BJ www.biochemj.org

BJ Gene

Biochem. J. (2009) **424**, 479–490 (Printed in Great Britain) doi:10.1042/BJ20090766

Nucleic-acid-binding properties of the C2-L1Tc nucleic acid chaperone encoded by L1Tc retrotransposon

Sara R. HERAS*, M. Carmen THOMAS*, Francisco MACIAS*, Manuel E. PATARROYO†‡, Carlos ALONSO§ and Manuel C. LÓPEZ*[1]

*Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina 'López Neyra', CSIC, 18001 Granada, Spain, †Fundación Instituto de Inmunología de Colombia, FIDIC, Bogotá, Colombia, ‡Universidad Nacional de Colombia, Bogotá, Colombia, and §Centro de Biología Molecular 'Severo Ochoa', CSIC, Madrid, Spain

It has been reported previously that the C2-L1Tc protein located in the *Trypanosoma cruzi* LINE (long interspersed nuclear element) L1Tc 3′ terminal end has NAC (nucleic acid chaperone) activity, an essential activity for retrotransposition of LINE-1. The C2-L1Tc protein contains two cysteine motifs of a C2H2 type, similar to those present in TFIIIA (transcription factor IIIA). The cysteine motifs are flanked by positively charged amino acid regions. The results of the present study show that the C2-L1Tc recombinant protein has at least a 16-fold higher affinity for single-stranded than for double-stranded nucleic acids, and that it exhibits a clear preference for RNA binding over DNA. The C2-L1Tc binding profile (to RNA and DNA) corresponds to a non-co-operative-binding model. The zinc fingers present in C2-L1Tc have a different binding affinity to nucleic acid molecules and also different NAC activity. The RRR and RRRKEK [NLS (nuclear localization sequence)] sequences, as well as the C2H2 zinc finger located immediately downstream of these basic stretches are the main motifs responsible for the strong affinity of C2-L1Tc to RNA. These domains also contribute to bind single- and double-stranded DNA and have a duplex-stabilizing effect. However, the peptide containing the zinc finger situated towards the C-terminal end of C2-L1Tc protein has a slight destabilization effect on a mismatched DNA duplex and shows a strong preference for single-stranded nucleic acids, such as C2-L1Tc. These results provide further insight into the essential properties of the C2-L1Tc protein as a NAC.

Key words: double-stranded nucleic acid, long interspersed nuclear element (LINE), nucleic acid binding, nucleic acid chaperone, retrotransposition, single-stranded nucleic acid, *Trypanosoma cruzi*, zinc finger.

INTRODUCTION

Retrotransposons are ubiquitous mobile genetic elements that transpose through an RNA intermediate. These genetic elements are present in the genome of most eukaryotes [1]. They can be classified into two different lineages based on the integration mechanism they utilize. The elements having LTRs (long terminal repeats) are similar in structure and retrotransposition mechanism to those of retroviruses [2]. The elements lacking LTRs, also called LINEs (long interspersed nuclear elements), are very diverse in structure probably due to host–mobile element co-evolution. LINEs use a transposition mechanism originally described for the insect R2Bm non-LTR element, termed TPRT (target-primed reverse transcription) [3].

Most non-LTR retrotransposons have two ORFs (open reading frames). The mechanism for their retrotransposition depends upon the enzymatic functions of the ORF1 and ORF2 encoded proteins [4]. The ORF2 has a high degree of similarity with the *pol* genes of retroviruses and encodes a protein that provides the reverse transcriptase and endonuclease activities required for TPRT [5]. The role of the ORF1-encoded protein (ORF1p) in LINE retrotransposition has been difficult to uncover since it has not been associated with the function of any known protein. It has been described that in mammalian LINE-L1 elements the ORF1 codes for a protein with RNA-binding activity [6–8] and that it facilitates rearrangements between nucleic acids behaving, thus, as a NAC (nucleic acid chaperone) [9]. The highly basic region located at the C-terminal half of the ORF1p is well conserved among all mammalian ORF1 proteins and it is involved in these activities [10]. The ORF1p protein encoded by the *Drosophila melanogaster* I factor LINE has also been shown to have nucleic-acid-binding capacity and to be endowed *in vitro* with NAC activity [11]. The ORF1p from I factor contains a zinc-finger motif (CCHC) similar to the zinc fingers present in the nucleocapside basic portion of the retroviral gag polyproteins. The motif is also present in the proteins encoded by the first ORF from most LINE-like elements [12].

The L1Tc element is the best represented autonomous non-LTR retrotransposon from the *Trypanosoma cruzi* genome, a protozoan parasite belonging to the *Trypanosomatidae* family. This parasite is the agent responsible for Chagas' disease, a parasitism that affects 16–18 million people, mainly in Central and South America (www.who.int/tdr/diseases/chagas/direction.htm). The relatively high content of retroelements in the *T. cruzi* genome has been related to the significant genomic polymorphism and high degree of plasticity that this protozoan pathogen presents [13,14]. L1Tc is actively transcribed in the three stages of the parasite life cycle [15]. Some L1Tc copies have been found to contain a single ORF encoding a 1574 amino acid protein that contains all functional domains [16]. They are considered, therefore, as active transposable elements [16,17]. L1Tc codes for the enzymatic machinery involved in its retrotransposition process including an AP (apurinic/apyrimidinic) endonuclease [18], a 3′ phosphatase, a 3′ phosphodiesterase [18,19], a reverse transcriptase [20] and an RNase H activity [21].

We have previously described that the L1Tc C-terminus encodes a protein, termed C2-L1Tc, which has NAC activity and binds to several types of nucleic acids [22]. C2-L1Tc catalyses

Abbreviations used: dsDNA, double-stranded DNA; DTT, dithiothreitol; EMSA, electrophoretic mobility-shift assay; LINE, long interspersed nuclear element; LTR, long terminal repeat; NAC, nucleic acid chaperone; NLS, nuclear localization sequence; ORF, open reading frame; RNP, ribonucleoparticle; ssDNA, single-stranded DNA; T_m, melting temperature; TFIIIA, transcription factor IIIA; TPRT, target-primed reverse transcription.

[1] To whom correspondence should be addressed (email mclopez@ipb.csic.es).

the rearrangement of nucleic acids preventing melting of perfect DNA duplexes and facilitates, moreover, the strand exchange between DNAs to form stable DNA duplexes [22]. The C2-L1Tc protein contains two cysteine motifs of the C2H2 type flanked by positively charged amino acid regions. In the context of retrotransposons this is, to our knowledge, the first description of a NAC activity mediated by a protein containing C2H2 zinc-finger motifs. It has been suggested that the two zinc fingers and the basic residues located upstream of the first zinc finger co-operate and are essential for the NAC activity [22]. The C2H2 motifs were first described in the *Xenopus laevis* TFIIIA (transcription factor IIIA). These motifs are also present in many transcription factors, as well as in other DNA-binding proteins [23]. Furthermore, the C2H2 motifs are also found in proteins encoded by other non-LTR retroelements such as R2 from arthropods, CRE/SLACS from trypanosomes, NeSL from *Caenorhabiditis elegans* and in the GENIE family from *Giardia lamblia* [5,24].

The nucleic-acid-binding properties and the different affinity that the NAC proteins have for single- and double-strand nucleic acids have been related to the mechanism of NAC activity of the L1 elements [10]; however, this mechanism has not yet been completely understood in molecular terms. In the present study we have analysed the binding of C2-L1Tc to single- and double-stranded nucleic acids and investigated the contribution of the functional domains of C2-L1Tc to duplex stabilization/destabilization. In addition, we have determined the implication of specific regions of C2-L1Tc in the nucleic-acid-binding properties of the protein and its relationship to the NAC activity.

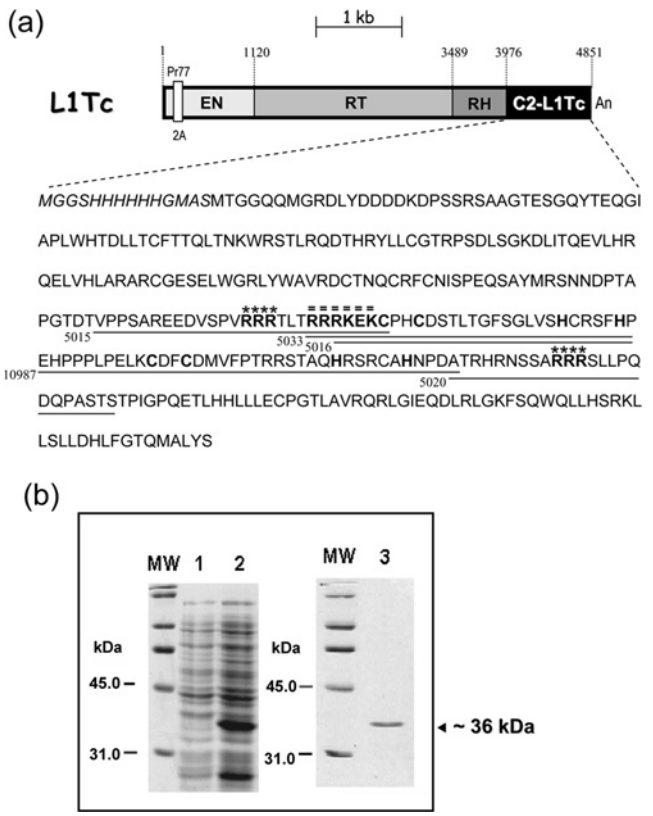

Figure 1 Structure of L1Tc and sequence of the recombinant C2-L1Tc protein (a), and purification of the recombinant protein (b)

Purification of the recombinant protein (**a**). The single ORF of the L1Tc element is represented by a rectangular box. The endonuclease (EN), reverse transcriptase (RT), RNase H (RH) and NAC (C2-L1Tc) domains [22] are separated by thin black lines indicating the nucleotide number of each one of the protein-encoding sequences. The white vertical rectangle indicates the position of the 2A self-processing sequence-mediated cleavage site [43]. The region comprising the C2-L1Tc domain is indicated as a black box. The deduced amino acid sequence is shown below. The sequence of the peptides used in the present study are underlined and numbered. The two CCHH zinc-finger domains are in bold, the RRR basic region and the NLS are in bold and labelled with (*) and (=) respectively. (**b**) *Escherichia coli* TAP-F strain (active motif) transformed with the pCAS-C2L1Tc vector was grown in SOB medium (20 g of Bacto-tryptone, 5 g of Bacto-yeast extract, 0.5 g of NaCl, double-distilled H_2O to 1 litre, pH 7.0). Total protein extracts from non-induced (lane 1) and induced (lane 2) cultures were resolved using SDS/PAGE (12 % gels). C2-L1Tc recombinant protein was purified to homogeneity by Ni^{2+}-affinity chromatography and gel filtration using a Superdex 75 column (lane 3). The molecular mass (MW) in kDa is indicated on the left-hand side of each gel.

