dombroski *et al.*, 1991).

Transcription of the *I* factor is initiated from an internal RNA polymerase II promoter contained within the first 30 bp of the element (McLean *et al.*, 1993) and ORF1 and ORF2 are thought to be translated from the resulting polycistronic messenger RNA (Bouhidel *et al.*, 1994). Transposition of these elements is regulated as it only occurs at high frequency in the ovaries of dysgenic females and is not seen in males or in females of an inducer strain. This is due, at least in part, to regulation of the *I* factor promoter (Chaboissier *et al.*, 1990) and sequences responsible for this lie between nucleotides 1 and 186 (McLean *et al.*, 1993; Udomkit *et al.*, 1996). Transcription and transposition occur at low levels in inducer females. This results from the presence of complete *I* factors, since a reactive strain can be converted to the inducer state simply by the addition of full-length elements (Pritchard *et al.*, 1988). This could be due to production of an *I* factor-encoded protein that functions as a transcriptional repressor (Bucheton *et al.*, 1984; Pelisson and Bregliano, 1987; Jensen *et al.*, 1995), but other explanations are possible (Udomkit *et al.*, 1996).

The precise mechanism by which the *I* factor transposes is not known but we have suggested that its RNA transposition intermediates are packaged into virus-like particles that transport them to the nucleus (Pritchard *et al.*, 1988). The similarity between ORF1 protein and both the nucleocapsid component of retroviral gag polyproteins and cauliflower mosaic virus coat protein suggests that it might be the major component of such particles. Translation of ORF1 is initiated more frequently than is that of ORF2 (Bouhidel *et al.*, 1994), as would be expected if these particles contained more copies of ORF1 protein than of the reverse transcriptase encoded by ORF2. Particles of this sort may be involved in transposition of mammalian LINE elements since human L1 RNA can be found in particles together with the product of the first open reading frame of L1Hs elements (Hohjoh and Singer, 1996) and transcripts of murine LINE-1 (L1Md) elements are also associated with ribonucleoprotein particles in teratocarcinoma cells (Martin, 1991).

In order to better understand the mechanism by which LINE elements transpose, we have investigated the properties of the *I* factor ORF1 protein expressed in *Escherichia coli* or *Spodoptera* cells. The results indicate that it is a non-specific nucleic acid binding protein that can form particles that could be intermediates in transposition.

Results

ORF1 encodes a protein with nucleic acid binding activity

We have expressed ORF1 in *E. coli* using pET-3 (Studier *et al.*, 1990). These cells produced a protein that reacted with antiserum raised against a peptide comprising the carboxy-terminal nine amino acids of ORF1 and that appeared to be of 33 kDa molecular weight as judged by SDS-PAGE. This protein, which we have named ORF1met₁₃₄, is smaller than that expected for full-length ORF1 protein, but corresponds in size to a protein resulting from translation initiating at the second methionine codon, codon 134. Presumably, the second AUG is preceded by a sequence that can act as a ribosome binding site in *E. coli* whereas the first is not. In order to express the

Fig. 1. Map of deletion derivatives of ORF1 protein that have been expressed in *E. coli*. The box indicates the structure of full-length ORF1 protein with the region containing the CCHC motifs shown in black. Derivatives 1–8 have been made in *E. coli* using the plasmid vector pET-3. These appear to start at the second AUG codon in ORF1, codon 134. Derivatives 9–12 have been expressed using the vector pRSETC and have an amino-terminal extension of six histidine residues. The full-length protein has also been expressed in *Spodoptera* cells. The column on the right indicates whether or not a protein binds double-stranded DNA. Proteins 1–8 have been assayed by Southwestern blotting and proteins 9–12 by gel retardation. The amino acids of ORF1 present in each of these polypeptides are indicated below. The amino acids deleted from each polypeptide are indicated in brackets. 1, ORF1met₁₃₄; 2, ORF1met₁₃₄ with a five amino acid insertion between residues 200 and 201; 3, ORF1met₁₃₄Δ(200–266); 4, ORF1met₁₃₄Δ(200–361); 5, ORF1met₁₃₄Δ(361–426); 6, ORF1met₁₃₄Δ(267–426); 7, ORF1met₁₃₄Δ(200–426); 8, ORF1met₁₃₄Δ(141–200); 9, 6hisORF1; 10, 6hisORF1Δ(361–426); 11, 6hisORF1Δ(157–302); 12, 6hisORF1Δ(134–426). The filled triangle in protein 2 indicates the position of the five amino acid insertion in protein 2.

complete protein we have subcloned ORF1 downstream of DNA coding for an amino-terminal hexa-histidine tag using the pRSETC vector (Invitrogen). After induction, cells carrying this construct produced a protein, 6hisORF1, of 50–55 kDa. This is the size expected for full-length ORF1 protein plus the amino-terminal tag and it reacted with anti-ORF1 antisera in Western blots.

We have investigated the ability of ORF1 protein to bind DNA both by Southwestern blotting and by gel retardation using the full-length protein and a series of in-frame deletions. Each of the derivatives of ORF1met₁₃₄ (Figure 1, proteins 2–8) reacted with anti-ORF1 antiserum and was the size expected if translation had initiated at the second AUG codon. Both ORF1met₁₃₄ and its deletion derivatives were insoluble, so we have investigated their DNA binding abilities by Southwestern blotting. Extracts of cells expressing these proteins were subjected to SDS-PAGE (Figure 2A) and then renatured and transferred to nitrocellulose and probed with nucleotides 1–186 of the *I* factor labelled with ³²P. In each case the ORF1-derived protein was able to bind to double-stranded DNA (Figure 2B). This is a non-specific interaction, since binding of the probe was competed to the same extent by equimolar concentrations of either specific or non-specific competitors. The lack of sequence-specific binding was not due to release of bound Zn²⁺ during the Southwestern procedure, since the same results were seen in the presence of 10 μM Zn²⁺ (data not shown).

