The EMBO Journal Vol.16 No.19 pp.6034–6043, 1997

Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon

Hirohiko Hohjoh[1] and Maxine F.Singer[1,2,3]

[1]Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892 and [2]Carnegie Institution of Washington, 1530 P Street NW, Washington, DC 20005, USA

[3]Corresponding author

Previous experiments using human teratocarcinoma cells indicated that p40, the protein encoded by the first open reading frame (ORF) of the human LINE-1 (L1Hs) retrotransposon, occurs in a large cytoplasmic ribonucleoprotein complex in direct association with L1Hs RNA(s), the p40 RNP complex. We have now investigated the interaction between partially purified p40 and L1Hs RNA *in vitro* using an RNA binding assay dependent on co-immunoprecipitation of p40 and bound RNA. These experiments identified two p40 binding sites on the full-length sense strand L1Hs RNA. Both sites are in the second ORF of the 6000 nt RNA: site A between residues 1999 and 2039 and site B between residues 4839 and 4875. The two RNA segments share homologous regions. Experiments involving UV cross-linking followed by immunoprecipitation indicate that p40 in the *in vitro* complex is directly associated with L1Hs RNA, as it is in the p40 RNP complex found in teratocarcinoma cells. Binding and competition experiments demonstrate that p40 binds to single-stranded RNA containing a p40 binding site, but not to single-stranded or double-stranded DNA, double-stranded RNA or a DNA–RNA hybrid containing a binding site sequence. Thus, p40 appears to be a sequence-specific, single-strand RNA binding protein.

Keywords: human high affinity binding site/LINE-1 (L1Hs)/non-LTR retrotransposon/RNA binding protein/ ribonucleoprotein complex

Introduction

LINE-1 (L1) is a retrotransposon found in all mammalian genomes; it belongs to the class of retrotransposons that lack long terminal repeats (LTRs) (Fanning and Singer, 1987; Hutchison *et al.*, 1989; Singer *et al.*, 1993). There are at least 100 000 L1 elements in the human genome (L1Hs) (Hwu *et al.*, 1986; Smit, 1996). Approximately 3–4% of these are full length, the remainder being truncated to varying extents, primarily at the 5′-end. Of the full-length elements, ~20–40 may be active, i.e. capable of retrotransposition (Sassaman *et al.*, 1997). Seven cloned L1Hs elements are known to retrotranspose when introduced into mammalian cell lines (Moran *et al.*, 1996; Sassaman *et al.*, 1997).

Active, full-length L1Hs elements have two open reading frames (ORFs). ORF1, the 5′-most ORF, encodes a 40 kDa protein, p40, which has been found in human teratocarcinoma and choriocarcinoma cell lines and in several kinds of tumor cells (Leibold *et al.*, 1990; Bratthauer and Fanning, 1992). ORF2 predicts an ~149 kDa protein with which are associated two activities, DNA endonuclease (Feng *et al.*, 1996) and reverse transcriptase (Dombroski *et al.*, 1991; Mathias *et al.*, 1991; Moran *et al.*, 1996; Sassaman *et al.*, 1997). Analysis of L1Hs elements with mutations in either p40 or the endonuclease or reverse transcriptase regions of the ORF2 protein indicates that all three are required for efficient transposition (Moran *et al.*, 1996).

Previously we showed that p40 occurs in a cytoplasmic ribonucleoprotein complex (called the p40 RNP complex) in human teratocarcinoma cells growing in culture (Hohjoh and Singer, 1996). This p40 RNP complex contains L1Hs RNA and the RNA appears to be directly bound to p40. The p40 RNP complex is large; it elutes in the void volume of a Sephacryl S-400 column and is thus likely to be >700 kDa. When the p40 RNP complex is treated with high salt, the RNA is released and p40 can be recovered as multimers in the size range 200 kDa; treatment of the p40 RNP complex with various ribonucleases yields similar multimers (Hohjoh and Singer, 1997). No components of the complexes other than L1Hs RNA and p40 have been identified. Similar RNP complexes, containing LINE-1 RNA and ORF1 protein, appear to occur in mouse embryonal carcinoma cells (Martin, 1991; Martin and Branciforte, 1993).

We have investigated the interaction of partially purified p40 multimers with L1Hs RNA *in vitro*. The results indicate that the p40 multimers obtained after high salt treatment of the p40 RNP complex bind single-stranded regions of L1Hs RNA in a sequence-specific manner.

Results

Preparation of p40

The source of p40 for these experiments was cytoplasmic extracts of the human teratocarcinoma cell line 2102Ep; these cells express a relatively high level of p40 (Leibold *et al.*, 1990). Similar extracts of HeLa cells, which contain little or no p40, were used as controls (Liebold *et al.*, 1990). Most of the p40 in the 2102Ep cell extracts is in the p40 RNP complex and sediments upon centrifugation at 160 000 *g* (Hohjoh and Singer, 1996). After incubation in 0.5 M NaCl this complex dissociates, releasing L1Hs RNA and p40 multimers (~200 kDa) that are significantly smaller than the original RNP complex (Hohjoh and Singer, 1997). Such p40 multimers were partly purified by fractionation on heparin columns for use in the current work (see Materials and methods for details). Those column fractions that contained p40, as determined by SDS–PAGE and Western blotting, were pooled and are

 © Oxford University Press

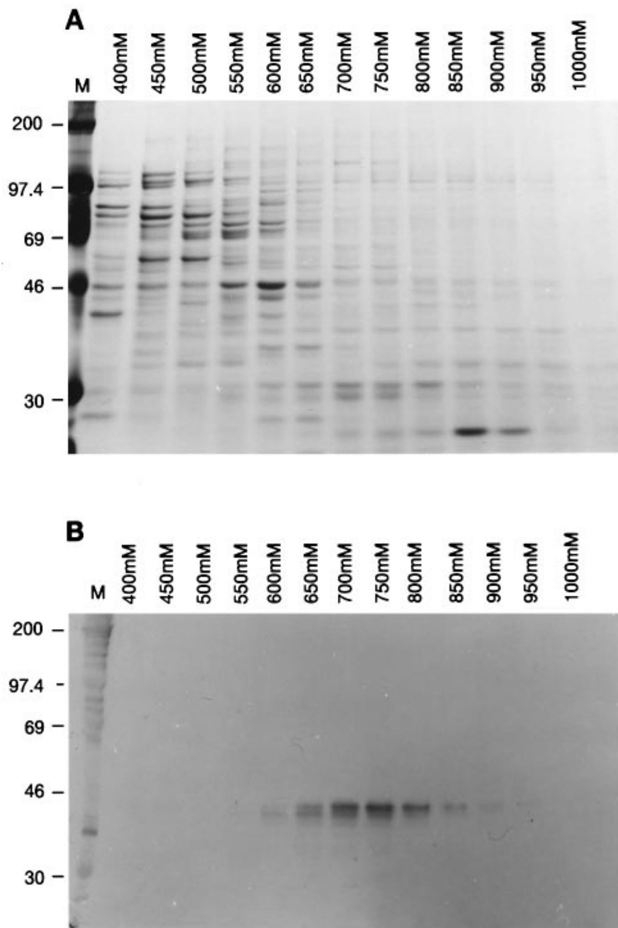

Fig. 1. Protein analysis of heparin column fractions. (**A**) SDS–PAGE profile of heparin column fractions. Twenty microliters of each fraction (0.5 ml) were used and gels were stained with Coomassie Brilliant Blue R-250 (Bio-Rad). The NaCl concentration of each fraction is indicated. Size marker proteins (kDa) in lane M are indicated by bars. (**B**) Western blot analysis of heparin column fractions. Proteins were separated as in (A) and electrophoretically blotted onto Immobilon-P membrane (Millipore). Membrane blocking, incubation with anti-p40 immune serum and biotinylated second antibodies and visualization of the antigen–antibody complexes by a colorimetric method were carried out as previously described (Hohjoh and Singer, 1996).

referred to here as the partially purified p40 multimer preparation. Figure 1A shows the results obtained when heparin column fractions were electrophoresed under denaturing conditions and stained for protein; the amount of p40 is too low to be detectable. Figure 1B shows p40 on the corresponding Western blots analyzed with anti-p40 IgG. While the heparin column separates p40 from substantial numbers of other proteins, many contaminating proteins are still present in the p40-containing fractions.