EXPERIMENTAL

Cloning and protein purification of the recombinant C2-L1Tc

The region of the L1Tc element between positions 3976 and 4851 (GenBank® accession number AF208537) [15] was cloned into the pCAS B vector (Active motif®) as previously described [22] (Figure 1a). The C2-L1Tc protein was produced in bacteria and purified under native conditions as previously described [22]. Thus C2-L1Tc recombinant protein was recovered with more than 95 % purity (Figure 1b).

Peptide synthesis

Peptides were synthesized by the simultaneous multiple solid-phase synthetic method [25]. The peptides were assembled using the standard t-Boc SPPS (solid-phase peptide synthesis) strategy on a MBHA (*p*-methylbezhydrilamide) resin [26]. Purity was checked by HPLC. Peptide sequences are shown in Table 1. Peptides were dissolved in sterile 1 × PBS containing 30 μM zinc chloride, at a final concentration of 500 μM.

RNA and DNA synthesis

144nt-RNA and 130nt-RNA were generated using a HindIII-digested pGR77 plasmid that contains 77 bp corresponding to the internal promoter of L1Tc [22,27] and a HindIII-digested TcKMP11n clone whose sequence is not related to L1Tc respectively (GenBank® accession number AJ000077) [28]. *In vitro* transcription was carried out using, as a template, 2 μg of linearized DNA and T7 RNA polymerase as described by Barroso-delJesus et al. [29]. Then, 30 μCi of [α-^{32}P]UTP (3000 Ci/mmol) were added to the reaction to radiolabel the *in-vitro*-synthesized transcripts. Specific activity was determined using a Bioscan QC2000 counter. The RNA was eluted from denaturing polyacrylamide gels, precipitated and resuspended in diethyl pyrocarbonate-treated water. A 2100nt-RNA containing a fragment of L1Tc mRNA from nucleotides 232–1468 was generated *in vitro* using the T7 polymerase from XbaI-linearized pCMV4NL1Tc as described above. Briefly, a 1234 bp KpnI–XbaI PCR fragment from L1Tc (GenBank® accession number AF208537) [15] was cloned into the expression vector pCMV4, resulting in pCMV4NL1Tc. The RNA was transcribed from the T7 promoter. Unincorporated nucleotides were removed by gel filtration (Sephadex G-50). Single-stranded RNAs (130 nt and 2100 nt; 130nt-denatured-RNA and 2100nt-denatured-RNA) were obtained by heating for 2 min at 65 °C, followed by cooling in ice for 2 min.

2.1kb-dsDNA (where dsDNA is double-stranded DNA) and 135bp-dsDNA were obtained by PCR using a T7 oligonucleotide (5′-GTAATACGACTCACTATAGGG-3′) as the sense primer. To amplify 2.1kb-dsDNA, pCMV4NL1Tc was used as a template and a 1446 bp L1Tc oligonucleotide

Table 1 Summary of the sequences and binding-affinity properties of the C2-L1Tc protein and C2-L1Tc-derived peptides

The number and the sequence of the peptides employed are indicated. The basic stretches of the C2-L1Tc protein are in bold and the residues forming the zinc fingers are underlined. The peptides derived from peptides 5015 and 5016 containing point mutations and deletions are labelled with *. Dissociation constants, K_d values, were obtained by fitting the experimental data to the Hill equation

$$y = \frac{B_{max} \cdot x^{\alpha_H}}{K_d^{\alpha_H} + x^{\alpha_H}}$$

where K_d is the ligand concentration at which 50 % of the nucleic acid is bound and B_{max} is the maximum binding (Figures 5, 6 and 7). The Hill formalism used has been described by Henriet et al. [44]. (a) The K_d of C2-L1Tc protein and of peptide 10987 and 5020 for dsDNA were obtained from the x intercept in the Hill transformation equation (Figures 2d and 7f). The dependent variable 'y' value is 0 when $Y/(1 - Y) = 1$. (b) Increase (+) or decrease (−) in temperature relative to the T_m of the 29/mm29c DNA duplex (40 °C) in the presence of 0.1 μM C2-L1Tc protein or 1 μM of each peptide. In the presence of peptide 5015 the duplex was not melted even at 55 °C. Conc50 represents the peptide concentration required to reach the formation of 50 % of stable duplexes as a measure of NAC activity [22]. Values are means ± S.D. NO, no observed activity.

Peptide	Sequence	RNA K_d (μM)	ssDNA K_d (μM)	dsDNA K_d (μM)	ΔT_m (°C) (b)	Conc50 (μM)
C2-L1Tc		$24 \pm 5 \times 10^{-3}$ (a)	$98 \pm 20 \times 10^{-3}$	1.78 ± 0.35	0	0.07
5033	**RRRKEK**CPHCDSTLTGFSGLVSHCRSFHP	1.21 ± 0.302	0.42 ± 0.07	1.34 ± 0.08	+ 12	0.4
5015	TVPPSAREEDVSPV**RRR**TLT**RRRKEK**C	0.43 ± 0.11	0.21 ± 0.06	1.51 ± 0.09	Not melted	0.51
5016	**RKEK**CPHCDSTLTGFSGLVSHCRSFHP	1.86 ± 0.02	1.03 ± 0.08	3.12 ± 1.01	+ 7	1.5
5031*	**RKEK**SPHSDSTLTGFSGLVSHCRSFHP	3.08 ± 0.11	2.29 ± 0.07	9.13 ± 0.16	+2	2.4
10987	EHPPPLPELKCDFCDMVFPT**RR**STAQHRSRCAHNPD	6.01 ± 0.05	0.30 ± 0.02	25.68 ± 4.22	−1	2.9
5020	ATRHRNSSA**RRR**SLLPQDQPAST	19.79 ± 3.18	10.48 ± 0.45	22.83 ± 9.78	−2	NO
5030*	TVPPSAREEDVSPV . . . TLT**RRRKEK**C	NO	NO	NO	0	NO
5032*	LT. . . . . CPHCDSTLTGFSGLVSHCRSFHP	NO	NO	NO	0	NO

(5′-GCTGATGCGGCGTAGATA-3′) as the antisense primer. A 135bp-dsDNA was amplified using kmp2 (5′-TTCCTCAA-GAGTGGTGGC-3′) as the antisense primer and the TcKMP11n clone as a template [25]. The PCR products were purified by gel filtration (Sephadex G-50). A 135 nt single-stranded DNA (135nt-ssDNA) fragment was generated by PCR and enzyme digestion. In this case, the Pfu DNA polymerase and the T4 polynucleotide kinase-phosphorylated kmp2 primer and the T7 primer were used to amplify blunt-end-135bp-dsDNA. Following the PCR amplification, the phosphorylated minus-strand of the PCR product was removed by digestion with λ-exonuclease (Fermentas). After the inactivation of the enzyme by heating at 80 °C for 10 min, the plus-strand was purified by gel filtration (Sephadex G-25) and precipitated with ethanol. Both of the 135bp-dsDNA and 135nt-ssDNA products were 5′-end labelled using [γ-^{32}P]ATP and T4 polynucleotide kinase (Roche). The unincorporated isotope was removed by gel-filtration chromatography (Sephadex G-25).

EMSAs (electrophoretic mobility-shift assays)

In the dsDNA- and ssDNA-binding experiments, ^{32}P-labelled 135bp-dsDNA or ^{32}P-labelled 135nt-ssDNA (0.5 nM) and increasing amounts of C2-L1Tc protein (0.02–2.8 μM), or the indicated concentration of each peptide, were incubated in 20 μl of binding buffer [20 mM Hepes (pH 7.5), 100 mM NaCl, 2 mM $MgCl_2$, 2 mM DTT (dithiothreitol), 5 % glycerol and 100 μg/ml BSA], for 30 min at 37 °C. For the RNA-binding experiments, ^{32}P-labelled 130nt-RNA (0.72 nM) was incubated with increasing concentrations of the C2-L1Tc (0.015–0.733 μM) protein or the synthetic peptides (1–30 μM) in 16 μl of binding buffer containing 5 units of RNasin (Ambion) for 30 min at 37 °C. To compare the affinity of C2-L1Tc for 130nt-RNA and 130nt-denatured-RNA (see Figure 3), the reactions containing the native or denatured *in vitro* transcripts and the indicated amount of C2-L1Tc protein were incubated for 5 min to avoid the formation of any secondary structure in the denatured 130nt-RNA. All of these reactions were incubated in ice and stopped by addition of 8 μl of dye solution (50 % glycerol, 0.1 % Bromophenol Blue and 0.1 % Xylene Cyanol). Nucleic-acid–protein complexes were resolved by electrophoresis on 5 % native polyacrylamide gels (39:1, acrylamide/bisacrylamide) containing 1 % glycerol. The gels were dried and phosphorimaged. The images were recovered on a Storm 820 and analysed with ImageQuant 5.2 (Amersham Biosciences).

Competition assays were performed by incubation of the C2-L1Tc protein (0.67 μM) with radiolabelled 130nt-RNA (0.72 nM) and increasing amounts of the non-radioactive 130nt-denatured-RNA and non-radioactive 2100nt-denatured-RNA in binding buffer at 37 °C for 5 min to avoid the formation of any secondary structure in the competitors. In a similar way, the binding affinity of the 130bp-dsDNA and 2.1kb-dsDNA fragments was calculated by mixing increasing concentrations of these molecules with the radiolabelled 144nt-RNA transcript (0.65 nM). These reactions were also incubated in binding buffer with the C2-L1Tc protein (0.67 μM) at 37 °C for 5 min. The reaction was stopped as described above. Electrophoretical analysis of the generated products was performed as described above.

DNA-melting assays

Assays were performed as described previously [9,22]. Briefly, a preannealed mismatched duplex was made by mixing 200 mM ^{32}P-labelled 29-mer oligonucleotide with its complementary oligonucleotide containing four internal mismatches (mm29c) in water. The mixture was heated for 5 min at 95 °C. NaCl was added to a concentration of 50 mM and the mixture was slowly cooled to room temperature (22 °C). Then, 1 nM of the ^{32}P-29mer/mm29c preannealed duplex was mixed with 1 μM of each peptide in 50 μl of buffer [20 mM Hepes (pH 7.5), 50 mM NaCl, 1 mM $MgCl_2$, 1 mM DTT and 0.1 % Triton X-100]. The sample was incubated for 5 min at temperatures ranging from 25 °C to 55 °C. At each 5 °C interval, a 5 μl aliquot was removed and mixed with 5 μl of ice-cold stop mix (0.4 mg/ml tRNA, 0.2 % SDS, 15 % Ficoll, 0.2 % Bromophenol Blue and 0.2 % Xylene Cyanol). Gels and analysis were performed as described. The melting effect was monitored on native 15 % polyacrylamide gels. The dried gel was analysed using a phosphorimager system.