The full-length protein, 6hisORF1, was partially soluble, allowing us to purify it on a nickel resin so that we could examine its DNA binding activity by gel retardation (Figure 3A). As the protein concentration was increased,

Fig. 2. Binding of DNA by ORF1 and its derivatives expressed in *E. coli*. (A) SDS-PAGE of extracts of *E. coli* expressing deletion derivatives of ORF1 proteins. Extracts of cells made after induction were fractionated on an SDS-polyacrylamide gel and stained with Coomassie blue. The position of each ORF1 derivative is indicated with an asterisk. The numbering of the proteins is the same as that in Figure 1. These are, lane 1, protein 7; lane 2, protein 3; lane 3, protein 1; lane 4, protein 4; lane 5, protein 8; lane 6, protein 6; lane 7, protein 10; lane 8, protein 2; lane 9, protein 9; lane 10, protein 5; lane 11, protein 11; lane 12, protein 12; lane 13, protein 11; lane 14, protein 6; lane 15, protein 4. (B) Autoradiogram of a Southwestern blot of the extracts shown in (A). These were transferred to nitrocellulose and treated with nucleotides 1–186 of the *I* factor labelled with ^{32}P . The positions of the ORF1 derivatives are indicated by asterisks. The position of protein 12 that does not bind the probe is marked by an arrowhead.

a high molecular weight complex was formed that remained at the top of the gel. Lower molecular weight complexes were seen at intermediate concentrations before appearing at the top of the gel. This would be expected if monomers bind co-operatively to the probe. When 6hisORF1 protein was incubated with an oligonucleotide 35 bp long, the probe was still retained at the top of the gel (data not shown). The fact that this occurs with such a short oligonucleotide suggests that ORF1 protein can also form high molecular weight complexes in the absence of DNA.

Binding was not affected by either deletion of the CCHC motifs together with the flanking basic region (Figure 2B, lane 3) or insertion of five amino acids within the first CCHC motif (Figure 2B, lane 8). Binding was also seen after deletion of the highly basic carboxy-terminal region (Figure 2B, lane 6). The only ORF1 derivative that did not bind DNA was 6hisORF1 Δ (134–426) (Figure 2B, lane 12) which comprises the amino-terminal one-third of the molecule. The probe was bound to an intermediate extent by ORF1met₁₃₄ Δ (267–426) in these experiments (Figure 2A, lanes 6 and 14). This is probably due to its inability to renature fully during the Southwestern procedure or to poor transfer or binding to nitrocellulose rather than to loss of a DNA binding domain, since it bound the probe less well than did ORF1met₁₃₄ Δ (200–426) protein (Figure 2B, lane 1) which is contained within it.

In gel retardation experiments, equimolar amounts of each of the deletion derivatives of 6hisORF1 protein

Fig. 3. Binding of nucleic acids by full-length ORF1 protein. (A) Binding of ORF1 protein to DNA. Nucleotides 1–186 of the *I* factor labelled with ^{32}P were mixed with varying concentrations of 6hisORF1 protein and then run on a polyacrylamide gel. The concentration of 6hisORF1 protein used is indicated below each track. (B) Binding of ORF1 protein to RNA. A ^{32}P -labelled RNA containing nucleotides 1–366 of the *I* factor was mixed with varying concentrations of 6hisORF1 protein and then run on a polyacrylamide gel. The position of free probe is indicated by the arrow and the concentration of 6hisORF1 protein used is indicated below each track.

bound to double-stranded DNA with affinities similar to that shown by the full-length protein. The only exception was 6hisORF1 Δ (134–426) which bound the probe with <1% the affinity of that of the other ORF1 derivatives. We conclude from these results that the CCHC motifs are not essential for DNA binding and that more than one region within the carboxy-terminal two-thirds of ORF1 protein is required. An *I* factor probe comprising nucleotides 1–186 was not bound by ORF1 proteins in the presence of an excess of non-specific competitor DNA, indicating that it does not recognize this region in a sequence-specific manner.

We have also investigated the ability of full-length ORF1 protein to bind RNA in gel retardation experiments (Figure 3B). The results were similar to those using DNA probes. A high molecular weight complex appeared as the concentration of protein increased and again there is evidence for an intermediate form. Similar results were obtained using probes from either the 5' or 3' end of the *I* factor, as well as with one from another *Drosophila* transposable element, *mariner* (Medhora *et al.*, 1991). The protein bound each of these RNAs in a similar manner, suggesting that it does not recognize a specific sequence within these regions, the *I* factor.

We have compared the affinity of 6hisORF1 protein for double-stranded DNA, single-stranded DNA and RNA in a similar way. The 1–186 double-stranded DNA probe was mixed with 6hisORF1 protein in the presence or absence of various unlabelled nucleic acid competitors. The results indicate that both single-stranded DNA and RNA can compete more efficiently than double-stranded DNA, indicating that the protein binds more efficiently to these nucleic acids (Figure 4). The same is true of the nucleocapsid protein of HIV-1 (Lapadat-Tapolsky *et al.*, 1993).

ORF1 encodes a protein capable of multimerization

To examine whether full-length ORF1 protein is capable of forming macromolecular complexes, as suggested by

Fig. 4. Affinity of ORF1 protein for different nucleic acids. The 1–186 fragment of the *I* factor labelled with ^{32}P and at a concentration of 0.3 ng/ml was mixed with either a 20-fold or 200-fold excess of various unlabelled competitor nucleic acids and then run on a polyacrylamide gel. In the absence of competitor, ORF1 protein formed a high molecular weight complex with the probe and this was competed to a varying extent by the unlabelled nucleic acids. The competitors were, dsDNA, the *I* factor 1–186 fragment; ssDNA, denatured 1–186 fragment; tRNA, total yeast tRNA (Sigma); RNA, total RNA from *S.frugiperda* cells.

the results of the gel retardation experiments described above, we have fractionated 6hisORF1 protein produced in *E.coli* by gel filtration and centrifugation through sucrose gradients. In each case, fractions containing ORF1 protein were identified by Western blotting using rabbit anti-ORF1 anti-serum. The rate of sedimentation of 6hisORF1 protein in sucrose gradients was similar to that of thyroglobulin, 670 kDa, indicating that the ORF1 protein was in a high molecular weight form. This is supported by its behaviour during HPLC using a gel filtration column. At an initial concentration of 15 mM, pH 7.0 and 0.3 M NaCl, most of the protein eluted at molecular weights corresponding to 62 kDa and >1 MDa, with minor peaks corresponding to 120 kDa and 173 kDa suggesting that, under these conditions, ORF1 protein is present as a mixture of monomers and high molecular weight forms with some dimers and trimers.