Chemical cross-linking with glutaraldehyde of the partially purified p40 multimer preparation followed by Western blot analysis (according to the procedures described by Hohjoh and Singer, 1996) confirmed that the p40 is in multimers (data not shown).

HeLa cell extracts were treated by the same procedures as the 2102Ep extracts and the pooled heparin column fractions, in which no p40 was detectable by Western blot analysis, were used as controls.

Binding of L1HsRNA to p40 multimer

The *in vitro* binding assay system used to study the interaction of p40 with L1Hs RNA depends on immunoprecipitation of p40 by specific anti-p40 IgG and co-immunoprecipitation of bound RNA segments. The strategy is detailed in Materials and methods. Briefly, the p40 preparation was incubated with radiolabeled sense strand L1Hs RNA synthesized *in vitro* using the cloned L1.2A L1Hs element as template; L1.2A is a full-length, active element (Dombroski *et al.*, 1991; Moran *et al.*, 1996). The mixture was then treated with RNase T1 (which specifically attacks the 3′-phosphate groups of guanine nucleotides) and subjected to immunoprecipitation with anti-p40 IgG. Finally, the L1.2A RNA fragments co-immunoprecipitated with p40 were analyzed on 8% sequencing gels.

When the p40 preparation was incubated with full-length L1.2A RNA as probe two intense bands (labeled A and B) were observed on the gel (Figure 2, lane 2). In addition, some faint bands of both greater and lesser mobility appeared (as well as a smear that was not reproducible). No RNA fragments were detected in the sample prepared with pre-immune IgG (Figure 2, lane 3), indicating that the RNA fragments seen in lane 2 are associated with p40 multimer. Lane 1 in Figure 2 displays all the fragments produced by RNase T1 digestion of full-length L1.2A RNA (no immunoprecipitation); these include bands with the same mobility as those co-immunoprecipitated with p40 as well as bands of lower and higher mobility. In other experiments (not shown) the RNA fragments were electrophoresed on 20% polyacrylamide gels in order to display any very small oligonucleotides that might be bound to p40 multimer; no additional bands of significant intensity were observed.

The longest stretch of L1.2A RNA between two guanine residues is from nucleotide 1999 to 2039, thus the longest RNA fragment expected after total digestion with RNase T1 is 41 nt. Fragments longer than 41 nt, such as those seen in the control sample without immunoprecipitation (Figure 2, lane 1), are likely to represent regions of the RNA protected from RNase T1 by secondary structure. Although the data in Figure 2, lane 1 suggest that full-length L1.2A RNA has a complex secondary structure, no bands >41 nt appear in lane 2, indicating that none of the regions protected by secondary structure interact with p40 multimer.

The specificity of the protein–RNA interaction was investigated further. When the protein preparation was obtained from HeLa cells, which contain little if any p40, no bands were observed in the samples immunoprecipitated by either immune or pre-immune IgG (Figure 2, lanes 5 and 6 respectively). This result is as expected if the presence of bands observed in lane 2 is dependent on p40. Earlier experiments demonstrated that β-actin mRNA is not associated with the large p40 RNP complex found in human teratocarcinoma cells (Hohjoh and Singer, 1996). When β-actin mRNA was used in the *in vitro* binding assay no oligonucleotide bands were observed in the immunoprecipitates obtained with either immune or pre-immune IgG (Figure 2, lanes 8 and 9 respectively). Altogether, the *in vitro* binding assay results indicate that there is specific binding between L1Hs RNA and p40 multimer, as suggested by earlier analysis of the cytoplasmic p40 RNP complex (Hohjoh and Singer, 1996).

Further, the data in Figure 2 suggest that association of p40 multimer with L1Hs RNA involves specific sequences on the RNA.

A series of additional experiments confirmed the results described in Figure 2. Results identical to those in Figure 2, lane 2 were obtained when the reaction mixture was immunoprecipitated prior to RNase T1 digestion rather than after digestion (data not shown); in this case unbound RNA fragments were removed by washing and recentrifugation of the immunoprecipitate. Thus binding of p40 multimers to the A and B segments occurs in the intact 6000 nt RNA chain and does not depend on prior digestion with RNase T1. The results were also the same if the reaction mixture was irradiated with UV prior to immunoprecipitation and subsequent RNase T1 digestion (data not shown). This indicates that no additional, relatively unstable interactions involving regions of L1.2A RNA other than those represented by the A and B segments occur.

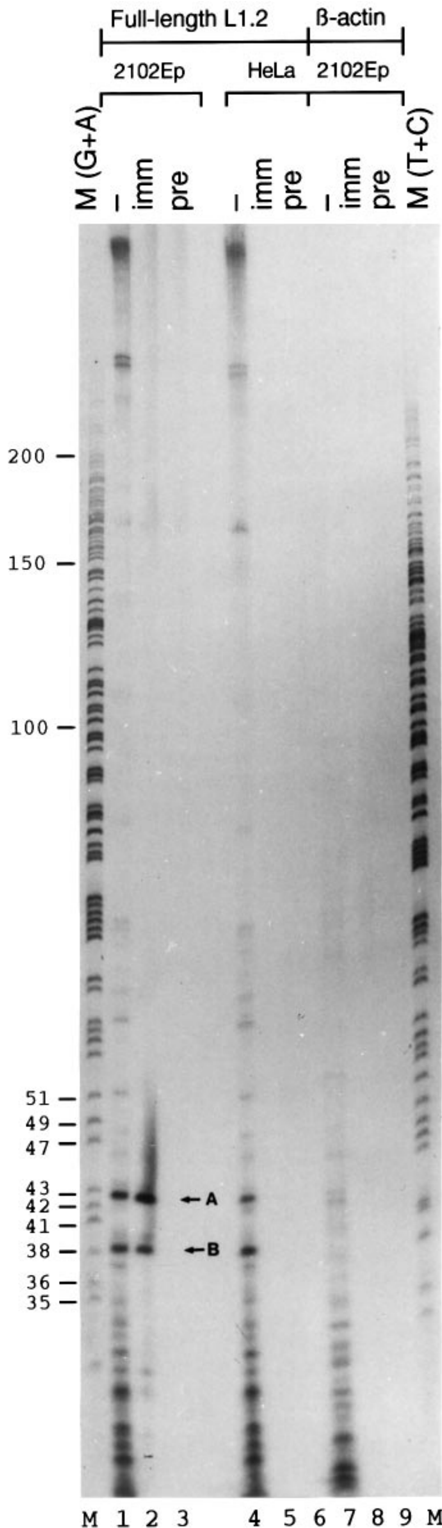

Fig. 2. *In vitro* RNA binding assay. The assay procedure is detailed in Materials and methods. The protein preparations used are indicated by the names of the cells, 2102Ep and HeLa, from which they were obtained. Radioactive RNA probes, full-length L1.2A and human β-actin, synthesized *in vitro* with [α-^{32}P]UTP are as indicated. The lanes containing RNA fragments collected in the absence of immunoprecipitation are indicated by –. The lanes containing RNA fragments obtained by immunoprecipitation with pre-immune and anti-p40 IgGs are indicated by pre and imm respectively. Arrows A and B show the intense bands co-immunoprecipitated with p40 by immune IgG. Marker lanes (M) are M13 sequence reaction mixtures (Amersham) and appropriate size marker bands (nt) are indicated by bars.

We next investigated whether p40 multimers would bind to the same fragments of L1.2A RNA if the protein was mixed with predigested RNA. L1.2A RNA was digested to completion with RNase T1 and mixed with the p40 preparation; the mixture was incubated and treated as described for the binding assay. The resulting gels again looked the same except that the B fragment appeared somewhat weaker than the A band (data not shown). Thus association of p40 multimer with the sequences in fragments A and B does not require complete full-length L1.2A RNA.