RESULTS

Binding properties of the C2-L1Tc protein to nucleic acids

We have previously shown that the C2-L1Tc protein encoded by L1Tc (Figures 1a and 1b), a non-LTR retrotransposon from *T. cruzi,* exhibits NAC activity and that it is able to bind to several

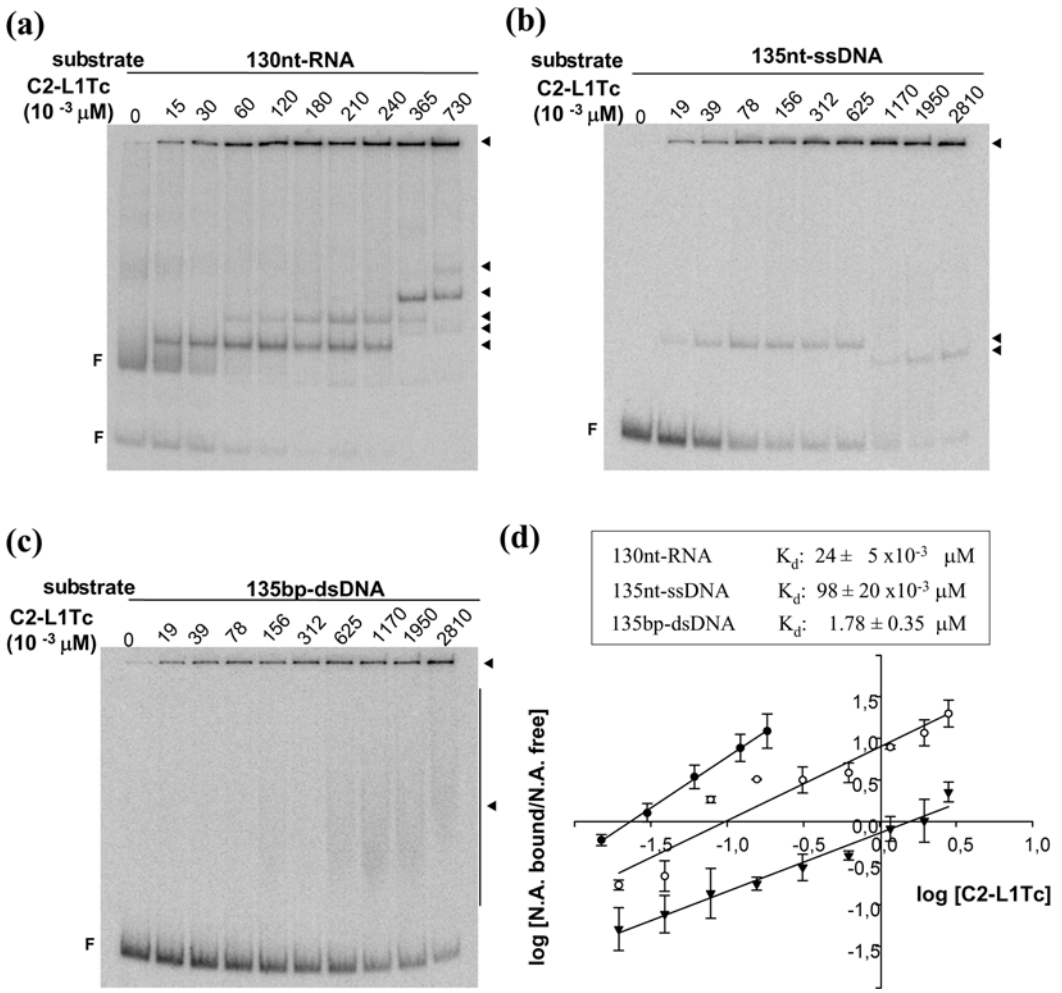

Figure 2 Nucleic-acid-binding analysis of C2-L1Tc protein

EMSA of C2-L1Tc binding to different types of nucleic acids with the same sequence composition and length. (**a**) RNA binding: 0.72 nM of ^{32}P-labelled *in vitro* transcript 130nt-RNA (0.5 ng) was incubated with an increasing concentration of C2-L1Tc (15–730 $\times$ 10^{-3} μM). (**b**) ssDNA binding: 0.5 nM ssDNA ($\sim$4 ng) (135nt-ssDNA) was incubated with increasing amounts of C2-L1Tc (19–2810 $\times$ 10^{-3} μM). (**c**) dsDNA binding: 0.5 nM dsDNA ($\sim$8 ng) (135bp-dsDNA) was incubated with increasing amounts of C2-L1Tc (19–2810 $\times$ 10^{-3} μM). (**d**) C2-L1Tc affinity for RNA, ssDNA and dsDNA. In the (**a**)–(**c**), a Hill transformation was applied to the data obtained from three independent EMSAs. The average of the three log ($Y/(1-Y)$) values were in all cases plotted against the log of C2-L1Tc concentration, where Y is the bound 130nt-RNA (●), 135nt-ssDNA (○) and the 135bp-dsDNA (▼) fraction [R^2 (coefficient of determination) = 0.99, R^2 = 0.91 and R^2 = 0.96 respectively]. K_d values were obtained from the x intercept of each equation obtained from three independent experiments since the dependent variable y value is 0 when $Y/(1-Y) = 1$ and calculated from the average of the three values in (**a**)–(**c**). The position of the 'complexes' and 'free' nucleic acids is indicated on the right-hand side (arrow heads) and on the left-hand side (F) of the panels respectively. N.A., nucleic acid.

types of nucleic acids with different affinity [22]. To determine whether the different affinity for nucleic-acid binding are due to specific features of the nucleic acids, EMSA experiments were carried out using increasing protein concentrations and several radiolabelled molecules (RNA, ssDNA and dsDNA), having the same sequence composition. As shown in Figure 2(a), and consistent with previous studies [22], when a low protein concentration of C2-L1Tc was incubated with RNA (130nt-RNA) a discrete product was formed. If the protein concentration increased, the amount of reduced-mobility products also increased (Figure 2a). When a ssDNA (135nt-ssDNA) was incubated with C2-L1Tc, a single sharp complex was detected. However, when the DNA was incubated with increasing concentrations of the protein, additional shifted bands were not formed (Figure 2b). In contrast, when dsDNA (135bp-dsDNA) was incubated with the protein, a faint smear was observed, together with a reduction of the amount of free-form dsDNA. The fraction of the protein-bound dsDNA retained on the wells increased as the concentration of protein increased (Figure 2c). These results suggest that the binding behaviour of C2-L1Tc for double- and single-stranded nucleic acids is different. The data from three independent experiments were used to generate linear-regression curves. The protein concentration at which half of each nucleic acid remained bound to the protein (K_d) was estimated to be $24 \pm 5 \times 10^{-3}$ μM for the RNA, $98 \pm 20 \times 10^{-3}$ μM for ssDNA and 1.78 ± 0.35 μM for dsDNA. Thus we may conclude that C2-L1Tc has at least a 16-fold higher affinity for ssDNA than for dsDNA, and that it exhibits a clear preference for RNA than for DNA binding.

To test whether the high affinity of the protein for RNA is influenced by the 2′-OH group (or the methyl group in thymidine) or by the specific secondary structure of the RNA used, we compared the binding capacity of C2-L1Tc to native 130nt-RNA and to the same RNA in a denatured state, 130nt-denatured-RNA (Figure 3). The slight decrease in affinity of C2-L1Tc for

the denatured RNA form ($K_d = 35.6 \pm 1.5 \times 10^{-3}$ μM for 130nt-denatured-RNA and $K_d = 24.2 \pm 0.1 \times 10^{-3}$ μM for 130nt-RNA) indicated that the RNA conformation influenced the C2-L1Tc binding capability (Figures 3a and 3b). However, the affinity of C2-L1Tc for the denatured 130nt-RNA was still higher than the observed affinity for 135nt-ssDNA ($K_d = 98 \pm 20 \times 10^{-3}$ μM) as an indication that the C2-L1Tc protein has a preference for RNA.

To determine whether the nucleic-acid length affects the capability of C2-L1Tc to bind RNA, competition experiments using competitors of different sizes were carried out (Figure 4). A constant amount of protein and a radiolabelled 130nt-RNA and, as non-radioactive competitors, a denatured 130nt-RNA (130nt-denatured-RNA) or a denatured 2100nt-RNA (2100nt-denatured-RNA) are shown in Figure 4(a). The experimental data were fitted to a four-parameter logistic curve. The EC_{50}, defined as the competitor concentration required to release half the amount of the protein bound to the radiolabelled RNA, and the Hill coefficient (α_H) which reflects co-operativity, were determined. The EC_{50} values were 0.43 ± 0.17 ng/μl for the 130nt-denatured-RNA and 0.12 ± 0.03 ng/μl for the 2100nt-denatured-RNA (Figure 4a). A similar assay was performed using a labelled 144nt-RNA and increasing concentrations of two unlabelled dsDNA molecules, 130bp-dsDNA and 2.1kb-dsDNA (Figure 4b). In this case, the EC_{50} was 24.69 ± 0.17 ng/μl and 10.34 ± 5.7 ng/μl respectively. These results revealed that, although C2-L1Tc has a lower affinity for double-stranded than for single-stranded nucleic acids, the binding affinity of C2-L1Tc to both types of nucleic acid increases as the length of the nucleic acid increases. Furthermore, the Hill coefficient obtained in all cases was close to ~ 1, corresponding to a non-co-operative-binding model. Thus the α_H values of denatured-RNA and dsDNA binding of higher length were 1.13 ± 0.24 for the 2100nt-denatured-RNA long molecule and 0.94 ± 0.39 for the 2.1kb-dsDNA long molecule, compared with 1.06 ± 0.29 and 1.14 ± 0.12 respectively, for nucleic acids of shorter length.