We have investigated which region of ORF1 protein is required for formation of this high molecular weight complex by testing the ability of 6hisORF1 and its derivatives 6hisORF1Δ(157–302) (Figure 1, sequence 11), 6hisORF1Δ(361–426) (Figure 1, sequence 10) and 6hisORF1Δ(134–426) (Figure 1, protein 12) to multimerize. While full-length and 6hisORF1Δ(361–426) proteins migrated in sucrose gradients in an equivalent manner (Figure 4), the 6hisORF1Δ(157–302) and 6hisORF1Δ(134–426) proteins, both of which have lost the CCHC motifs, remained at the top of the gradient (Figure 5), indicating that they had not formed high molecular weight complexes.

We have investigated this further by glutaraldehyde cross-linking. As the concentration of glutaraldehyde was increased from 0.0005% to 0.05%, 6hisORF1 protein decreased in mobility until it remained at the top of the gel (Figure 6A). After treatment with the lower concentrations of glutaraldehyde it migrated at positions correspond-

Fig. 5. The ability of ORF1 protein and its derivatives to associate in high molecular weight forms. Full-length 6hisORF1 protein and its deletion derivative, 6hisORF1Δ(361–426), 6hisORF1Δ(157–302) and 6hisORF1Δ(134–426) were purified from *E.coli* cells carrying the corresponding plasmids. These proteins were then loaded together on a 20–60% sucrose gradient and centrifuged as described in Materials and methods. Proteins in each fraction collected from the gradient were fractionated on a 13.5% SDS–polyacrylamide gel together with tracks containing each of the purified proteins. The gel was then stained with Coomassie blue. Full-length 6hisORF1 protein sedimented as a high molecular weight complex together with the deletion derivative 6hisORF1Δ(361–426) that retains the CCHC sequence, whereas proteins 6hisORF1Δ(157–302) and 6hisORF1Δ(134–426) which have lost this sequence sedimented as structures with lower molecular weights.

Fig. 6. Glutaraldehyde cross-linking of ORF1 protein. (A) Full-length ORF1 protein expressed in *E.coli* (protein 8 in Figure 1) treated with varying concentrations of glutaraldehyde as described in Materials and methods and run on a polyacrylamide gel. The positions of protein complexes are indicated by arrows. The position of monomers is indicated by I and of putative dimers, trimers and tetramers by II, III and IV, respectively. (B) ORF1 protein lacking the CCHC motifs (protein 11 in Figure 1) treated with varying concentrations of glutaraldehyde and run on a polyacrylamide gel. The position of monomers is indicated by I and of putative dimers by II. Form III that migrates more rapidly than form I may be due to intra-molecular cross-linking of ORF1 protein.

ing to dimers, trimers and tetramers (Figure 6A, complexes II, III and IV). In contrast, 6hisORF1Δ(157–302) remained as a monomer even after treatment with 0.005% glutaraldehyde (Figure 6B). A putative dimeric form was produced after treatment with 0.05% glutaraldehyde, but no high molecular complexes were detected. This treatment also gave a form migrating more rapidly than monomers. This could be due to formation of intra-molecular lysine cross-links.

Taken together, these results suggest that the CCHC motifs, which have been deleted from 6hisORF1Δ(157–302) and 6hisORF1Δ(134–426), may be required for the protein–protein interactions that form ORF1 multimeric complexes.

Fig. 7. ORF1 protein made in insect cells forms discrete high molecular weight particles. Extracts of *Spodoptera* cells infected with baculovirus containing ORF1 downstream of the polyhedrin promoter (upper panel) or wild-type virus (lower panel) were centrifuged through 20–60% sucrose gradients. The protein contained in fractions collected from these gradients after centrifugation were run on SDS–polyacrylamide gels together with purified ORF1 protein (lane M). The proteins were transferred to nitrocellulose and probed with anti-ORF1 antibody.

Behaviour of ORF1 expressed in insect cells

We have investigated whether ORF1 protein can form macromolecular complexes in insect cells as well as *E. coli* by inserting ORF1 within the baculovirus virus expression vector pAcSGHisNT-B (PharMinogen) and growing this in *Spodoptera frugiperda* cells. This resulted in production of soluble full-length protein with an amino-terminal tag similar to that of 6hisORF1. Cells infected with either wild-type or recombinant virus were harvested 64 h after infection and extracts enriched for soluble high molecular weight material were loaded onto 20–60% linear sucrose gradients. Western blotting of fractions collected after centrifugation indicated the presence of ORF1 protein in a high molecular weight complex not seen in extracts of cells infected with wild-type virus (Figure 7). This complex sedimented more slowly than that isolated from *E. coli* cells expressing full-length ORF1 protein. The reason for this difference is unclear, but may lie in the more severe extraction procedures used to isolate the protein from bacteria.

ORF1 protein promotes the annealing of complementary oligonucleotides

During the course of this work we noticed that full-length ORF1 protein can form complexes with DNA that can be pelleted by low-speed centrifugation. This is also true of the nucleocapsid protein of HIV-1 (Lapadat-Tapolsky *et al.*, 1993) and is associated with an ability to accelerate the annealing of complementary single-strand nucleic acids (Tsuchihashi and Brown, 1994; George *et al.*, 1996). We have tested the effect of ORF1 protein on the annealing of complementary oligonucleotides 186 nucleotides long and have found that this process is stimulated ~300-fold as compared with the initial rate in the absence of protein. Figure 8 shows a time course for this reaction in the presence or absence of 20 nM 6hisORF1 protein. This activity was not present in identically prepared fractions from the same *E. coli* strain carrying the plasmid vector but no *I* factor DNA. The rate of annealing was greatest at protein concentrations between 11 and 23 nM.