In another series of experiments we obtained similar results using L1.2A RNA labeled with [α-^{32}P]GTP rather than the usual [α-^{32}P]UTP (data not shown).

Characterization of the A and B RNA fragments

The longest oligonucleotides expected from complete RNase T1 digestion of L1.2A RNA are as follows (the numbers in parentheses indicate the residue numbers assigned in Dombroski *et al.*, 1991): 41 nt (1999–2039); 37 nt (4839–4875); 28 nt (3044–3071); 27 nt (2577–2603, 4896–4922, 5177–5203). The control digests of total L1Hs RNA shown in Figure 2 (lanes 1 and 4) contain prominent bands with mobilities corresponding to the 42–43 and 38–39 nt markers derived from M13 DNA (lanes M). As will be apparent from the experiments described below, these prominent bands are the expected RNase T1-resistant 41 and 37 nt fragments; the slight difference in mobility compared with the markers likely reflects the different mobilities of RNA and DNA segments. Because the L1.2A RNA fragments co-immunoprecipitated with p40 multimers by immune IgGs had the same mobility as the 41 and 37 nt fragments in the total digest, it seemed possible that bound bands A and B were themselves these fragments. Consistent with this possibility, no further digestion of bands A and B was observed when the RNA fragments were collected after immunoprecipitation, heat denatured, redigested with RNase T1 and then separated on a sequencing gel.

RNAs were prepared corresponding to four separate regions of L1.2A RNA which together span the entire

6000 nt RNA (Figure 3A). The standard binding assay was carried out with each of the RNA preparations (Figure 3B). Intense bands with the same mobility as the A and B bands observed when full-length L1.2A RNA was used (lane 2) were detected in the immunoprecipitates formed in the presence of fragments M1 (residues 1627–3425) (lane 7) and 3′L1 (residues 4327–6060) (lane 13) respectively. In contrast, no and a few barely detectable bands were observed with the 5′L1 (residues 1–1627) (lane 4) and M2 (residues 3185–4327) (lane 10) RNA probes respectively. No band was detected in the immunoprecipitates formed by pre-immune IgG with any of the four probes (lanes 5, 8, 11 and 14). Redigestion of the immunoprecipitated RNA fragments with RNase T1, as previously described, again left the A and B bands (from M1 and 3′L1 RNAs respectively) intact (data not shown). Thus the A and B fragments each reside in that portion of L1.2A RNA predicted to contain the RNase T1-resistant fragment of corresponding length. To confirm the location of the A and B fragments in the RNA, we examined the M1 and 3′L1 regions in more detail.

Figure 4A shows the results of binding assays carried out with RNAs representing subregions of M1. The A band was detected in the immunoprecipitates whenever residues 1999–2039 were contained within the probe (Figure 4A, lanes 2, 5 and 11). In contrast, no RNA band was co-immunoprecipitated when these residues were missing from the probe (lanes 8 and 14). The results confirm that the A band derives from the region between residues 1894 and 2172, the location of the 41 nt RNase T1-resistant segment.

Similar experiments using subregions of the 3′L1 segment as RNA probes are shown in Figure 4B. The B band was detected in the immunoprecipitates when the probe contained residues 4816–5322 (lane 5), whereas probes representing residues 4327–4837 (lane 2) or 5323 to the 3′-end of L1.2A (lane 8) showed no or a few barely detectable bands respectively. These results confirm that the B fragment derives from the region between residues 4838 and 5322, the location of the 37 nt RNase T1-resistant fragment.

The results described thus far suggest that the p40 multimer may be a sequence-specific nucleic acid binding protein. It is interesting in this regard that there are several relatively long RNase T1-resistant segments within L1.2A RNA other than the 41 and 37 nt fragments. These include, for example: a 19 nt sequence in 5′L1; 28 nt and 27 nt sequences in M1; two 23 nt sequences in M2; two 27 nt sequences in 3′L1. None of these segments are preferentially co-immunoprecipitated (Figure 3B, lanes 4, 7, 10 and 13). These observations suggest that the p40 multimer has high affinity for interaction with the 41 and 37 nt RNA regions. The faint bands of other lengths that were observed in the immunoprecipitates may represent regions with low affinity for the p40 multimer. We also note that the data in Figures 3B and 4 demonstrate that binding to each of sites A and B is efficient in the absence of the second site. Thus p40 multimers bind independently to each site.

As already pointed out, RNase T1 digestion of total L1.2A RNA (without immunoprecipitation) suggested that the 6000 nt RNA likely has a complex secondary structure (Figure 2). When total RNA digests of full-length L1.2A

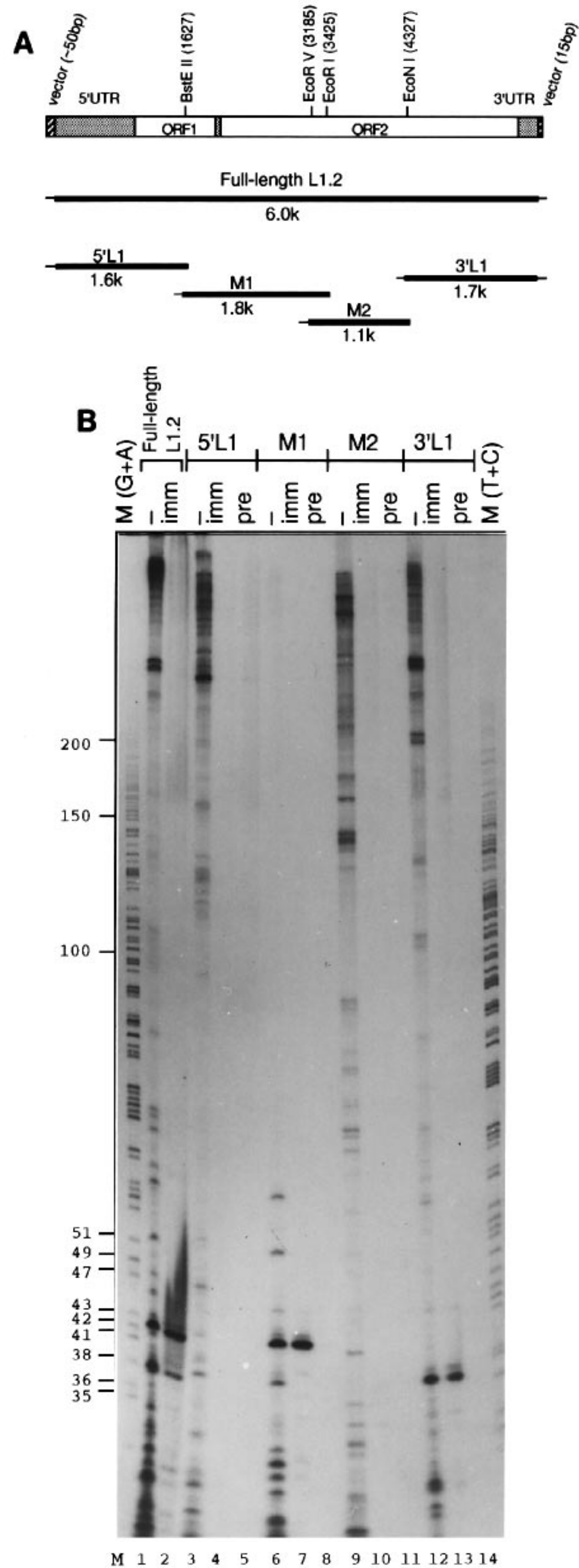

Fig. 3. *In vitro* RNA binding assay with subfragments of L1.2A RNA as probes. (**A**) Schematic drawing of the probes. The names and sizes (nt) are indicated above and below the solid bars respectively. The hatched boxes and thin lines show vector sequences. Figures in parentheses indicate the nucleotide positions and are based on the numbering used in Dombroski *et al.* (1991). UTR, untranslated region. (**B**) Binding assay. The procedures are as in Figure 2.