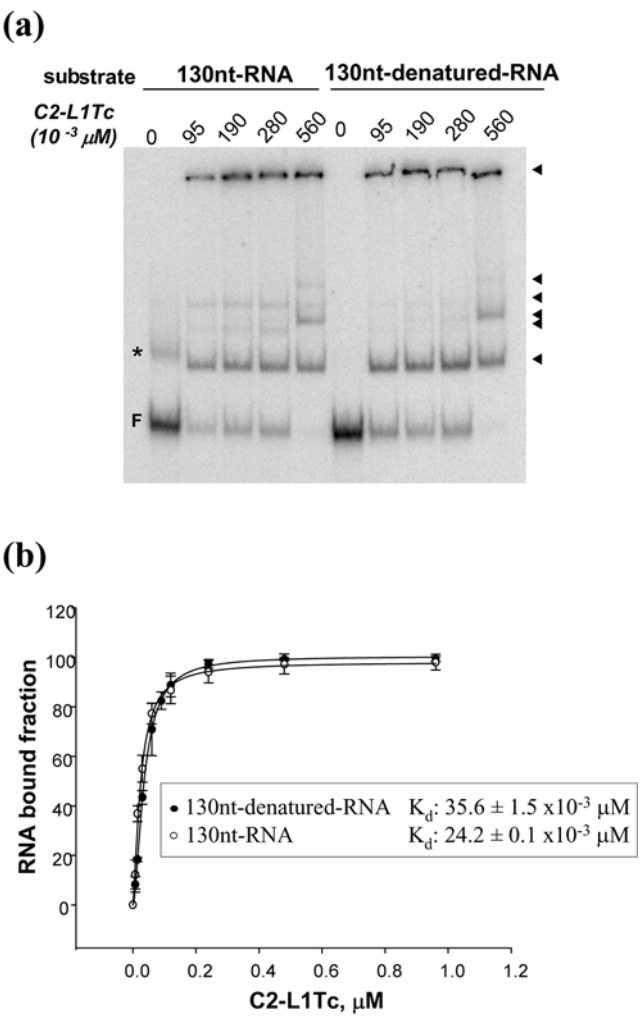

Figure 3 Influence of the secondary structure on the binding capacity of C2-L1Tc to RNA

(**a**) EMSA measuring the binding capacity of C2-L1Tc to native and denatured 130nt-RNA: 0.72 nM native 130nt-RNA (panel 130nt-RNA) or denatured 130nt-RNA (panel 130nt-denatured-RNA) fragments were incubated with increasing concentrations of C2-L1Tc ($95–560 \times 10^{-3}$ μM). The mixtures were incubated at 37 °C for 5 min. Lanes 0 show the reactions carried out in the absence of protein. The asterisk indicates the reduced-mobility free conformation of the 130nt-RNA in the native state. (**b**) Titration curves were obtained from the average of three independent band-shift assays as those shown in (**a**). The average values corresponding to the fraction of radiolabelled 130nt-RNA (○) and 130nt-denatured-RNA (●) bound to C2-L1Tc are plotted against the protein concentration. The curves correspond to the best fit of the Hill equation to the experimental data [R^2 (coefficient of determination) = 0.99 and $R^2 = 0.99$ respectively]. The equation used is as follows:

$$y = \frac{B_{max} \cdot x^{\alpha_H}}{K_d{}^{\alpha_H} + x^{\alpha_H}}$$

where x is the C2-L1Tc concentration and y the fraction of the bound radiolabelled RNA. B_{max}, the maximum value for y, is approx. 1; K_d, defined as the C2-L1Tc concentration at which 50 % of the RNA is bound, is indicated. Arrow heads in (**a**) indicate the shifted bands generated in each case. The electrophoretic mobility of the nucleic acid free forms is also indicated (F).

Mapping of C2-L1Tc-binding domains to RNA

C2-L1Tc has two C2H2 zinc-finger motifs [22] flanked by domains enriched in basic residues such as RRR and RRRKEK (Figure 1a and Table 1). The RRRKEK domain has been described as a NLS (nuclear localization sequence) and also as a DNA-binding motif [30]. To determine the implication of these domains in the binding of C2-L1Tc to nucleic acids, peptides mapping the zinc fingers and the basic stretches (see sequence details in Figure 1a) were incubated with labelled 130nt-RNA. Figures 5(a)–5(c) show the band-shift assays. In order to determine the dissociation constant, K_d, the peptide-bound RNA fraction was quantified and plotted against the peptide concentration and fitted to the Hill equation (Figure 5d). The analysis indicated that peptide 5015 (Figure 5a), which covers the NLS and RRR stretches, has the strongest affinity for RNA with a K_d value of 0.43 ± 0.11 μM. Peptide 5033 containing the NLS motif and the zinc finger located immediately downstream of this domain (named upstream-finger) was also shown to have a high affinity for the RNA molecule with a K_d value of 1.21 ± 0.02 μM (Figure 5b). However, peptide 10987 containing the zinc finger situated towards the C-terminal end of the C2-L1Tc protein (named downstream-finger) and the peptide covering the region located downstream of this zinc finger, peptide 5020, had a lower affinity ($K_d = 6.01 \pm 0.05$ μM and $K_d = 19.79 \pm 3.18$ μM respectively), for the RNA molecule (Figure 5d).

To further analyse the implication of the upstream-finger and the basic regions located at the C2H2 N-terminal end on the RNA-binding capacity of C2-L1Tc, several peptides containing deletions or substitutions (see sequence details in Table 1) were studied using EMSAs (Figures 5a and 5c). The results showed that the partial or complete deletion of NLS (peptides 5016 and 5032) resulted in a decrease ($K_d = 1.86 \pm 0.02$ μM) and a complete loss of affinity for RNA respectively, in spite of the

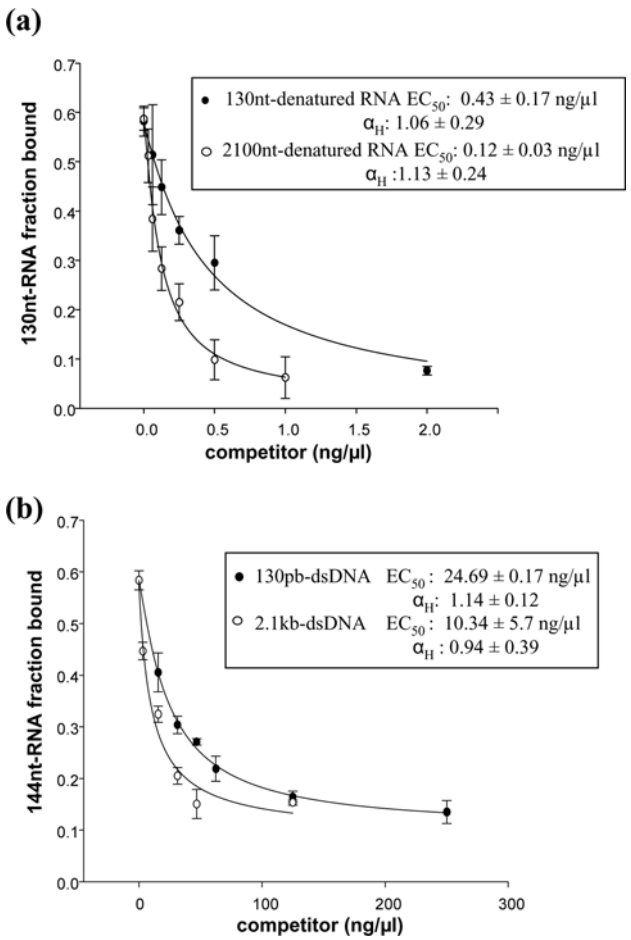

Figure 4 Effect of the nucleic-acid length on the C2-L1Tc nucleic-acid-binding capacity

Titration curves were obtained by EMSA. A total of 0.5 ng of ^{32}P-labelled *in-vitro*-transcribed 130nt-RNA (**a**) was incubated with increasing amounts of unlabelled competitors: 130nt-denatured-RNA (●) and 2100nt-denatured-RNA (o). A total of 0.5 ng of ^{32}P-labelled *in-vitro*-transcribed 144nt-RNA (**b**) was incubated with increasing amounts of unlabelled competitors: 130bp-dsDNA (●) and 2.1kb-dsDNA (o). The samples were further incubated with 0.67 μM C2-L1Tc. The average values corresponding to the fraction of bound radiolabelled 130nt-RNA (**a**) or 144nt-RNA (**b**) from three independent experiments are plotted against the concentration of the nucleic acid employed as the competitor. The curves correspond to the best fit of a four-parameter logistic equation of the experimental data [R^2 (coefficient of determination) = 0.99, $R^2 = 0.99$, $R^2 = 0.99$ and $R^2 = 0.97$ respectively]. The equation used is as follows:

$$y = B_{min} + \frac{B_{max} - B_{min}}{1 + 10^{(\log EC_{50} - x) - \alpha_H}}$$

where x is the competitor concentration, y the fraction of bound-radiolabelled RNA, B_{max} is the maximum binding and B_{min} is the minimum binding. The 50% effective concentration EC_{50}, defined as the concentration of unlabelled nucleic acid required to produce 50% displacement between the upper and lower plateaus of a dose–response curve and the Hill coefficient (α_H), which is an indicative index of co-operativity, are indicated.

fact that the upstream-finger was conserved. In addition, when the RRR stretch was removed from peptide 5015, to generate peptide 5030, the RNA-binding capacity was eliminated although the NLS domain was present (Figure 5c). The substitution of the CCHH motif for SSHH in peptide 5016, peptide 5031, resulted in a 2-fold decrease in binding affinity ($K_d = 3.08 \pm 0.11$ μM) to RNA. These data confirm that the central region of C2-L1Tc containing the upstream-finger and the basic RRR and RRRKEK stretches are the main regions responsible for the binding of the protein to RNA.

Mapping of C2-L1Tc-binding domains to ssDNA

A similar approach to that described above was carried out in order to determine the implication of the C2-L1Tc motifs for ssDNA-binding affinity. Thus band-shift assays were performed using a constant amount of radiolabelled 135nt-ssDNA and an increasing amount of each peptide (Figures 6a–6d). To calculate the K_d, the peptide-bound DNA fraction was quantified and plotted against the peptide concentration (Figure 6e). This analysis showed that peptide 5015 containing the RRR and NLS stretches, and peptide 10987 containing the downstream-finger, have, among the assayed peptides, the highest binding affinity for ssDNA ($K_d = 0.21 \pm 0.06$ μM and $K_d = 0.30 \pm 0.02$ μM respectively) (Figure 6e). The presence of a single-shifted band was observed at different concentrations of peptide 10987 (Figure 6b). Despite the fact that the 10987 peptide concentration increased from 0.09 to 1.8 μM (Figures 6b and 6e), 25% of the ssDNA remained unbound ($B_{max} = 0.74 \pm 0.02$). Interestingly, in spite of the very low affinity of peptide 5032 (containing the upstream-finger) to ssDNA, a similar shifted band was detected at a high concentration of this peptide (Figure 6c).