Fig. 8. ORF1 protein stimulates the annealing of single-stranded DNA. A double-stranded oligonucleotide comprising nucleotides 1–186 of the *I* factor was denatured and allowed to reanneal in the presence (□) or absence (◇) of 20 nM ORF1 protein for varying periods of time. The proportion of the single-stranded DNA that had reannealed at each time point was assayed by separating single and double strands on a polyacrylamide gel.

Discussion

We have investigated the properties of the ORF1 protein of the *I* factor, a *Drosophila* LINE-like element, using both full-length and deleted proteins expressed in *E. coli* and insect cells. The *I* factor, like the majority of LINE-like elements, has two open reading frames, the second of which encodes a reverse transcriptase required for transposition. There is no direct evidence as to the function of the product of the first open reading frame, but it contains CCHC motifs like those in the nucleocapsid domains of retroviral gag polyproteins. The precise function of the CCHC motifs in gag polyproteins is uncertain but they appear to be required for stable packaging of genomic RNA in the first steps of virus formation (Aronoff, *et al.*, 1993; Dorfman *et al.*, 1993) and may be required for gag polyprotein multimerization (Franke *et al.*, 1994). The *I* factor ORF1 protein may play a similar role in packaging RNA transposition intermediates. If so, we should expect it to bind nucleic acids and be able to associate to form high molecular weight particles. The results of the experiments described above indicate that ORF1 protein has both of these properties.

Full-length ORF1 protein expressed in *E. coli* or *S. frugiperda* cells binds DNA, both double-stranded and single-stranded, and RNA. We have tried to map the region of the protein responsible for DNA binding using derivatives of the full-length protein. The results of these experiments indicate that there is no single sequence responsible for DNA binding and that the CCHC motifs are not essential for this. There must be at least two DNA binding sequences both located in the carboxy-terminal two-thirds of the molecule.

We have also found that ORF1 monomers can associate to form high molecular weight complexes in both *E. coli* and *S. frugiperda* cells. These may be similar to particles formed in *E. coli* cells by the *gag* gene product of Mason-Pfizer monkey virus (Klikova *et al.*, 1995) and the product of the first open reading frame of the yeast retrotransposon *Ty1* (Luschnig *et al.*, 1995). The region of ORF1 that includes the CCHC motifs is required for formation of

these complexes; they are not formed by ORF1 derivatives from which the CCHC motifs have been removed. There are CCHC motifs in similar positions in the ORF1 proteins of each of the LINE-like elements so far described in the genome of *D.melanogaster*, as well as in those detected in plants and fungi, but there is nothing similar in the ORF1 proteins of mammalian LINE elements, even though these have been found as ribonucleoprotein particles in the cytoplasm of human teratocarcinoma cells and mouse embryonic carcinoma cells (Martin, 1991; Hohjoh and Singer, 1996). There is a potential 'leucine zipper' motif in ORF1 protein of human LINE elements that, like the CCHC motif of the *I* factor ORF1 protein, is in a region of the molecule required for stable particle formation (Hohjoh and Singer, 1996), suggesting that these motifs may have similar functions. If this is the case, then we must ask why human LINE elements use a leucine zipper for a function that, in other LINE-like elements and in the more distantly related retroviruses, is fulfilled by CCHC motifs.

These results are consistent with our proposal that ORF1 protein helps to package *I* factor RNA and transport it to the nucleus where reverse transcription and integration can take place. Several groups have suggested that these events take place simultaneously at staggered single-strand breaks in chromosomal DNA (Schwartz-Sommer *et al.*, 1987; Pritchard *et al.*, 1988; Hutchison *et al.*, 1989; Bucheton, 1990), the 3' ends of which serve as primers for reverse transcription. A similar mechanism has been demonstrated *in vitro* for R2Bm elements that insert at a specific site in 28S rRNA genes (Dongmei *et al.*, 1993; Luan and Eickbush, 1995) and a nuclease that might initiate this process has been identified in the product of the second open reading frame of the human LINE element *LIHs* (Feng *et al.*, 1996). The ability of ORF1 protein to stimulate the annealing of complementary strands of DNA (Figure 8) may be required to condense the RNA transposition intermediate to allow it to be packaged into ribonucleoprotein particles. Alternatively, it may act during integration to stabilize the association of the 3' end of the *I* factor RNA transposition intermediate with the chromosomal primers of reverse transcription and second strand synthesis.

Transcription from the *I* factor promoter is reduced ~25-fold in ovaries of inducer as compared with reactive females due to the presence of complete *I* factors (Udomkit *et al.*, 1996). The most likely explanation for this was that *I* factor encodes a *trans*-acting transcriptional regulator. The product of ORF2 is unlikely to have this function, since expression of *I* factors as judged by their ability to induce hybrid dysgenesis can be regulated by an element with a defective copy of ORF2 (Jensen *et al.*, 1995). We have shown that at least one of the sequences responsible for regulating *I* factor expression in the presence of complete elements is located within the first 186 nucleotides of the *I* factor (Udomkit *et al.*, 1996). If the product of ORF1 were a *trans*-acting repressor we should expect it to bind to this region in a sequence-specific manner. We have found no evidence for this and presume that either ORF1 protein does not regulate *I* factor expression in this way, or that it does so in conjunction with another protein that can recognize this sequence. We conclude that the main function of ORF1 protein is to associate

with full-length *I* factor RNA to form transposition intermediates that facilitate reverse transcription and integration at new sites in the genome.