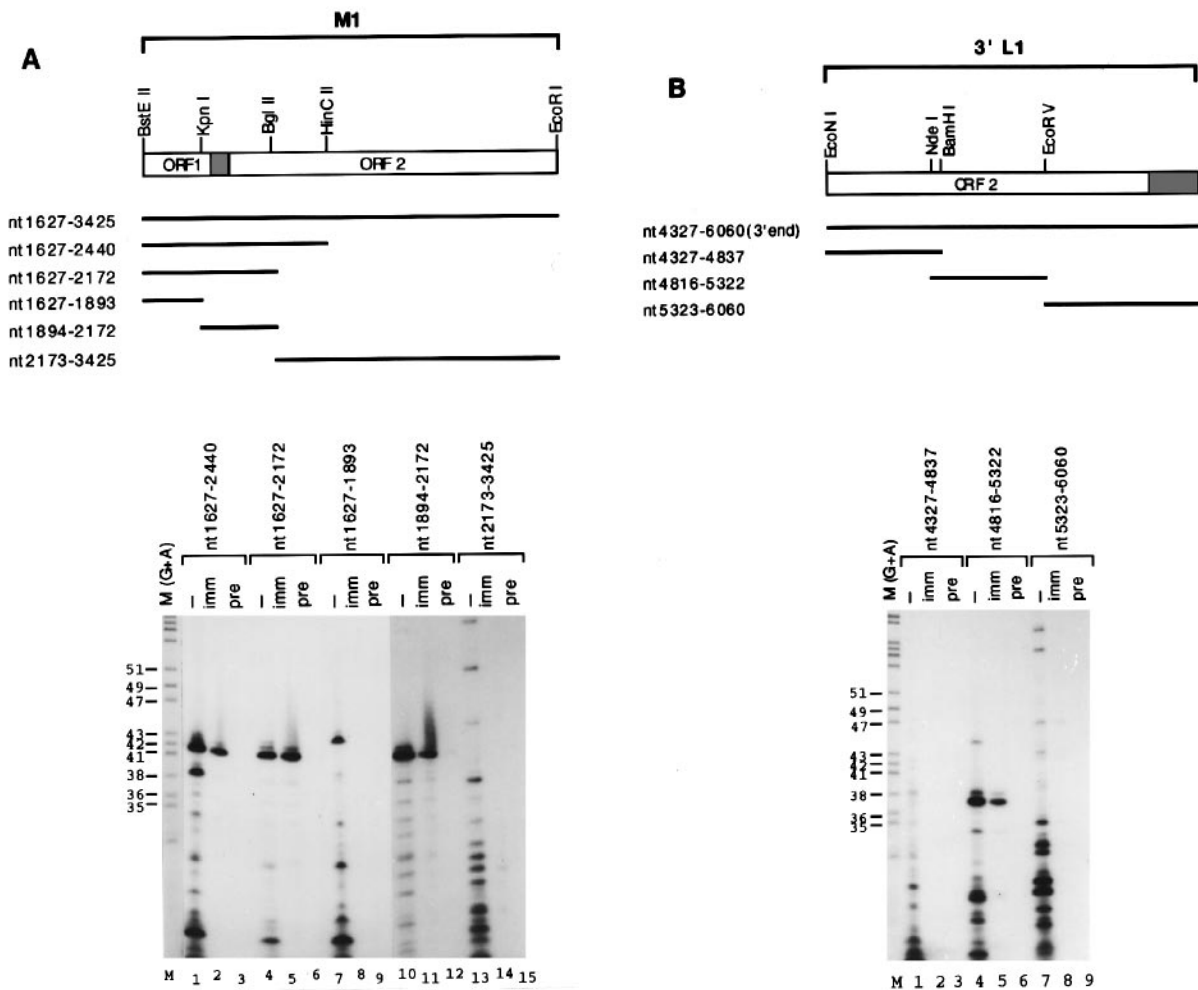

Fig. 4. *In vitro* RNA binding assay with various RNA segments of the M1 (**A**) and 3′L1 (**B**) regions as probes. The probes are indicated schematically and the nucleotide positions are indicated. The procedures are as in Figure 2.

RNA and the four segments 5′L1, M1, M2 and 3′L1 are compared (Figure 3B, lanes 1, 3, 6, 9 and 12 respectively), it is apparent that while each of the segments produced some of the bands seen in digested full-length L1.2A RNA, new bands also appeared. Many of the new bands are longer than 41 nt, suggesting that they arise from regions protected by secondary structure. Therefore, it appears that some regions in the L1.2A RNA segments can, in the absence of other regions of the molecule, form distinctive and reasonably stable secondary structures. Moreover, experiments with the subsegments of L1.2A RNA confirm the earlier suggestion that the p40 multimer does not tend to form stable associations with regions of secondary structure; little if any of the bands longer than 41 nt are co-immunoprecipitated with p40 (Figure 3B, lanes 4, 7, 10 and 13).

Competition assays

To confirm the specificity of p40 multimer binding, we performed competition assays. Four non-radioactive L1.2A RNAs were used as competitors: residues 1627–1893, 1894–2172, 4327–4837 and 4816–5322 (Figure 5A). Binding reaction mixtures were prepared with cold competitor RNAs and incubated for 5 min. Thereafter radiolabeled full-length L1.2A RNA was added as a probe and the standard binding assay procedure was carried out. As shown in Figure 5B, when either residues 1894–2172 (which contains the 41 nt segment) or 4816–5322 (which contains the 37 nt segment) were used as competitors, binding of both the A and B bands was inhibited. No inhibition was observed when either residues 1627–1893 or 4327–4837 were used as competitor. These results support the conclusion that there is a sequence-specific association of p40 multimer with L1.2A RNA and that regions 1894–2172 and 4838–5322, which contain the 41 and 37 nt segments respectively, include the binding sites. We note that while RNA fragment 1894–2172 competes with binding of both the A and B regions at both concentrations of competitor used, RNA fragment 4816–5322 competes efficiently with B but relatively weakly with A. Because binding to the A and B regions are independent of one another (Figures 3B and 4), this result suggests that the 41 nt segment has a higher affinity than the 37 nt segment for the p40 multimer.

Binding of the 41 nt sequence to p40

All the data presented thus far indicate that the A band is the 41 nt sequence from residue 1999 to 2039. This oligonucleotide was therefore used for additional studies on specificity of the interaction. Binding assays were carried out with radiolabeled sense and antisense single-stranded DNAs (ssDNA) and RNAs. When the 41 nt

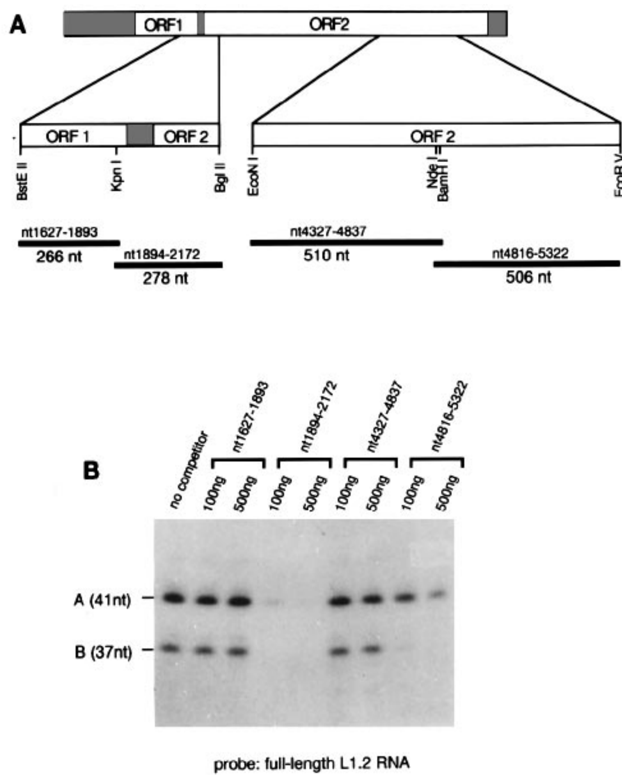

Fig. 5. Competition assay. (**A**) Schematic drawing of L1.2A RNA fragments used as competitors. Solid bars show RNA fragments with the nucleotide positions and sizes indicated above and below the bars respectively. These unlabeled competitor RNAs were synthesized *in vitro*. (**B**) Indicated amounts of several competitor RNAs were added to the binding reaction mixture containing the p40 preparation. After 5 min incubation at room temperature labeled full-length L1.2A RNA (100 ng) was added as probe and the standard RNA binding assay as in Figure 2 was followed. The intense bands A and B co-immunoprecipitated with p40 are indicated by bars.