Peptide 5033 containing the upstream-finger and the complete NLS motif also has a high affinity for ssDNA ($K_d = 0.42 \pm 0.07$ μM). Deletions and mutations of these two motifs resulted in a significant reduction in binding affinity. Thus peptide 5016 which contains a partial deletion of the NLS, but maintains the zinc-finger motif, had less affinity for ssDNA ($K_d = 1.03 \pm 0.08$ μM) than peptide 5033. Peptide 5031, in which the cysteine residues of the upstream-finger were substituted by serine residues, had a 2-fold lower affinity for ssDNA than the 5016 peptide ($K_d = 2.29 \pm 0.07$ μM) (Figure 6d). Peptide 5020, containing the RRR stretch located downstream of the downstream-finger, exhibited a low ssDNA-binding affinity ($K_d = 10.48 \pm 0.45$ μM). Thus, most probably, both zinc-fingers and the basic stretches flanking them (the two RRR and the RRRKEK sequences) participate in the binding of C2-L1Tc to ssDNA.

Mapping of C2-L1Tc-binding domains to dsDNA

The role of the C2-L1Tc motifs in binding to dsDNA was also investigated by incubating 135bp-dsDNA and increasing concentrations of each peptide (Figures 7a–7d). The data were analysed using the Hill equation (Figure 7e). Peptide 10987, which contains the downstream-finger, had only a slight affinity for dsDNA ($K_d = 25.68 \pm 4.22$ μM). However, the peptides containing the upstream-finger and the complete or partial NLS motif, peptides 5033 and 5016, had a high affinity ($K_d = 1.34 \pm 0.08$ μM and 3.12 ± 1.01 μM respectively). The substitution of CCHH for SSHH in this zinc finger (peptide 5031) resulted in a 5-fold decrease in binding affinity ($K_d = 9.13 \pm 0.16$ μM), corroborating the important role of this motif in the binding of C2-L1Tc to dsDNA. Nevertheless, peptide 5032, which lacks the NLS, but maintains the zinc finger, had no affinity for dsDNA. Thus the basic stretches upstream of the zinc fingers (RRR and NLS) also participate in the binding of C2-L1Tc to dsDNA. In fact, peptide 5015 (which contains both regions) showed among the assayed peptides a high affinity ($K_d = 1.51 \pm 0.09$ μM) for dsDNA. Moreover, the deletion of the RRR stretch (peptide 5030) resulted in a complete loss of binding affinity. However, the peptide containing the region located downstream of both C2H2 zinc fingers (peptide 5020), bearing also a basic stretch (RRR), showed a low affinity for dsDNA. The K_d for peptide 5020 was estimated to be 22.83 ± 9.78 μM.

Remarkably, the concentration-dependent binding curves obtained for the binding of each peptide to dsDNA displayed

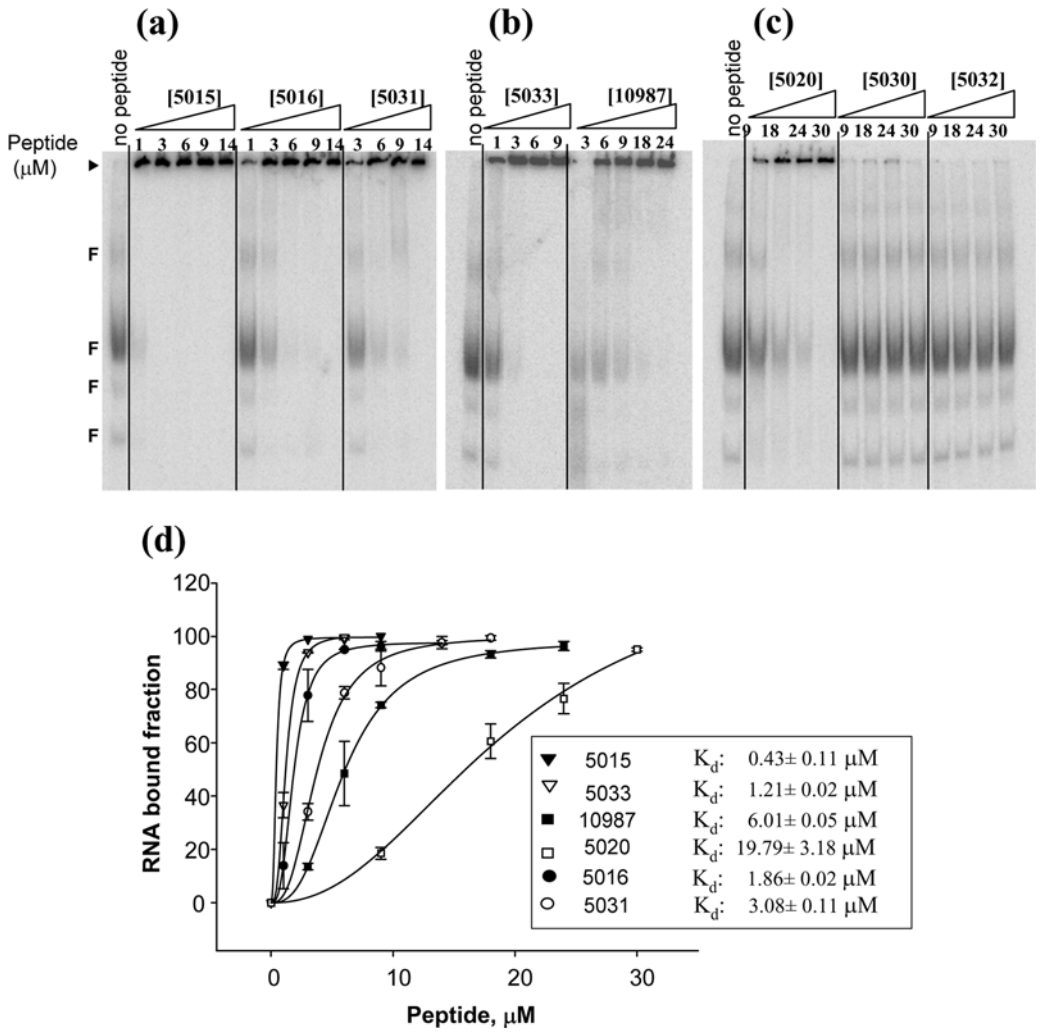

Figure 5 Binding analysis of C2-L1Tc-derived peptides to RNA by EMSA

A 0.72 nM concentration of *in-vitro*-transcribed ^{32}P-labelled 130nt-RNA was pre-incubated with increasing peptide concentrations at 37 °C for 30 min. (**a**) Peptides 5015, 5016 and 5031, from 1 to 14 μM; (**b**) peptide 5033 from 1 to 9 μM, and peptide 10987 from 3 to 24 μM; (**c**) peptides 5020, 5030 and 5032, from 9 to 30 μM (see Figure 1 and Table 1 for the peptide sequence details). Control reactions were performed without peptide (no peptide). The samples were run on 5 % native polyacrylamide gels. (**d**) Curves representing the binding of peptides 5015, 5033, 10987, 5020, 5016 and 5031 to the 130nt-RNA. The data were obtained from three independent experiments as those shown in (**a**)–(**c**). The average values corresponding to the bound RNA fraction were plotted against the peptide concentration. The curves correspond to the best fit of the Hill equation to the experimental data [R^2 (coefficient of determination) $\geqslant$ 0.99]. The equation used was as follows

$$y = \frac{B_{max} \cdot x^{\alpha_H}}{K_d{}^{\alpha_H} + x^{\alpha_H}}$$

where x is the peptide concentration, y is the radiolabelled RNA-bound fraction and B_{max} is the maximum binding. K_d, defined as the peptide concentration at which 50 % of the RNA is bound, is indicated. The position of the 'complex' and 'free' nucleic acids is indicated on the left-hand side of the panels as an arrowhead and F respectively.

significantly different shapes (Figure 7e). The sigmoidal or hyperbolic shapes of the curves indicate whether the binding is co-operative and non-co-operative respectively. To obtain the co-operativity value, a Hill transformation was applied to the dsDNA-binding data (Figure 7f). The Hill coefficients (α_H) for peptides 5015, 5033 and 5020 (containing at least one of the complete basic stretches) were 3.34 ± 0.21, 3.35 ± 0.42 and 2.47 ± 0.48 respectively, indicating a high degree of dsDNA binding co-operativity. However, peptide 10987 (containing the downstream-finger) binds to dsDNA with low co-operativity ($\alpha_H = 1.29 \pm 0.35$). The binding co-operativity of peptide 5016 [which contains the upstream-finger and a fraction of the NLS motif (RKEK)] is also low ($\alpha_H = 1.63 \pm 0.35$). The relatively low-binding co-operativity of 10987 and 5016 peptides is likely to be due to the presence of the zinc fingers. We observed that the substitution of the CCHH in peptide 5016 for SSHH (peptide 5031) increased the degree of co-operativity more than 2-fold ($\alpha_H = 1.63 \pm 0.16$ and $\alpha_H = 3.75 \pm 0.19$ respectively).

Effect on duplex stability of the protein motifs of C2-L1Tc

We have previously shown that the C2-L1Tc protein promotes the exchange of strands on mismatched DNA duplexes in the presence of an excess of single-stranded complementary DNA, even though the protein has no effect on the T_m (melting temperature) of the mismatched duplex [22]. This process probably requires the conjunction of several effects, such as those endowed with stabilization and destabilization properties on a mismatched DNA. In order to evaluate the implication of the C2-L1Tc motifs on the T_m of mismatched duplexes, peptides bearing these motifs were tested in melting assays. As expected, peptides 5030 and 5032, lacking both ssDNA- and dsDNA-binding capability does