Materials and methods

DNA constructs

A 1.8 kb fragment encoding ORF1 was subcloned from pI407 (Bucheton *et al.*, 1984) into pET-3 (Novagen) to make plasmid pIF, the plasmid that expresses protein ORF1_{met134} (Figure 1, protein 1). Plasmid pL4 was generated by PCR. It codes for protein 2 in Figure 1 that contains amino acids PGNSV between residues 200 and 201 of ORF1. This insertion lies within the CCHC motif. Deletions of pIF were derived as follows, the numbers refer to the position in the sequence of the complete *I* factor, p71 (Figure 1, protein 3) deletion of a 200 bp internal *NdeI* (784)–*HincII* (984) fragment; pΔ3a (Figure 1, protein 4), deletion of a 483 bp internal *NdeI* (784)–*MaeI* (1267) fragment; pΔCt–*HincII* (Figure 1, protein 6), deletion of the sequence downstream of the *HincII* site at nucleotide 984; pCtΔ–*MaeI* (Figure 1, protein 5), deletion of the sequence downstream of the *MaeI* site at nucleotide 1267; pCtΔ–*NdeI* (Figure 1, protein 7), deletion of the sequence downstream of the *NdeI* site at nucleotide 784. Plasmid pPCRXI (Figure 1, protein 8) has deleted nucleotides 607–786 together with addition of 15 bases at the deletion breakpoint coding for PGNSV. This is a derivative of pL4. All internal deletions were constructed so that they are in frame.

Plasmid p6hisORF1 encoding full-length ORF1 protein with an amino-terminal tag of six histidine residues (Figure 1, protein 9) was generated by inserting a fragment containing the complete ORF1 sequence into the expression vector pRSETC (Invitrogen). This fragment was made by PCR using pI407 (Bucheton *et al.*, 1984) as template and oligonucleotides 335T (CGGGATCCTGACAGACCCACCAAAC) and 336T (GCGAATTCTATATGGATTGCTG) as primers. The plasmid p6hisORF1Δ (157–302) containing the corresponding fragment encoding ORF1 deleted for 146 amino acids including the CX₂CX₄HX₄C motif (Figure 1, protein 11) was made as follows. Two overlapping fragments were synthesized spanning the deleted region from nucleotides 654–1093, using oligonucleotides 335T and G6279 (GTTGAAACTGGACTCATACATCACCTAATATTCAC), and G6280 (GTCAATATTAGGTGATGGTATGAGTCCAGTTTCAAC) and 336T as primers. The products were combined as templates for a third PCR region using primers 335T and 336T. Plasmid p6hisORF1Δ(134–426) (Figure 1, protein 12) was made by PCR using pI407 as template and oligonucleotides 335T and M4126 (GCGAATTCTATATTTTAACTTCAGATAC) as primers. Fragments containing either the complete or deleted ORF1 sequence were purified by agarose gel electrophoresis, digested with *Bam*HI and *Eco*RI and then ligated to pRSETC digested with the same enzymes. Plasmid p6hisORF1Δ(361–426) (Figure 1, protein 10) was made by deleting the sequence downstream of the *MaeI* site at nucleotide 1267.

All recombinant plasmids were initially grown in *E.coli* AG-1 (*recA1*, *endA1*, *gyrA96*, *thi-1 hsdR17* [rk[−]mk⁺], *supE44*, *relA1*) and then transferred to *E.coli* BL21 DE3 pLysS (Studier *et al.*, 1990) for protein expression.

Protein purification

Escherichia coli BL21 DE3 pLysS strain carrying the recombinant plasmid were grown in an orbital shaker to OD₅₅₀ of 0.8 when they were induced for 4 h with 1 mM IPTG. Cells were harvested and lysed by sonication in buffer A (50 mM Tris–HCl, pH 8.0, 2 mM 2-mercaptoethanol, 1 mM PMSF, 1 mM benzamidine, 20% glycerol) containing 0.5 M NaCl, 1% Triton X-100, 1 mM imidazole. Following centrifugation to pellet insoluble material, the cell-free extract was applied to nickel affinity resin (QIA-express, Qiagen) pre-equilibrated with buffer A containing 0.5 M NaCl and 40 mM imidazole. Bound protein was washed with buffer A containing 1 M NaCl and 40 mM imidazole before eluting in 0.5 M NaCl buffer A with a gradient of imidazole from 40 to 500 mM. Peak fractions were pooled and eluted protein passed down a desalting column (Pharmacia, PD-10) and finally resuspended in 50 mM Tris–HCl, pH 8.0, 100 mM NaCl, 2 mM 2-mercaptoethanol, 20% glycerol. Aliquots were snap-frozen and stored at −70°C.

Gel retardation

A 186 bp fragment of DNA representing the 5' UTR of the *I* factor DNA was generated by PCR using primers 664C (GTCTAGACATTACCACTCAACCTCCG) and 403D (CGGATCCGATTGTT-

GGTTAAGGGCTTG) which contained sites for restriction enzymes *Xba*I and *Bam*HI respectively, and pI407 as template. The product was digested with these enzymes, purified by agarose gel electrophoresis and then end-labelled with [α - 32 P]dCTP using Klenow polymerase. The labelled DNA was purified by electrophoresis through a 4% native polyacrylamide gel and eluted overnight in 0.5 M sodium acetate, 0.1% SDS. Following removal of the acrylamide by spinning through glass wool, the probe was ethanol-precipitated and resuspended in TE buffer. For gel retardation experiments, the probe was incubated on ice for 20 min with purified protein at the stated concentration in binding buffer (50 mM HEPES, pH 8.0, 100 mM NaCl, 10 mM MgCl₂, 2 mM DTT, 10% glycerol, 0.1 μ g/ml BSA) in a final volume of 20 μ l.