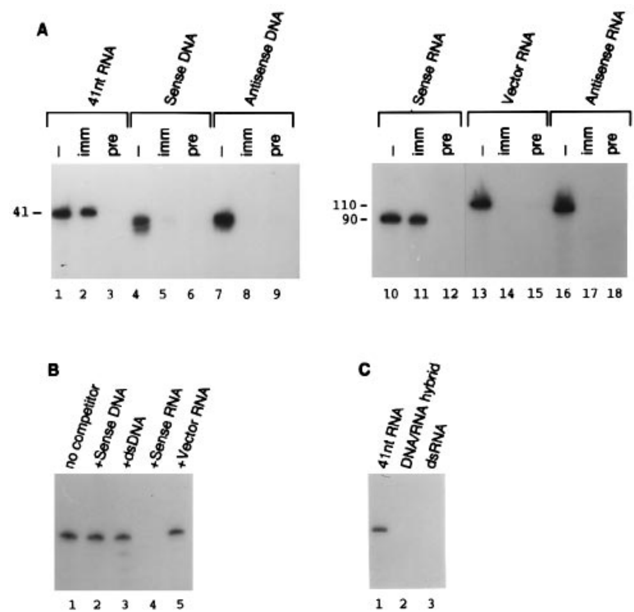

Fig. 6. Binding and competition assay with various kinds of polynucleotides as probes and competitors. (**A**) *In vitro* binding assay with sense and antisense ssDNAs and RNAs as probes. The 41 nt RNase T1-resistant fragment (41nt RNA) was purified by polyacrylamide gel electrophoresis under denaturing conditions after RNase T1 digestion of the [α-^{32}P]GTP-labeled 1894–2172 L1.2A RNA segment. DNA oligomers S42 and AS42 were used as the sense and antisense DNAs containing the 41 nt and its complementary sequence respectively. Unlabeled sense and antisense RNAs were synthesized *in vitro*, using as templates pF41 plasmid DNAs digested with *Eco*RI and *Xba*I respectively. The resultant sense and antisense RNAs contain 52 and 62 nt vector sequences at their 5′-ends respectively. The vector RNA was synthesized *in vitro* in the same direction as the sense RNA with Bluescript SK(–) vector DNA digested by *Asp*718 as template; the transcript is 121 nt long and shares the first 52 nt sequence with the sense RNA. The sense and antisense DNAs and RNAs and the vector RNA were 5′-end-labeled with [γ-^{32}P]ATP and purified as described in Materials and methods. Binding reactions were carried out with the indicated probes (0.3 pmol/300 μl reaction) as described in Figure 2 except for omission of RNase T1 digestion. Samples were examined in 20% polyacrylamide denaturing gels. Size markers (nt) are indicated by bars. (**B**) Competition assay. The *Xba*I–*Asp*718 fragment DNA (97 bp) of pF41 was isolated and used as a dsDNA competitor. Indicated unlabeled polynucleotides (10 pmol each) were added to the standard binding reaction mixture (100 μl). After 5 min incubation at room temperature, radioactive 41 nt RNA (0.1 pmol) purified as in (A) was added and the procedure described in (A) followed. (**C**) Binding assay with dsRNA and DNA–RNA hybrid as probes. The radioactive 41 nt RNA (0.1 pmol) purified as in (A) was mixed with 5 pmol of either antisense DNA (lane 2) or RNA (lane 3), heat denatured at 65°C for 3 min, annealed by cooling to room temperature over 30 min and then used as probe in the binding assay (100 μl reaction) as in (B). Lane 1 (41 nt RNA) is the sample obtained in the absence of antisense polynucleotides.

RNA purified from an RNase T1 digest was used as probe it was recovered after immunoprecipitation with immune IgG (Figure 6A, lane 2). This experiment confirms identification of the A band in the experiments in which the 41 nt oligomer was selected for binding from a large mixture of RNase T1-digested fragments; these results indicate that no additional flanking sequences are required for efficient binding. Binding also occurred when the 41 nt RNA was joined (at its 5′-end) to a sequence derived from the vector (Figure 6A, lane 11); thus the presence of a randomly selected flanking RNA sequence does not interfere with binding. In contrast, neither the sense (Figure 6A, lane 5) nor antisense (Figure 6A, lane 8) ssDNA, the antisense RNA (Figure 6A, lane 17) nor an RNA representing vector sequence alone (Figure 6A, lane 14) was co-immunoprecipitated by immune IgG. Thus binding to the p40 multimer appears to be specific for RNA as well as for specific sequences.

To confirm this conclusion, we performed binding assays using the 41 nt RNA segment as radiolabeled probe and either sense ssDNA, sense RNA, vector RNA or double-stranded DNA (dsDNA) containing the 41 nt sequence and its complement as competitors (Figure 6B). As expected, the 41 nt sense RNA segment joined to the vector sequence competed with the probe (Figure 6B, lane 4). However, no competition was observed with either the sense ssDNA (Figure 6B, lane 2), dsDNA (Figure 6B, lane 3) or vector RNA (Figure 6B, lane 5).

We next examined whether the p40 multimer can bind to a DNA–RNA hybrid or dsRNA containing the 41 nt RNA sequence. The radioactive 41 nt RNA was annealed to either antisense ssDNA or RNA and the resulting duplexes were used as probes in the binding assay. No 41 nt RNA was co-immunoprecipitated by immune IgG in these experiments (Figure 6C, lanes 2 and 3), indicating that the p40 multimer does not associate with the 41 nt RNA segment if it is in a duplex with either DNA or RNA.

Fig. 7. Analysis of the protein associated with the 41 nt RNA. The 41 nt RNA labeled with [α-^{32}P]UTP was purified as in Figure 6A and used (0.3 pmol) in the standard binding reaction (300 µl). After 15 min incubation at room temperature the reaction was irradiated with UV as described previously (Hohjoh and Singer, 1996) and divided into three 100 µl portions. Two portions were subjected to immunoprecipitation with pre-immune (Pre) or anti-p40 (Imm) IgGs. Proteins associated with RNA were collected by ethanol precipitation (Total) from the remainder. The collected samples were treated with RNase A (0.5 µg/µl) at 37°C for 30 min and analyzed by SDS–PAGE followed by autoradiography. Sizes of marker proteins (kDa) in lane M are indicated by bars.

```
B

41nt fragment, nt1999-2039   AUCAAAUUCACACAUA--ACAAUAUUAACUUUAAAUAUAAAUG
                                *    * **** *   **** * * * ** **
37nt fragment, nt4839-4875  AUCCCUUCCUUACACCUUAUACAAAAAUCAAUUCAAG

             nt1999-2008  AUCAAAUUCA
                          ***** ****
             nt4865-4873  AUCAA-UUCA
```

Fig. 8. Sequences of the 41 and 37 nt RNase T1-resistant fragments. (**A**) Schematic drawing of the full-length L1Hs element showing the positions of the 41 and 37 nt fragments. Endonuclease, reverse transcriptase and a zinc finger-like domain in ORF2 are indicated by EN, RT and ZN respectively. UTR, untranslated region; An, poly(A) tail. (**B**) The 41 and 37 nt RNA sequences. Gaps are indicated by –. Identical nucleotides between the 41 and 37 nt fragments are indicated by asterisks. The nucleotide positions of these segments in L1.2A are indicated.