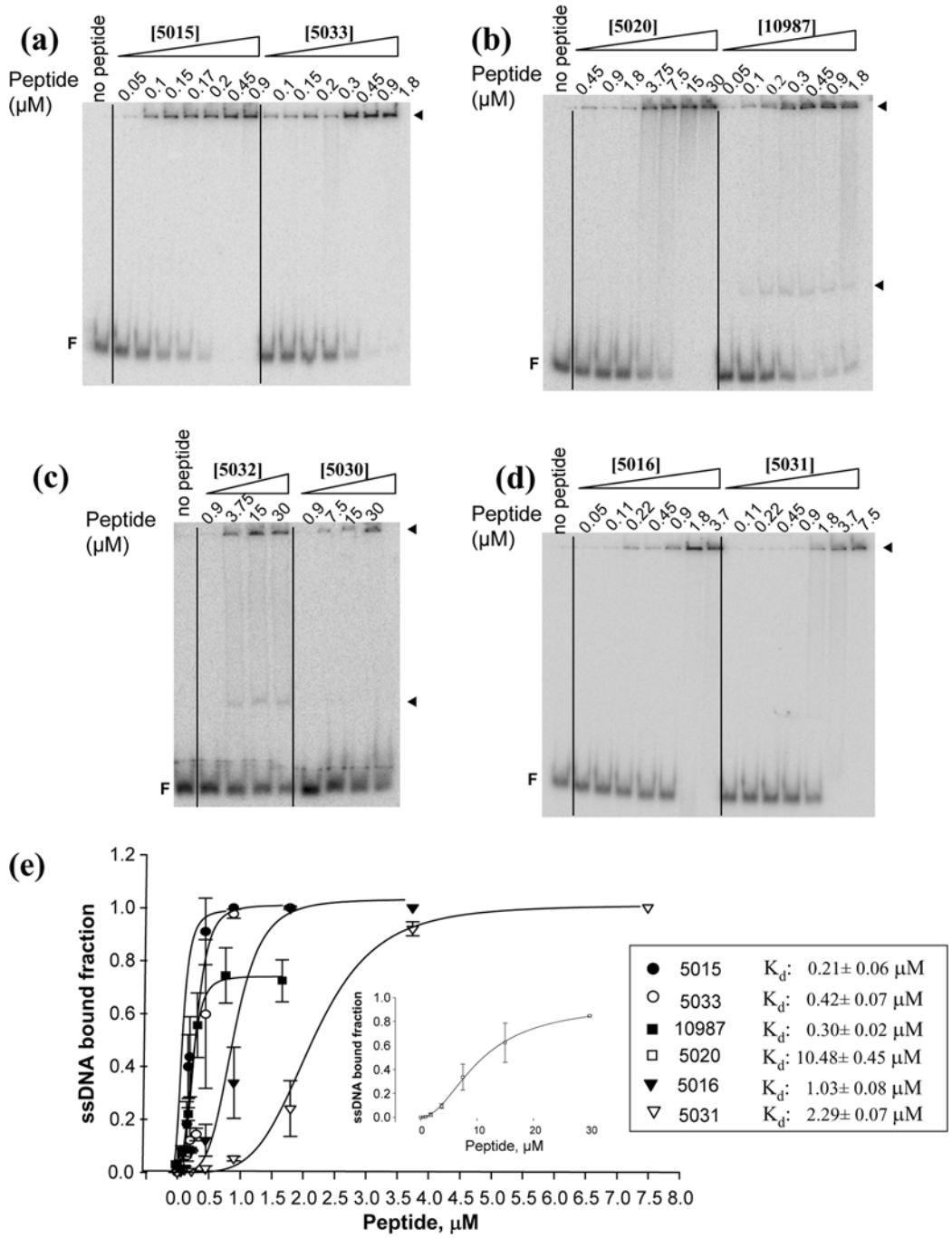

Figure 6 Binding analysis of C2-L1Tc-derived peptides to ssDNA by EMSA

A 0.5 nM concentration of ^{32}P-labelled 135nt-ssDNA was pre-incubated with the indicated increasing concentrations of peptides 5015, 5033 (**a**); 5020, 10987 (**b**); 5032, 5030 (**c**); and 5016 and 5031 (**d**); at 37°C for 30 min (see Figure 1 and Table 1 for peptide sequence details). Control reactions were performed without peptide (no peptide). The samples were run on 5% native polyacrylamide gels. The detected complexes are labelled with an arrowhead and the ssDNA free form as F. (**e**) Curves representing the binding of peptides 5015, 5033, 10987, 5016 and 5031 to the 135nt-ssDNA are represented. The data corresponding to peptide 5020 are represented on the right-hand side. The results were obtained from three independent experiments as those shown in (**a**)–(**d**). The average values corresponding to the bound ssDNA fraction were plotted against the peptide concentration. The curves correspond to the best fit of the Hill equation to the experimental data [R^2 (coefficient of determination) $\geqslant 0.98$]. The equation used was as follows:

$$y = \frac{B_{max} \cdot x^{\alpha_H}}{K_d{}^{\alpha_H} + x^{\alpha_H}}$$

where x is the peptide concentration, y is the radiolabelled RNA-bound fraction and B_{max} is the maximum binding. K_d, defined as the peptide concentration at which 50% of the RNA is bound, is indicated.

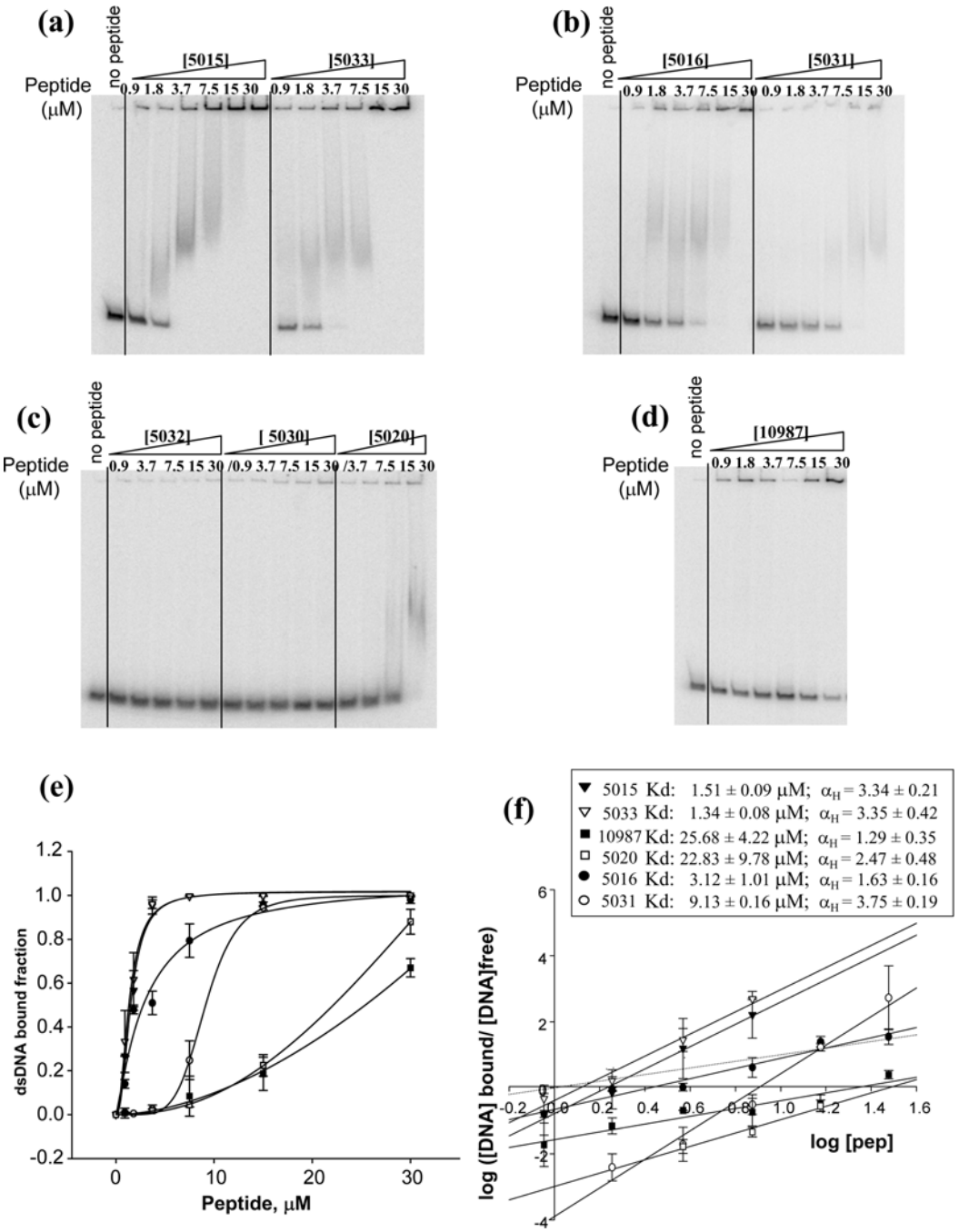

Figure 7 Analysis of the binding affinity of C2-L1Tc-derived peptides to dsDNA by EMSA

A 0.5 nM concentration of ^{32}P-labelled 135bp-dsDNA was pre-incubated with increasing concentrations (0.9–30 μM) of C2-L1Tc-derived peptides: 5015, 5033 (**a**); 5016, 5031 (**b**); 5032, 5030 and 5020 (**c**) and 10987 (**d**) (see Table 1 for sequence composition details); at 37 °C for 30 min. Control reactions were performed without peptide addition (no peptide). Reactions were loaded on to 5 % native polyacrylamide gels and quantification was carried out in a phosphorimager. (**e**) Binding curves representing the binding of peptides 5015, 5033, 10987, 5020, 5016 and 5031 to the 135bp-dsDNA. The results were obtained from three independent experiments as those shown in (**a**)–(**d**). The average values corresponding to the bound dsDNA fraction were plotted against the peptide concentration. The curves correspond to the best fit of the Hill equation to the experimental data [R^2 (coefficient of determination) $\geqslant$ 0.97]. The equation used was as follows:

$$y = \frac{B_{max} \cdot x^{\alpha_H}}{K_d{}^{\alpha_H} + x^{\alpha_H}}$$

where x is the peptide concentration and y is the radiolabelled dsDNA bound fraction. K_d, defined as the peptide concentration at which 50 % of the dsDNA is bound, is indicated in (**e**). (**f**) A Hill transformation was applied to the data obtained from three independent EMSAs. The $\log(Y/(1-Y))$ average values were plotted against the log of the C2-L1Tc concentration, where Y is the bound 135bp-dsDNA fraction. The thin lines correspond to the best fit determined by linear regression ($R^2 \geqslant 0.93$). The slope of the best-fit equation determines the Hill coefficient (α_H) and indicates the degree of co-operativity. The dotted line is the theoretical binding curve for a non-co-operative interaction. The K_d of C2-L1Tc peptides 10987 and 5020 were obtained from the x intercept. These parameters are the average of the values obtained from the equations of three independent experiments. pep, peptide.