The RNA probe containing sequences from the 5' end of the *I* factor was made using T7 RNA polymerase to transcribe the vector pSPT19 (Boehringer) containing a 5.6 kb *Sal*I fragment including a complete *I* factor. This was digested with *Rsa*I and then transcribed using the Riboprobe System T7 (Promega) in the presence of [α - 32 P]UTP. This gave a 495 nucleotide RNA including the first 366 nucleotide of the *I* factor. The probe containing sequences from the 3' end was made using a derivative of pSPT15.6 that had lost sequences up to the *Xba*I site at position 5208. This plasmid was digested with *Nru*I and then transcribed as described above. This gave a 517 nucleotide RNA including the last 263 nucleotide of the *I* factor. In each case the labelled RNA was purified on a 6% polyacrylamide gel before use.

Southwestern blots

Southwestern blots were performed on *E. coli* extracts expressing various regions of ORF1 protein. After separation by electrophoresis through a 13.5% SDS-polyacrylamide gel, the protein was transferred to nitrocellulose filters by electro-blotting. Immobilized protein was first denatured in 6 M guanidine hydrochloride in binding buffer (20 mM HEPES, pH 7.9, 3 mM MgCl₂, 40 mM NaCl, 10 mM 2-mercaptoethanol). The filter was then treated with successive 2-fold serial dilutions of guanidine hydrochloride in binding buffer and finally incubated for 2 h in binding buffer containing 4% non-fat dried milk. The filter was washed to remove excess blocking agent and the filter resuspended in binding buffer. The probe was added and incubated for 1–2 h at room temperature with gentle agitation. The filter was washed several times with wash buffer containing 0.01% Triton X-100 and air-dried before exposure to X-ray film.

Annealing assays

Annealing assays were performed essentially as described by Tsuchihashi and Brown (1994). A double-stranded DNA comprising nucleotides 1–186 of the *I* factor was denatured by incubation at 94°C for 5 min in TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). The standard reaction mixture (10 ml) contained 50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM dithiothreitol, 0.1% Triton X-100 and 3 mM MgCl₂. Reactions were stopped by the addition of 5 ml of 'stop' solution [0.25% bromophenol blue, 20% glycerol, 20 mM EDTA, 0.2% SDS, 0.4 mg yeast tRNA (Boehringer-Mannheim) per ml]. Samples were electrophoresed through an 8% polyacrylamide gel in 1× TBE (0.045 M Tris-borate, 0.001 M EDTA) and then the gel was dried for autoradiography and quantification using a Molecular Dynamics Phosphorimager.

Gel filtration

Gel filtration was carried out using a Rainin Dynamax 4.6 250 mm Hydropore-5-SEC column run at 30 ml/h. The column was equilibrated in buffer (50 mM Tris-HCl, pH 7.0, 100 mM NaCl, 5 mM EDTA, 5 mM 2-mercaptoethanol, 10% glycerol). Protein (10 μ g, 20 μ l) was applied directly to the column. The column was run at 0.4 ml/min. Eluted protein was collected in 50 μ l fractions and detected by measuring the absorbency at 280 nm. The molecular mass standards used (Bio-Rad) were bovine thyroglobulin, 670 kDa; bovine gamma-globulin, 150 kDa; chicken ovalbumin, 44 kDa; horse myoglobin, 17 kDa and vitamin B₁₂, 1.35 kDa.

Expression of ORF1 protein in *S.frugiperda* (Sf9) cells

Sf9 insect cells (a gift from Gibco-BRL) were propagated at 27°C in TC100 medium (Gibco-BRL) containing 10% fetal bovine serum. The ORF1 coding sequence was subcloned into the baculovirus transfer vector pAcSGHis NT-B (PharMingen) which had been digested with the restriction enzymes *Eco*RI and *Bgl*II. Recombinant baculovirus expressing ORF1 protein was then generated by co-transfection of 2 mg of purified recombinant plasmid together with 0.5 mg Baculogold™ virus DNA using the protocol and buffers supplied. Recombinant baculovirus was purified through two rounds of plaque purification and the final virus stock amplified from a single plaque. Infection of Sf9

cells with virus was carried out at a multiplicity of infection of 5 for 65–72 h.

Sucrose gradient analysis

Sucrose gradient analysis of recombinant baculovirus expressing ORF1 was performed essentially as described by Thomsen *et al.* (1992) except the procedure was modified to include a lysis step since ORF1 protein is not excreted into the medium. Briefly, following infection of Sf9 for 65 h with recombinant baculovirus, cells were harvested by centrifugation at 2000 r.p.m. (Denly BRL01) at 4°C. All manipulations were henceforth carried out on ice and all buffers pre-chilled. The pellet was resuspended in 1–2 ml TN buffer (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 200 mM PMSF) and cells lysed by vortexing with an equal volume of glass beads (40 mesh, Bio-Rad). Lysed cells were left on ice for 30 min. The lysate was pipetted up and down several times through a drawn-out Pasteur pipette and then clarified by centrifugation (3000 r.p.m., 4°C followed by 12 000 g). The high molecular weight component of this cleared lysate was then pelleted by centrifugation at 107 000 g, 4°C for 1 h. The pellet was resuspended in 500 μ l of TN buffer and layered carefully on to the top of a 14 ml 20–60% linear sucrose gradient in TN buffer. Following centrifugation at 35 000 r.p.m., 4°C for 15 h (Sorval, TH641 rotor), fractions of 400 μ l were collected from the bottom and stored on ice. The percentage sucrose of fractions following centrifugation was estimated using a refractometer (Abbe).

Fractions were analysed either by discontinuous SDS-PAGE (Laemmli, 1970) on a 10 or 12% gel followed by staining with Coomassie blue R250 or by fixing followed by electrophoretic transfer to nitrocellulose (Hybond C, 0.45 micron, Amersham) and Western blotting (Amersham, ECL) using polyclonal antibody raised against protein pIF derivative of full-length ORF1 protein.

Purified ORF1 protein expressed in bacteria was applied (10–20 mg) directly to the top of the sucrose gradient and then analysed as described above.

Glutaraldehyde cross-linking

Chemical cross-linking of ORF1 proteins with glutaraldehyde was carried out for 2 min in DNA binding buffer containing 100 mM NaCl at room temperature with varying concentrations of glutaraldehyde. The reactions were quenched by the addition of 2 M sodium borohydride dissolved in 0.1 M NaOH to give a final concentration of 100 mM. Cross-linked products were examined by SDS-PAGE and Western blotting with anti-ORF1 anti-serum.