Direct association of p40 with L1Hs RNA

Earlier experiments indicated that p40 is directly associated with L1Hs RNA in the p40 RNP complex found in human teratocarcinoma cells (Hohjoh and Singer, 1996). We carried out experiments to determine if this was also the case for the complex formed *in vitro* between p40 multimer and RNA. A standard binding reaction mixture was prepared with the 41 nt RNA labeled with [α-^{32}P]UTP and irradiated with UV. After irradiation the reaction was divided into three equal portions. Ethanol was added to one portion to precipitate all proteins and associated RNA; the other two portions were subjected to immunoprecipitation with pre-immune and immune IgGs respectively. All the precipitates were collected and treated with RNase A and those proteins that were cross-linked to the labeled RNA were examined by SDS–PAGE followed by autoradiography (Figure 7). In the sample precipitated with ethanol two bands with mobilities corresponding to ~80 and 39 kDa respectively were observed against a smeared background (Figure 7, Total). The 39 kDa band was the only one observed in the sample immunoprecipitated with anti-p40 IgG (Figure 7, Imm) and no band was seen in the sample immunoprecipitated with pre-immune IgG (Figure 7, Pre). Monomeric p40 is expected to migrate as an ~38–39 kDa protein in SDS gels (Leibold *et al.*, 1990; Holmes *et al.*, 1992). Thus these results suggest that p40 in the multimer associates directly with the 41 nt RNA.

Discussion

Binding sites for p40 in L1Hs RNA

Three forms of p40 have been identified. On SDS–PAGE gels under reducing conditions the polypeptide extracted from human teratocarcinoma cells migrates as a monomer with a mobility somewhat faster than expected (Leibold *et al.*, 1990; Holmes *et al.*, 1992). When cytoplasmic extracts are not subjected to denaturing or reducing conditions or treatment with high salt or ribonuclease, p40 is found in a large RNP complex in association with L1Hs RNA (Hohjoh and Singer, 1996). Treatment of the p40 RNP complex with several kinds of RNases or high salt yields p40 multimers in the ~200 kDa size range (Hohjoh and Singer, 1997). High salt treatment also releases RNA from the RNP complex.

This paper reports experiments designed to study the association of p40 multimers with L1Hs RNA. For this purpose we developed an *in vitro* RNA binding assay system that depends on co-immunoprecipitation of p40 and bound RNA and digestion of the RNA with RNase T1 to identify the bound regions. The results indicate that there are two high affinity binding sites for p40 in full-length L1Hs RNA synthesized from an active L1Hs element, L1.2A (Dombroski *et al.*, 1991; Moran *et al.*, 1996).

Initial identification of the location of the binding sites was based on the observation that two RNase T1-resistant fragments of the 6000 nt L1Hs RNA are co-immunoprecipitated with p40. These fragments, A and B, are 41 and 37 nt long and derive from ORF2 region residues 1999–2039 and 4839–4875 respectively (Figure 8A). The experiments reported here also indicate that the isolated 41 nt RNA is efficiently bound by p40. Thus no additional sequences flanking this region are required for efficient binding. Moreover, both the 41 and 37 nt segments were specifically bound even when presented to p40 in a mixture of all the RNase T1-resistant fragments produced from L1.2A RNA. Our experiments also demonstrate that each of the segments binds efficiently in the absence of the other.

The sequences of fragments A and B are, in part, homologous (Figure 8B), but until the detailed binding sites are determined, the significance of the homologies will be unknown. Preliminary experiments (H.Hohjoh) suggest that the binding site is within the 3′-terminal two

A

```
                       *     * **** * **** * * * ** **
                augacaggAUCAAAUUCACACAUAACAAUAUUAACUUUAAAUAUAAAUG  L1Hs (nt1991-2039)
augccaacucuaacaacaaaa..a.a...GAGC..CAAUU..U.UU.CU..A..U..C.U.....C....  L1Md (nt2855-2924)
           augaau..a..g...AGC...AAU...UAUU.CU..A..U..C.C..C..C....  L1Rn (nt2519-2573)
                ...g....GCA..G.CACU..G..U....UG.C..A..G..C..U....  L1Oc (nt1235-1283)

                    M  T  G  S  N  S  H  I  T  I  L  T  L  N  I  N   L1Hs
M  P  T  L  T  T  K I  K  •  •  •  N  Y  F  •  •  •  •  •  •  •  •   L1Md
                 M  N I  •  •  •  •  N  •  Y  •  •  •  •  •  •  •  •   L1Rn
                    •  A  •  Q  S  H  Y  V  •  •  V  •  •  •  •  •   L1Oc
```

B

```
   *    * **** *   **** * * * ** **
AUCCCUUCCUUACACCUUAUACAAAAAUCAAUUCAAG  L1Hs (nt4839-4875)
....A.A.U.AU.U....G...U..GG....A..U.A  L1Md (nt5724-5760)
....A.C.U.AU....C.G......GC.U..G..C.A  L1Rn (nt5370-5406)
.C....A.......U..C.C.........C.C....C  L1Oc (nt4086-4122)

  P  F  L  T  P  Y  T  K  I  N  S    L1Hs
  •  •  •  •  •  C  •  •  V  K  •    L1Md
  •  S  •  •  •  C  •  •  •  K  •    L1Rn
  •  •  •  •  S  H  •  •  •  H  •    L1Oc
```

Fig. 9. Nucleotide and amino acid sequences in the regions corresponding to the 41 (**A**) and 37 nt (**B**) segments in mammalian L1 elements. Sequence data are derived as follows; human L1 (L1Hs), Dombroski *et al.* (1991); mouse L1 (L1Md), Loeb *et al.* (1986); rat L1 (L1Rn), D'Ambrosio *et al.* (1986); rabbit L1 (L1Oc); Demers *et al.* (1989). Alignment of the sequences is based on that described by Demers *et al.* (1989). Nucleotide positions of the fragments in their elements are indicated in parentheses. The sequences in (A) begin with the first methionine codon in L1Hs ORF2 and the nucleotide sequences of the 41 nt segment are indicated by capital letters. Only nucleotides and amino acids that are different from the sequences of L1Hs are indicated. Small and large dots in nucleotide and amino acid sequences respectively indicate identical residues. Underlined dots indicate similar amino acids (F = Y, I = L and S = T) to those of L1Hs. Homologous nucleotides between the 41 and 37 nt segments in L1Hs RNA are indicated by asterisks.

thirds of fragment A. We note that the next longest four RNase T1-resistant products predicted from the L1.2A sequence do not appear in the immunoprecipitates (even after UV cross-linking), although they share some short sequence similarities with the A and B fragments. The few RNA fragments other than A and B which are occasionally observed as minor bands associated with the immunoprecipitates may represent additional, lower affinity p40 binding sites. Altogether, the results indicate a specific interaction between the p40 multimers and sequences within the A and B segments. Some segments within fragment A (e.g. AUAAA and UUUA) are similar to protein binding sites in (pre-)mRNAs (McCarthy and Kollmus, 1995).

Since the A segment begins within a few codons of the first methionine of ORF2 (Figures 8A and 9), it is possible that the interaction between L1Hs RNA and p40 may influence translation of ORF2. However, previous *in vitro* translation experiments showed no such effect (McMillan and Singer, 1993).

Evidence that p40 is a sequence-specific, single-strand RNA binding protein

RNase T1 digests of full-length L1.2A RNA suggest that the regions containing the A and B segments are not protected from digestion by secondary structure and computer analysis indicates no significant secondary structure exists within the segments themselves. Moreover, our experiments demonstrate that p40 multimers do not bind efficiently to the 41 nt sequence corresponding to fragment A if it is presented in a duplex with either DNA or RNA or as single- or double-stranded DNA. These data suggest that p40 (in the multimer) is a sequence-specific, single-strand RNA binding protein.