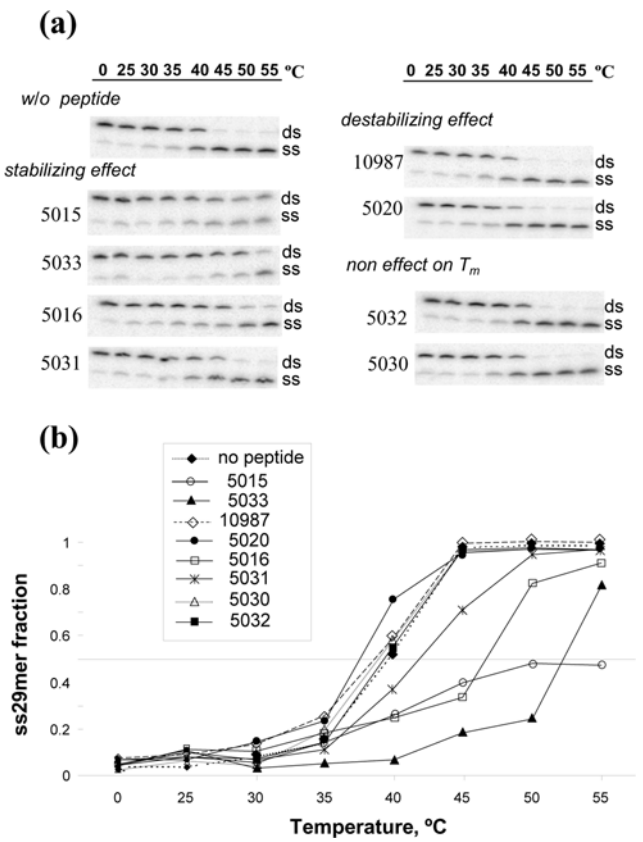

Figure 8 Effect of C2-L1Tc-derived peptides on the T_m of a preformed mismatched dsDNA duplex

(**a**) Phosphorimages of dried gels show the ^{32}P-labelled DNA products formed after incubation for 5 min of 1 mM 29 nt DNA duplex containing four mismatches (^{32}P-29mer/mm29c) with 1 μM C2-L1Tc-derived peptides or without (w/o) peptide at the indicated temperatures. The peptides are grouped according to their influence on the ^{32}P-29mer/mm29c duplex as measured by decreasing (destabilizing), increasing (stabilizing) or no effect on the T_m. The ^{32}P-29mer/mm29c duplex (ds) and the ^{32}P-29mer oligonucleotide (ss) are indicated on the right-hand side of the panels. (**b**) Fraction of unbound ^{32}P-29mer oligonucleotide plotted against temperature. The variation in T_m of the assays, carried out from three independent experiments, was never higher than 1.5 °C. In all assays the same stabilization/destabilization effect was observed for each peptide. Key feature of assayed peptides: 5033, contains the complete NLS motif and the upstream-finger; peptides 5016 and 5032 contain the upstream-finger, but with the partial or complete deletion of NLS respectively; 5015, leads the two basic stretches (RRR and NLS) located upstream of the zinc finger; 5030, similar peptide to 5015 containing the NLS, but lacking the RRR stretch; 5031, similar peptide to 5016, but with a substitution of CCHH for SSHH in the upstream-finger; 10987, contains the zinc finger situated towards the C-terminal end of the C2-L1Tc protein (named downstream-finger); 5020, peptide covering the region located downstream of this zinc finger.

not have any influence on the T_m of this imperfect DNA duplex (Figure 8 and Table 1). Peptide 5015, containing both basic motifs (RRR and NLS sequence), strongly prevented the melting of the duplex. Peptide 5033, containing the NLS sequence together with the upstream-finger, also prevented the melting of the duplex, increasing the T_m from 40 °C to 52 °C. The partial deletion of the NLS motif, peptide 5016, and the additional substitution of cysteine residues by serine residues in the upstream-finger, peptide 5031, resulted in a reduction in the T_m (47 °C and 42 °C respectively compared with 52 °C of the 5033 peptide), although both of them maintained the stabilization effect. In contrast, the peptides containing the downstream-finger or the RRR region located downstream of both zinc fingers (peptides 10987 and 5020 respectively) induced a slight decrease in the T_m of the duplex containing four internal mismatches (Figures 8a and 8b).

DISCUSSION

Retrotransposition of LINE requires the interaction at different steps of some of the proteins that the elements code for and the nucleic acids. Thus the interaction between the proteins encoded by these elements and the intermediate RNA forming a RNP (ribonucleoparticle), as well as those between the newly formed RNP with the target DNA, are obligated processes [31,32]. Even though there is a large diversity among non-LTR retrotranposable elements, a conserved domain containing a potential nucleic-acid-binding motif is, however, retained in most of them [24]. It has been previously shown that the C2-L1Tc protein encoded by the sequence located at the 3′-end of the *T. cruzi* L1Tc element binds to nucleic acids and that it has NAC activity [22]. The results of the present study show that C2-L1Tc exhibits a preference for RNA binding. The high binding affinity that C2-L1Tc shows for RNA suggests that it may have an important role *in vivo* for L1Tc mRNA binding. This capability to bind RNA is also present in other proteins encoded by other non-LTR retroelements, such as ORF1-derived proteins (ORF1p) from the human and mouse L1 elements and from that encoded in the *Drosophila* I factor [7,32,33]. Thus the high affinity that these proteins have for RNA seems to be essential for mobilization of RNA-intermediate-mediated transposable elements. In fact, some specific point mutations in the ORF1p that reduce the binding affinity of the protein for RNA lead to the formation of altered RNPs and to a severe reduction of the retrotransposition efficiency [32].

The C2-L1Tc-binding profile to RNA corresponds to a non-co-operative-binding model. However, the affinity of the C2-L1Tc protein for this type of nucleic acid molecule increases with the RNA size, showing a clear non-specific sequence affinity. Consequently, since the L1Tc mRNA has a large size (5 kb), it is expected that the C2-L1Tc protein may bind to several positions on the RNA. C2-L1Tc may have binding preferences for specific conformations of the L1Tc RNA. In fact, we have observed that the RNA conformation influences the binding capacity of the C2-L1Tc protein for RNA molecules. Thus the affinity of C2-L1Tc for the RNA molecule is reduced when the RNA molecule is in a denatured state. Taken together, the results presented suggest that the C2-L1Tc protein plays an important role in the binding to the L1Tc transcript and consequently in the RNP formation. Furthermore, the data also suggest that the C2-L1Tc–RNA association may protect the L1Tc transcript from degradation.

We believe that the RRR and RRRKEK (NLS) domains, as well as the zinc finger located immediately downstream of these basic stretches (upstream-finger), are the main motifs responsible of the high binding affinity that the C2-L1Tc protein exhibits for the RNA molecule. The co-operativity of these motifs and the relative position of them in the protein should probably play an essential role in the binding of the protein to RNA as the affinity of these peptides for the RNA is substantially lower than that shown by the full-length protein. Previously reported studies have indicated that the C2-L1Tc protein endowed with NAC activity stabilizes complementary DNA duplexes and does not modify the T_m of mismatched duplexes [22]. In the present study we provide some insights into the molecular mechanisms involved in this activity. In fact, C2-L1Tc has a 16-fold higher affinity for ssDNA than for dsDNA. Also, a higher affinity for binding single-stranded nucleic acids than double-stranded has also been described to exist in the NAC proteins encoded by HIV-1 NC [34,35], the LTR retrotransposons from the yeast Ty-1 NC-like protein [36] and other non-LTR retrotransposons, such as the mouse L1 protein and *Drosophila* I factor ORF1p [3,11]. Since some L1 ORF1p mutant proteins, in which neither the binding

affinity for RNA nor the RNP formation have been altered, exhibit a reduced retrotransposition rate [10,32], it has been suggested that this protein having NAC activity should have an additional function in the TPRT mechanism at a subsequent step of that involving the RNP formation. This additional function may be correlated with the capability of the protein for maintaining a delicate balance between its ability to promote both stabilization and destabilization of the helix as it has been demonstrated to be required for effective L1 retrotransposition [37].

EMSA analysis using synthetic peptides covering different active regions present in the C2-L1Tc protein show that the region containing the RRR and RRRKEK (NLS) domains has a high capability for DNA binding. The affinity is higher for single-stranded than for double-stranded molecules. Deletion of any of these motifs produces a complete loss of the binding capacity for both types of nucleic acids. The binding affinity for the DNA molecules, but not for RNA molecules, of peptides bearing the mentioned motifs is similar to that shown by the full-length protein. This fact suggests that the synergic effect of the motifs present in the protein is not essential for DNA binding. The C2-L1Tc region that covers both the NLS domain and the zinc finger located immediately downstream of these basic stretches (upstream-finger) also shows a high affinity for both nucleic acid molecules. The substitution of cysteine residues in that finger for serine residues induces a significantly higher decrease in binding affinity for dsDNA than for single-stranded nucleic acids. This indicates that the upstream-finger plays a relevant role in binding to dsDNA. However, the peptide bearing the zinc finger situated towards the C-terminal end of the C2-L1Tc protein (downstream-finger) has the ability to bind mainly ssDNA. These results indicate the two zinc fingers present in C2-L1Tc have a differential behaviour relative to nucleic-acid binding. Other proteins containing multiple C2H2 zinc fingers (e.g. TFIIIA) are also able to bind both DNA and RNA molecules [38,39], although the zinc fingers involved in this binding are specific for each type of nucleic acid molecules [35,40].

Previous strand-exchange experiments have shown that the two C2H2 zinc fingers and the basic domains located upstream of the first C2H2 cysteine motif are involved in the NAC activity [22]. Our results show that the protein regions responsible for the nucleic-acid binding are the same as those previously described to be implicated in NAC activity. However, the data presented in the present study (Table 1) suggest that both processes are uncoupled. Thus, although the binding to nucleic acids is essential for NAC activity, the affinity of isolated motifs for double- or single-stranded nucleic acids is not directly correlated with NAC activity. This fact seems to be a general characteristic of the NAC proteins. This feature is probably due to the need for establishing a proper balance between single-stranded and double-stranded interactions. This balance seems to be required to promote both stabilization and destabilization of the nucleic acid helix [35,41]. Previous studies have shown that the effective NAC activity (strand-annealing function) of the HIV-1 nucleocapsid protein is correlated with the protein ability to rapidly bind and dissociate from nucleic acids [42]. Our results show that the RRR and RRRKEK (NLS) domains and the upstream-finger implicated in binding to both ssDNA and dsDNA have a stabilizing effect on mismatched duplexes. However, both the downstream-finger and the basic motif located downstream of the finger, mainly implicated in binding to ssDNA, have a destabilization effect. We suggest, therefore, that in order to properly function as a NAC, specific motifs of the C2-L1Tc protein must maintain a proper balance between the binding affinity for single- and double-strand nucleic acids and the capability for stabilizing and destabilizing the nucleic acid helix.

AUTHOR CONTRIBUTION

Sara Heras designed and performed most of the experiments, analysed the data and wrote the initial draft of the manuscript. Carmen Thomas designed the experiments, analysed the data, discussed results and contributed in writing the manuscript. Francisco Macias carried out experiments and contributed to manuscript preparation. Manuel Patarroyo carried out the synthesis of the peptides. Carlos Alonso discussed and corrected the manuscript prior to submission. Manuel López conceived the research, designed experiments, analysed the data, was involved in scientific discussion and also wrote the manuscript.

ACKNOWLEDGEMENTS

We thank M. Caro for technical assistance in the purification of the C2-L1Tc recombinant protein. We also thank Dr Cristina Romero for her help with data analysis and Dr Javier Cáceres for critical reading of the manuscript.