Acknowledgements

We are grateful to D.Dryden for help with HPLC and to I.Clark, L.Powell and J.Smith for comments on a draft of this manuscript. This work was supported by grants from the European Union, the Human Frontier Science Programme, the Medical Research Council and the Wellcome Trust (036248/Z/92).

References

- Aksoy,S., Williams,S., Chang,S. and Richards,F.F. (1990) SLACS retrotransposon from *Trypanosoma brucei gambiense* is similar to mammalian LINEs. *Nucleic Acids Res.*, **18**, 785–792.
- Aronoff,R., Hajjar,A.M. and Linial,M.L. (1993) Avian retroviral RNA encapsidation: re-examination of functional 5' RNA sequences and the role of nucleocapsid Cys-His motifs. *J. Virol.*, **67**, 178–188.
- Berg,J.M. (1990) Zinc fingers and other metal-binding domains. *J. Biol. Chem.*, **265**, 6513–6516.
- Boeke,J.D. and Corces,V.G. (1989) Transcription and reverse transcription of retrotransposons. *Annu. Rev. Microbiol.*, **43**, 403–434.
- Bouhidel,K., Terzian,C. and Pinon,H. (1994) The full-length transcript of the *I* factor, a LINE element of *Drosophila melanogaster*, is a potential bicistronic RNA messenger. *Nucleic Acids Res.*, **12**, 2370–2374.
- Bucheton,A. (1990) I transposable elements and I-R hybrid dysgenesis in *Drosophila*. *Trends Genet.*, **6**, 16–21.
- Bucheton,A., Paro,R., Sang,H., Pelisson,A. and Finnegan,D.J. (1984) The molecular basis of I-R hybrid dysgenesis in *Drosophila melanogaster*: identification, cloning and properties of the *I* factor. *Cell*, **38**, 153–163.
- Cambareri,E.B., Helber,J. and Kinsey,J.A. (1994) Tad1-1, an active LINE-like element of *Neurospora crassa*. *Mol. Gen. Genet.*, **242**, 658–665.