This and previous work demonstrate that p40 is directly associated with L1Hs RNA both *in vitro* (Figure 7) and *in vivo* (Hohjoh and Singer, 1996). We attempted to investigate which region(s) of p40 is important for RNA binding using the *in vitro* RNA binding assay and p40 synthesized in bacteria with L1.2A template (Hohjoh and Singer, 1996). However, no binding of L1.2A RNA was observed, although bacterial p40 forms multimers and is immunoprecipitated by our antibody preparations (data not shown). p40 in teratocarcinoma cells is known to be phosphorylated (Hohjoh and Singer, 1996) and may also be post-translationally modified in other ways; it is possible that such differences between the p40 isolated from teratocarcinoma cells and that synthesized in *Escherichia coli* account for the different results. It is also possible that other cellular factors present in the p40 preparation from teratocarcinoma cells might be required for binding.

As already reported (Holmes *et al.*, 1992; Hohjoh and Singer, 1996), p40 has no sequence homology to known proteins and cDNAs except for the C-terminal halves of the ORF1 polypeptides predicted to be encoded by L1s in other mammalian species. L1Hs p40 has a leucine zipper motif at amino acid residues 90–131 (Holmes *et al.*, 1992) and the potential to form α-helical coiled coils throughout the molecule; these contribute to p40–p40 interactions and formation of multimeric complexes *in vitro* (Hohjoh and Singer, 1996). We have detected no amino acid sequence similarity between p40 and reported RNA binding motifs (Kenan *et al.*, 1991; Burd and Dreyfuss, 1994; Nagai *et al.*, 1995). The leucine zipper motif in the N-terminal half of p40 is not preceded by the basic region typical of DNA binding transcription factors that contain coiled leucine zippers (Lamb and McKnight, 1991). Thus p40 appears to be a novel RNA binding protein.

The role of LINE-1 ORF1 proteins

Because it contains L1Hs RNA, it seems likely that the large p40 RNP complex participates in retrotransposition of L1Hs sequences. One possible model for formation of the large p40 RNP complex which is consistent with current information is as follows. p40 monomers themselves (or together with other proteins) are assembled into ~200 kDa multimers by interactions such as those typical of coiled coil structures involving leucine zippers and α-helices. The p40 multimers associate with specific

binding sites on L1Hs RNA and then combine, perhaps with additional p40 or other protein molecules, to form the large p40 RNP complex.

It is of interest to consider whether the other mammalian L1s form similar RNP complexes. No leucine zipper motif has been observed in the polypeptides predicted by the ORF1s of other mammalian L1s, but the amino acid sequence of all of these proteins is consistent with extensive α-helical regions (Demers *et al.*, 1989; Hohjoh and Singer, 1996). The protein encoded by ORF1 of *Mus domesticus* L1 (L1Md) is also found in an RNP complex in association with L1Md RNA in mouse embryonal carcinoma cell line F9 (Martin, 1991; Martin and Branciforte, 1993). This suggests that such complexes may be typical of mammalian L1s. Possibly the conserved C-terminal regions of mammalian ORF1 polypeptides, which include a high proportion of basic amino acids, are involved in binding to RNA. Moreover, mutations known to suppress L1Hs transposition occur in the conserved region of p40 (Moran *et al.*, 1996).

We compared the A and B segments of L1Hs with the corresponding regions of ORF2 in other mammalian L1 elements (Figure 9). Although there are some similarities in nucleotide sequence, it is not particularly striking. Comparing L1Hs with L1Md, for example, the nucleotide sequences are only ~50% identical. In contrast, the amino acid sequences, counting both identical and similar amino acids, are 79% conserved. Thus conservation appears to reflect the importance of the protein structure, not the nucleotide sequence. It may be that the conserved nucleotide residues reflect specific binding sites for all these ORF1 proteins. However, it is also possible that specific binding sites on L1 RNAs of other mammals, if they exist, are unrelated to the L1Hs RNA binding sites for p40.

The ORF1 proteins of mammalian L1s have no homology to gag and gag-like proteins and thus the RNP complexes they form are expected to be different from the RNP complexes formed by retroviruses, LTR retrotransposons and non-LTR retrotransposons found in invertebrates and plants. The data reported here, as well as previous observations, confirm this expectation. Thus intact ORF1 protein interacts directly with L1Hs RNA and is not first cleaved to smaller polypeptides. The RNA in the large RNP complex is accessible even to rather large ribonuclease molecules, which is not the case for virus and virus-like particles (Hohjoh and Singer, 1997).

Finally, we point out that the *in vitro* RNA binding assay method described in this paper may be generally useful to identify binding sites for particular proteins within long RNA chains. Immunoprecipitation provides specificity for the protein moiety and the use of RNases of known cleavage specificity aids in identification of binding sites on RNAs of known sequence. This offers advantages over conventional binding assays such as gel retardation and filter binding when using relatively long RNA sequences.

Materials and methods

Cell culture

2102Ep and HeLa cells were grown as previously described (Swergold, 1990).

Preparation of p40 multimer

Cells (1–1.4×10^8) were harvested and disrupted by Dounce homogenization as previously described (Hohjoh and Singer, 1996). The cell extract was subjected to sequential centrifugations at 12 000 *g* for 10 min and 200 000 *g* for 2 h (SW-60 rotor; Beckman) at 4°C. The pellet after centrifugation at 200 000 *g* was suspended in a high salt buffer (3 ml) containing 10 mM Tris–HCl, pH 8.0, 1 mM EDTA, 10 mM $MgCl_2$ and 0.5 M NaCl, placed on ice for 1 h and centrifuged at 200 000 *g* for 2 h at 4°C. The supernatant was diluted with three times its volume of buffer A (20 mM Tris–HCl, pH 8.0, 1 mM EDTA, 10 mM $MgCl_2$, 0.2% Nonidet P-40 and 20% glycerol) and applied to a HiTrap heparin column (Pharmacia) equilibrated with buffer A. Protein elution was carried out by increasing concentrations of NaCl (300 mM–1 M by step gradient) in buffer A (0.5 ml/fraction). Samples of the fractions were analyzed by SDS–PAGE and Western blotting using anti-p40 IgG (Figure 1B). The fractions containing p40 (650–800 mM NaCl) were pooled and dialyzed against buffer containing 20 mM Tris–HCl, pH 8.0, 0.2 mM EDTA, 140 mM KCl, 10 mM NaCl, 10 mM $MgCl_2$, 0.5 mM DTT and 20% glycerol.

Purification of IgG

IgGs were purified from AH40.1 rabbit immune serum against p40 (Leibold *et al.*, 1990) and pre-immune serum with protein A–agarose (Gibco BRL) as described previously (Hohjoh and Singer, 1996). The eluted IgG was dialyzed against phosphate-buffered saline (PBS) at 4°C.

Plasmids

pL1.2A was used as template for full-length L1Hs RNA synthesis *in vitro* (Dombroski *et al.*, 1991). Segments of pL1.2A were eliminated to provide templates for synthesis of less than full-length L1Hs RNA. To synthesize human β-actin RNA *in vitro*, plasmid pHCDβA-1 was constructed by inserting the *Pst*I–*Xho*I DNA fragment of pHFβA-1 (Gunning *et al.*, 1983) into the same enzyme sites of Bluescript SK(–) vector (Stratagene). Plasmid pF41, containing the L1Hs 41 nt RNase T1-resistant sequence, was constructed by inserting annealed DNA oligomers S42 and AS42EC into the *Bam*HI and *Eco*RI sites of Bluescript vector SK(–).

DNA oligomers

DNA oligomers synthesized for this study were: S42, 5′-GAT CAA ATT CAC ACA TAA CAA TAT TAA CTT TAA ATA TAA ATG-3′; AS42, 5′-CAT TTA TAT TTA AAG TTA ATA TTG TTA TGT GTG AAT TTG ATC-3′; AS42EC, 5′-AAT TCA TTT ATA TTT AAA GTT AAT ATT GTT ATG TGT GAA TTT-3′. The oligomers were obtained from Gibco BRL.