FUNDING

This work was supported by Plan Nacional I+D+I [MICINN (Ministerio de Ciencia e Innovación)] [grant numbers BFU2006-07972, BFU2007-64999]; PAI (Plan Andaluz de Investigación; Junta de Andalucía) [grant number P05-CVI-01227]; ISCIII-RETIC (Instituto de Salud Carlos III-Redes Temáticas de Investigación Cooperativa en Salud), Spain [grant number RD06/0021/0014]; a PAI Predoctoral Fellowship [grant number P05-CVI-01227 (to F.M.)]; and Colciencias [grant number RC-2007 (to M.E.P.)].

REFERENCES

1 Curcio, M. J. and Derbyshire, K. M. (2003) The outs and ins of transposition: from mu to kangaroo. Nat. Rev. Mol. Cell. Biol. **4**, 865–877
2 Boeke, J. D. and Corces, V. G. (1989) Transcription and reverse transcription of retrotransposons. Annu. Rev. Microbiol. **43**, 403–434
3 Luan, D. D., Korman, M. H., 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
4 Moran, J. V., Holmes, S. E., Naas, T. P., DeBerardinis, R. J., Boeke, J. D. and Kazazian, Jr, H. H. (1996) High frequency retrotransposition in cultured mammalian cells. Cell **87**, 917–927
5 Burke, W. D., Malik, H. S., Rich, S. M. and Eickbush, T. H. (2002) Ancient lineages of non-LTR retrotransposons in the primitive eukaryote, *Giardia lamblia*. Mol. Biol. Evol. **19**, 619–630
6 Hohjoh, H. and Singer, M. F. (1996) Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. EMBO J. **15**, 630–639
7 Hohjoh, H. and Singer, M. F. (1997) Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. EMBO J. **16**, 6034–6043
8 Kolosha, V. O. and Martin, S. L. (1997) *In vitro* properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. Proc. Natl. Acad. Sci. U.S.A. **94**, 10155–10160
9 Martin, S. L. and Bushman, F. D. (2001) Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. Mol. Cell. Biol. **21**, 467–475
10 Martin, S. L., Cruceanu, M., Branciforte, D., Wai-Lun Li, P., Kwok, S. C., Hodges, R. S. and Williams, M. C. (2005) LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. J. Mol. Biol. **348**, 549–561
11 Dawson, A., Hartswood, E., Paterson, T. and Finnegan, D. J. (1997) A LINE-like transposable element in *Drosophila*, the I factor, encodes a protein with properties similar to those of retroviral nucleocapsids. EMBO J. **16**, 4448–4455
12 Eickbush, T. H. and Malik, H. S. (2002) Origins and evolution of retrotransposons. In Mobile DNA II (Craig, N. L., Craigie, R., Gellert, M. and Lambowitz, A. M., eds), pp. 1111–1146, ASM Press, Washington DC
13 El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E. A., Hertz-Fowler, C. et al. (2005) Comparative genomics of trypanosomatid parasitic protozoa. Science **309**, 404–409
14 Donelson, J. E. (1996) Genome research and evolution in trypanosomes. Curr. Opin. Genet. Dev. **6**, 699–703
15 Martin, F., Maranon, C., Olivares, M., Alonso, C. and Lopez, M. C. (1995) Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzymes. J. Mol. Biol. **247**, 49–59
16 Olivares, M., Thomas, M. C., Lopez-Barajas, A., Requena, J. M., Garcia-Perez, J. L., Angel, S., Alonso, C. and Lopez, M. C. (2000) Genomic clustering of the *Trypanosoma cruzi* non-long terminal L1Tc retrotransposon with defined interspersed repeated DNA elements. Electrophoresis **21**, 2973–2982

17 El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G. et al. (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science **309**, 409–415
18 Olivares, M., Alonso, C. and Lopez, M. C. (1997) The open reading frame 1 of the L1Tc retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic-apyrimidinic nuclease activity. J. Biol. Chem. **272**, 25224–25228
19 Olivares, M., Thomas, M. C., Alonso, C. and Lopez, M. C. (1999) The L1Tc, long interspersed nucleotide element from *Trypanosoma cruzi*, encodes a protein with 3′-phosphatase and 3′-phosphodiesterase enzymatic activities. J. Biol. Chem. **274**, 23883–23886
20 Garcia-Perez, J. L., Gonzalez, C. I., Thomas, M. C., Olivares, M. and Lopez, M. C. (2003) Characterization of reverse transcriptase activity of the L1Tc retroelement from *Trypanosoma cruzi*. Cell. Mol. Life Sci. **60**, 2692–2701
21 Olivares, M., Garcia-Perez, J. L., Thomas, M. C., Heras, S. R. and Lopez, M. C. (2002) The non-LTR (long terminal repeat) retrotransposon L1Tc from *Trypanosoma cruzi* codes for a protein with RNase H activity. J. Biol. Chem. **277**, 28025–28030
22 Heras, S. R., Lopez, M. C., Garcia-Perez, J. L., Martin, S. L. and Thomas, M. C. (2005) The L1Tc C-terminal domain from *Trypanosoma cruzi* non-long terminal repeat retrotransposon codes for a protein that bears two C2H2 zinc finger motifs and is endowed with nucleic acid chaperone activity. Mol. Cell. Biol. **25**, 9209–9220
23 Krishna, S. S., Majumdar, I. and Grishin, N. V. (2003) Structural classification of zinc fingers: survey and summary. Nucleic Acids Res. **31**, 532–550
24 Malik, H. S., Burke, W. D. and Eickbush, T. H. (1999) The age and evolution of non-LTR retrotransposable elements. Mol. Biol. Evol. **16**, 793–805
25 Sarin, V. K., Kent, S. B., Tam, J. P. and Merrifield, R. B. (1981) Quantitative monitoring of solid-phase peptide synthesis by the ninhydrin reaction. Anal. Biochem. **117**, 147–157
26 Wang, S. S., Wang, B. S., Hughes, J. L., Leopold, E. J., Wu, C. R. and Tam, J. P. (1992) Cleavage and deprotection of peptides on MBHA-resin with hydrogen bromide. Int. Pept. Protein Res. **40**, 344–349
27 Heras, S. R., Lopez, M. C., Olivares, M. and Thomas, M. C. (2007) The L1Tc non-LTR retrotransposon of *Trypanosoma cruzi* contains an internal RNA-pol II-dependent promoter that strongly activates gene transcription and generates unspliced transcripts. Nucleic Acids Res. **35**, 2199–2214
28 Thomas, M. C., Garcia-Perez, J. L., Alonso, C. and Lopez, M. C. (2000) Molecular characterization of KMP11 from *Trypanosoma cruzi*: a cytoskeleton-associated protein regulated at the translational level. DNA Cell Biol. **19**, 47–57
29 Barroso-delJesus, A., Tabler, M. and Berzal-Herranz, A. (1999) Comparative kinetic analysis of structural variants of the hairpin ribozyme reveals further potential to optimize its catalytic performance. Antisense Nucleic Acid Drug Dev. **9**, 433–440
30 Cokol, M., Nair, R. and Rost, B. (2000) Finding nuclear localization signals. EMBO Rep. **1**, 411–415
31 Seleme, M. C., Disson, O., Robin, S., Brun, C., Teninges, D. and Bucheton, A. (2005) *In vivo* RNA localization of I factor, a non-LTR retrotransposon, requires a cis-acting signal in ORF2 and ORF1 protein. Nucleic Acids Res. **33**, 776–785
32 Kulpa, D. A. and Moran, J. V. (2005) Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. Hum. Mol. Genet. **14**, 3237–3248
33 Martin, S. L. (2006) The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. J. Biomed. Biotechnol. **2006**, 45621
34 Surovoy, A., Dannull, J., Moelling, K. and Jung, G. (1993) Conformational and nucleic acid binding studies on the synthetic nucleocapsid protein of HIV-1. J. Mol. Biol. **229**, 94–104
35 Levin, J. G., Guo, J., Rouzina, I. and Musier-Forsyth, K (2005) Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. Prog. Nucleic Acid Res. Mol. Biol. **80**, 217–286
36 Cristofari, G., Ficheux, D. and Darlix, J. L. (2000) The GAG-like protein of the yeast Ty1 retrotransposon contains a nucleic acid chaperone domain analogous to retroviral nucleocapsid proteins. J. Biol. Chem. **275**, 19210–19217
37 Martin, S. L., Bushman, D., Wang, F., Wai-Lun Li, P., Walker, A., Cummiskey, J., Branciforte, D. and Williams, M. C. (2008) A single amino acid substitution in ORF1 dramatically decreases L1 retrotransposiciton and provides insight into nucleic acid chaperone activity. Nucleic Acids Res. **36**, 5845–5854
38 Christensen, J. H., Hansen, P. K., Lillelund, O. and Thogersen, H. C. (1991) Sequence-specific binding of the N-terminal three-finger fragment of *Xenopus* transcription factor IIIA to the internal control region of a 5S RNA gene. FEBS Lett. **281**, 181–184
39 Searles, M. A., Lu, D. and Klug, A. (2000) The role of the central zinc fingers of transcription factor IIIA in binding to 5S RNA. J. Mol. Biol. **301**, 47–60
40 Matthews, J. M. and Sunde, M. (2002) Zinc fingers: folds for many occasions. IUBMB Life **54**, 351–355
41 Hargittai, M. R., Gorelick, R. J., Rouzina, I. and Musier-Forsyth, K. (2004) Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. J. Mol. Biol. **337**, 951–968
42 Cruceanu, M., Gorelick, R. J., Musier-Forsyth, K., Rouzina, I. and Williams, M. C. (2006) Rapid kinetics of protein–nucleic acid interation is a major component of HIV-1 nucleocapsid protein's nucelic acid chaperone function. J. Mol. Biol. **363**, 867–877
43 Heras R.S., Thomas M. C., García M., Felipe P., García-Pérez J.L., Ryan M. and López, M. C. (2006) L1Tc non-LTR retrotransposons contain a functional viral self-cleaving 2A sequence in frame with the active proteins they encode. Cell. Mol. Life Sci, **63**, 1449–1160
44 Henriet, S., Richer, D., Bernacchi, S., Decroly, E., Vigne, R., Ehresmann, B., Ehresmann, C., Paillart, J. C. and Marquet, R. (2005) Cooperative and specific binding of Vif to the 5′ region of HIV-1 genomic RNA. J. Mol. Biol. **354**, 55–72

Received 26 May 2009/11 September 2009; accepted 14 September 2009
Published as BJ Immediate Publication 14 September 2009, doi:10.1042/BJ20090766