- Chaboissier, M.-C., Busseau, I., Prosser, J., Finnegan, D.J. and Bucheton, A. (1990) Identification of a potential RNA intermediate for transposition of the LINE-like element I factor in *Drosophila melanogaster*. *EMBO J.*, **9**, 3557–3563.
- D'Ambrosio, E., Waitzkins, S.D., Witney, F.R., Salemm, A. and Furano, A.V. (1986) Structure of the highly repeated long interspersed repetitive DNA family (LINE or L1Rn) of the rat. *Mol. Cell. Biol.*, **6**, 411–424.
- Demers, G.W., Matunis, M.J. and Hardison, R.C. (1989) The L1 family of long interspersed repetitive DNA in rabbits: sequence, copy number, conserved open reading frames, and similarity to keratin. *J. Mol. Evol.*, **29**, 3–19.
- Di Nocera, P. (1988) Close relationship between non-viral retroposons in *Drosophila melanogaster*. *Nucleic Acids Res.*, **16**, 4041–4052.
- Di Nocera, P.P. and Casari, G. (1987) Related polypeptides are encoded by *Drosophila* F elements, I factors, and mammalian L1 sequences. *Proc. Natl Acad. Sci. USA*, **84**, 5483–5487.
- Dombroski, B.A., Mathias, S.L., Nanthakumar, E., Scott, A.F. and Kazazian, H.H. (1991) Isolation of an active human transposable element. *Science*, **254**, 1805–1808.
- Dongmei, D.L., Korman, M.K., Jakubczak, J.L. and Eickbush, T.H. (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell*, **72**, 595–605.
- Dorfman, T., Luban, J., Goff, S.P., Haseltine, W.A. and Gottlinger, H.G. (1993) Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus Type 1 nucleocapsid protein. *J. Virol.*, **67**, 6159–6169.
- Fawcett, D.H., Lister, C.K., Kellett, E. and Finnegan, D.J. (1986) Transposable elements controlling I-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian LINEs. *Cell*, **47**, 1007–1015.
- Feng, Q., Moran, J.V., Kazazian, H.H. and Boeke, J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, **87**, 905–916.
- 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.
- Franck, A., Guille, H., Jonard, C., Richards, K. and Hirth, L. (1980) Nucleotide sequence of cauliflower mosaic virus DNA. *Cell*, **21**, 285–294.
- Franke, E.K., Yuan, H.E.H., Leigh, B., Goff, S.P. and Luban, J. (1994) Specificity and sequence requirements for interactions between various retroviral Gag proteins. *J. Virol.*, **68**, 5300–5305.
- Gabriel, A. and Boeke, J.D. (1991) Reverse transcriptase encoded by a retrotransposon from the trypanosomid *Crithidia fasciculata*. *Proc. Natl Acad. Sci. USA*, **88**, 9794–9798.
- Gabriel, A., Yen, T.J., Schwartz, D.C., Smith, C.L., Boeke, J.D., Sollner-Webb, B. and Cleveland, D.W. (1990) A rapidly rearranging retrotransposon within the minixon gene locus of *Crithidia fasciculata*. *Mol. Cell. Biol.*, **10**, 615–624.
- George, J.A., Burke, W.D. and Eickbush, T.H. (1996) Analysis of the 5' junctions of R2 insertions with the 28S gene: implications for non-LTR retrotransposition. *Genetics*, **142**, 853–863.
- Hohjoh, H. and Singer, M.F. (1996) Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J.*, **3**, 630–639.
- Hutchison, C.A. III, Hardies, S.C., Loeb, D.D., Sehee, W.R. and Edgell, M.H. (1989) LINEs and related retroposons: long interspersed repeated sequences in the eucaryotic genome. In Berg, D.E. and Howe, M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 593–617.
- Jakubczak, J.J., Xiong, Y. and Eickbush, T.H. (1990) Type I (R1) and type II (R2) ribosomal DNA insertions of *Drosophila melanogaster* are retrotransposable elements closely related to those of *Bombyx mori*. *J. Mol. Biol.*, **212**, 37–52.
- Jensen, S., Cavarec, L., Gassama, M.-P. and Heidmann, T. (1995) Defective I elements introduced into *Drosophila* as transgenes can regulate reactivity and prevent I-R hybrid dysgenesis. *Mol. Gen. Genet.*, **248**, 381–390.
- Klikova, M., Rhee, S.S., Hunter, E. and Ruml, T. (1995) Efficient *in vivo* and *in vitro* assembly of retroviral capsids from gag precursor proteins expressed in bacteria. *J. Virol.*, **69**, 1093–1098.
- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.
- Lapadat-Tapolsky, M.H., Do Rocquigny, D., van Gent, B., Roques, B. and Plasterk, R. (1993) Interaction between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle. *Nucleic Acids Res.*, **21**, 831–839.
- Leeton, P.R.J. and Smyth, D.R. (1993) An abundant LINE-like element amplified in the genome of *Lilium speciosum*. *Mol. Gen. Genet.*, **237**, 97–104.
- Loeb, D.D., Padgett, R.W., Hardies, S.C., Sheher, W.R., Comer, M.B., Edgell, M.H. and Hutchison, C.A. (1986) The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons. *Mol. Cell. Biol.*, **6**, 168–182.
- Luan, D.D. and Eickbush, T.H. (1995) RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. *Mol. Cell. Biol.*, **15**, 3882–3891.
- Luschnig, C., Hess, M., Pusch, O., Brookman, J. and Bachmir, A. (1995) The gag homologue of retrotransposon Ty1 assembles into spherical particles in *Escherichia coli*. *Eur. J. Biochem.*, **228**, 739–744.
- Martin, S.L. (1991) Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol. Cell. Biol.*, **11**, 4804–4807.
- Mathias, M.L., Scott, A.F., Kazazian, H.H., Boeke, J.D. and Gabriel, A. (1991) Reverse transcriptase encoded by a human transposable element. *Science*, **254**, 1808–1810.
- McLean, C., Bucheton, A. and Finnegan, D.J. (1993) The 5' end of the I factor, a LINE-like retrotransposon of *Drosophila*, contains an internal promoter and sequences that regulate expression. *Mol. Cell. Biol.*, **13**, 1042–1050.
- Medhora, M., Mruyama, K. and Hartl, D.L. (1991) Molecular and functional analysis of the *mariner* mutator element *Mos 1* in *Drosophila*. *Genetics*, **128**, 311–318.
- O'Hare, K., Alley, M.R.K., Cullingford, T.E., Driver, A. and Sanderson, M.J. (1991) DNA sequence of the Doc retroposon in the *white-one* mutant of *Drosophila melanogaster* and of secondary insertions in the phenotypically altered derivatives *white-honey* and *white-eosin*. *Mol. Gen. Genet.*, **225**, 17–24.
- Pelisson, A. and Bregliano, J.-C. (1987) Evidence for rapid limitation of the I element copy number in a genome submitted to several generations of I-R hybrid dysgenesis in *Drosophila melanogaster*. *Mol. Gen. Genet.*, **207**, 306–313.
- Pelisson, A., Song, S.U., Prud'homme, N., Smith, P.A., Bucheton, A. and Corces, V.G. (1994) Gypsy transposition correlates with the production of a retroviral envelope-like protein under tissue specific control of the *Drosophila flamenco* gene. *EMBO J.*, **13**, 4401–4411.
- Primagii, A.F., Mizrokhi, L.J. and Ilyin, Y.V. (1988) The *Drosophila* mobile element *jockey* belongs to LINEs and contains coding sequences homologous to some retroviral proteins. *Gene*, **70**, 253–262.
- 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.
- Schwartz-Sommer, Z., Leclercq, L., Gobel, F. and Saedler, H. (1987) *Cin4*, an insert altering the structure of the *A1* gene in *Zea mays* exhibits properties of non-viral retrotransposons. *EMBO J.*, **6**, 3873–3880.
- Singer, M.F. and Skowronski, J. (1985) Making sense out of LINEs: long interspersed repeat sequences in mammalian genomes. *Trends Biochem. Sci.*, **10**, 119–122.
- Song, S.U., Kurukulos, M., Gerasimova, T., Boeke, J.D. and Corces, V.G. (1994) An *env*-like protein encoded by a *Drosophila* retroelement: evidence that gypsy is an infectious retrovirus. *Genes Dev.*, **8**, 2046–2057.
- Studier, W.F., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, **185**, 60–89.
- Thomsen, D.R., Meyer, A.L. and Post, L.E. (1992) Expression of feline leukaemia virus gp95 and gag proteins and assembly into virus-like particles using the baculovirus expression vector system. *J. Gen. Virol.*, **73**, 1819–1824.
- Tsuhiihashi, Z. and Brown, P.O. (1994) DNA strand exchange and selective DNA annealing promoted by the Human Immunodeficiency Virus Type 1 nucleocapsid protein. *J. Virol.*, **68**, 5863–5870.
- Udomkit, A., Forbes, S., Dalglish, G. and Finnegan, D.J. (1995) *BS* a novel LINE-like element in *Drosophila melanogaster*. *Nucleic Acids Res.*, **23**, 1354–1358.
- Udomkit, A., Forbes, S., Arkhipova, I., McLean, C. and Finnegan, D.J. (1996) Control of expression of the I factor, a LINE-like transposable element in *Drosophila melanogaster*. *EMBO J.*, **15**, 3174–3181.
- Wright, D.A., Ke, N., Smalle, B.M., Goodman, H.M. and Voytas, D.F. (1996) Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana*. *Genetics*, **142**, 569–578.

Received on December 2, 1996; revised on April 17, 1997