In vitro transcription and RNA labeling

Radiolabeled RNA was prepared *in vitro* by transcription of cloned L1.2A and subsegments of L1.2A in 30 μl reactions containing 40 mM Tris–HCl, pH 8.0, 8 mM $MgCl_2$, 50 mM NaCl, 2 mM spermidine, 5 mM DTT, 670 μM ATP, CTP and GTP, 200 μM UTP, 50 μCi [α-^{32}P]UTP (3000 Ci/mmol) {or 670 μM ATP, CTP and UTP, 200 μM GTP, 50 μCi [α-^{32}P]GTP (3000 Ci/mmol)}, ~1 μg template DNA, 1 U/μl RNase inhibitor and 20 U T7 or T3 RNA polymerase. After transcription the reaction was passed through a G-50 spin column (5Prime-3Prime) and RNA was collected by ethanol precipitation and dissolved in H_2O (40 μl). A small aliquot of the RNA was examined by agarose gel electrophoresis after denaturation with glyoxal and dimethyl sulfoxide (Maniatis *et al.*, 1982). For 5′-end-labeling synthesized DNA oligomers and RNAs treated with alkaline phosphatase (Boehringer Mannheim) were incubated with [γ-^{32}P]ATP and T4 DNA kinase (Gibco BRL). The 5′-end-labeled DNAs and RNAs and the 41 nt RNase T1-resistant fragment generated by digestion of labeled L1.2A RNA with RNase T1 were purified by polyacrylamide gel electrophoresis under denaturing conditions. The RNAs were eluted from gels in elution buffer (0.5 M ammonium acetate, 1 mM EDTA and 0.2% SDS), collected by ethanol precipitation and dissolved in H_2O.

In vitro RNA binding assay

Labeled RNAs (100 ng) were incubated with p40 preparation (~10 μg protein) in 300 μl binding buffer [20mM Tris–HCl, pH 7.5, 1 mM EDTA, 140 mM KCl, 10 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 100 μg/ml yeast total RNA (Sigma)] at room temperature for 15 min and digested with RNase T1 (300 U; Boehringer Mannheim) at room temperature for 30 min. Forty microliters of the reaction were removed and the rest was divided into two equal portions. From the 40 μl aliquot total digested RNA was isolated as a control. The equally divided

portions were subjected to immunoprecipitation with either pre-immune or anti-p40 IgGs. The purified IgGs (10 μg) were added to the samples and after 30 min incubation at room temperature 50 μl protein A–agarose (Gibco BRL) were added and incubation continued for 30 min at room temperature with gentle shaking. The agarose beads were washed four times with PBS containing 0.1% Tween 20 and twice with PBS and then suspended in 100 μl PBS containing 0.5 mg/ml proteinase K and 0.5% SDS. After incubation at 37°C for 30 min the aqueous solution was extracted with phenol and chloroform and RNAs in the solution were collected by ethanol precipitation with 10 μg yeast total RNA as carrier. The collected RNAs were separated on 8% sequencing gels and exposed to X-ray film.

UV cross-linking, SDS–PAGE and Western blotting

UV cross-linking, SDS–PAGE and Western blotting were carried out as previously described (Hohjoh and Singer, 1996).

Acknowledgements

We would like to thank H.H.Kazazian Jr for providing plasmid pL1.2A and B.Paterson for providing plasmid pHFβA-1. We also thank A.Clements, C.Klee, D.Hursh, G.D.Swergold and T.G.Fanning for their advice and helpful discussion.

References

Bratthauer,G.L. and Fanning,T.G. (1992) Active LINE-1 retrotransposons in human testicular cancer. *Oncogene*, **7**, 507–510.

Burd,C.G. and Dreyfuss,G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.

D'Ambrosio,E., Waitzkin,S.D., Witney,F.R., Salemme,A. and Furano, A.V. (1986) Structure of the highly repeated, long interspersed 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.

Dombroski,B.A., Mathias,S.L., Nanthakumar,E., Scott,A.F. and Kazazian,H.H.,Jr (1991) Isolation of an active human transposable element. *Science*, **254**, 1805–1808.

Fanning,T.G. and Singer,M.F. (1987) LINE-1: a mammalian transposable element. *Biochim. Biophys. Acta*, **910**, 203–212.

Feng,Q., Moran,J.V., Kazazian,H.H.,Jr and Boeke,J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, **87**, 905–916.

Gunning,P., Ponte,P., Okayama,H., Engel,J., Blau,H. and Kedes,L. (1983) Isolation and characterization of full-length cDNA clones for human α-, β-, and γ-actin mRNAs: skeletal but not cytoplasmic actins have an amino-terminal cysteine that is subsequently removed. *Mol. Cell. Biol*., **3**, 787–795.

Hohjoh,H. and Singer,M.F. (1996) Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J*., **15**, 630–639.

Hohjoh,H. and Singer,M.F. (1997) Ribonuclease and high salt sensitivity of the ribonucleoprotein complex formed by the human LINE-1 retrotransposon. *J. Mol. Biol*., **271**, 7–12.

Holmes,S.E., Singer,M.F. and Swergold,G.D. (1992) Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. *J. Biol. Chem*., **267**, 19765–19768.

Hutchison,C.A., Hardies,S.C., Loeb,D.D., Shehee,W.R. and Edgell,M.H. (1989) LINEs and related retrotransposons: 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.

Hwu,H.R., Roberts,J.W., Davidson,E.H. and Britten,R.J. (1986) Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. *Proc. Natl Acad. Sci. USA*, **83**, 3875–3879.

Kenan,D.J., Query,C.C. and Keene,J.D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci*., **16**, 214–220.

Lamb,P. and McKnight,S.L. (1991) Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerization. *Trends Biochem. Sci*., **16**, 417–422.

Leibold,D.M., Swergold,G.D., Singer,M.F., Thayer,R.E., Dombroski, B.A. and Fanning,T.G. (1990) Translation of LINE-1 DNA elements *in vitro* and in human cells. *Proc. Natl Acad. Sci. USA*, **87**, 6990–6994.

Loeb,D.D., Padgett,R.W., Hardies,S.C., Shehee,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.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 200–201.

Martin,S.L. (1991) Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol. Cell. Biol*., **11**, 4804–4807.

Martin,S.L. and Branciforte,D. (1993) Synchronous expression of LINE-1 RNA and protein in mouse embryonal carcinoma cells. *Mol. Cell. Biol*., **13**, 5383–5392.

Mathias,S.L., Scott,A.F., Kazazian,H.H.,Jr, Boeke,J.D. and Gabriel,A. (1991) Reverse transcriptase encoded by a human transposable element. *Science*, **254**, 1808–1810.

McCarthy,J.E.G. and Kollmus,H. (1995) Cytoplasmic mRNA–protein interactions in eukaryotic gene expression. *Trends Biochem. Sci*., **20**, 191–197.

McMillan,J.P. and Singer,M.F. (1993) Translation of the human LINE-1 element, L1Hs. *Proc. Natl Acad. Sci. USA*, **90**, 11533–11537.

Moran,J.V., Holmes,S.E., Naas,T.P., DeBerardinis,R.J., Boeke,J.D. and Kazazian,H.H.,Jr (1996) High frequency retrotransposition in cultured mammalian cells. *Cell*, **87**, 917–927.

Nagai,K., Oubridge,C., Ito,N., Avis,J. and Evans,P. (1995) The RNP domain: a sequence-specific RNA-binding domain involved in processing and transport of RNA. *Trends Biochem. Sci*., **20**, 235–240.

Sassaman.D.M., Dombroski,B.S., Moran,J.V., Kimberland,M.L., Nass,T.P., DeBerardinis,R.J., Gabriel,A., Swergold,G.D. and Kazazian,H.H.,Jr (1997) Many human L1 elements are capable of retrotransposition. *Nature Genet*., **16**, 37–43.

Singer,M.F., Krek,V., McMillan,J.P., Swergold,G.D. and Thayer,R.E. (1993) *LINE-1*: a human transposable element. *Gene*, **135**, 183–188.

Smit,A.F.A. (1996) The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev*., **6**, 743–748.

Swergold,G.D. (1990) Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol*., **10**, 6718–6729.

Received on May 27, 1997