



HAL
open science

Nucleic acids 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, Manuel C López

► **To cite this version:**

Sara R Heras, M. Carmen Thomas, Francisco Macias, Manuel E. Patarroyo, Carlos Alonso, et al.. Nucleic acids binding properties of the C2-L1Tc Nucleic Acid Chaperone encoded by L1Tc Retrotransposon. *Biochemical Journal*, 2009, 424 (3), pp.479-490. 10.1042/BJ20090766 . hal-00479197

HAL Id: hal-00479197

<https://hal.science/hal-00479197>

Submitted on 30 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nucleic acids 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[§], 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 and [‡]Universidad Nacional de Colombia. [§]Centro de Biología Molecular "Severo Ochoa", CSIC, Madrid –Spain.

Address correspondence to: ¹M.C. López, Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina "López Neyra", CSIC. Parque Tecnológico de Ciencias de la Salud, Avda. del Conocimiento s/n, 18100 Granada, Spain. Tel: +34 958 181 661. Fax: +34 958 181 632. E-mail: mclopez@ipb.csic.es

Short title: Nucleic acids binding properties of the C2-L1Tc

SYNOPSIS

It has been reported that the C2-L1Tc protein located in the *T. cruzi* LINE L1Tc 3' terminal end has nucleic acid chaperone (NAC) 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 the transcription factor TFIIIA. The cysteine motifs are flanked by positively charged amino acid regions. The data presented in this paper shows that the C2-L1Tc recombinant protein has at least 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-cooperative binding model. The zinc fingers present in C2-L1Tc have different binding affinity to nucleic acid molecules and also different NAC activity. The RRR and RRRKEK (NLS) 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 carboxyl 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, like C2-L1Tc. These results provide further insights into the essential properties of the C2-L1Tc protein as a nucleic acid chaperone.

Key words: Nucleic acid chaperone, LINE; Binding to Single and Double Stranded Nucleic Acids, Zinc finger, *Trypanosoma cruzi*, Retrotransposition.

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 long terminal repeats (LTR) are similar in structure and retrotransposition mechanism to those of retroviruses [2]. The elements lacking LTR, also called LINE, 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 target-primed reverse transcription (TPRT) [3].

Most non-LTR retrotransposons have two ORFs. 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 for 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 to 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 nucleic acid chaperone [9]. The highly basic region located at the COOH-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 been also 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 nucleocapsid basic portion of the retroviral *gag* polyproteins. The motif is also present in the proteins encoded by the first open reading frame 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 *Trypanosomatidae* family. This parasite is the agent responsible for the 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 relative 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 endonuclease [18], a 3' phosphatase, a 3' phosphodiesterase [18,19], a reverse transcriptase [20] and a RNase H activity [21].

We have recently described that the L1Tc carboxy-terminal encodes a protein, termed C2-L1Tc, has nucleic acid chaperone activity (NAC) and binds to several types of nucleic acids [22]. C2-L1Tc catalyzes the rearrangements 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 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 cooperate and are essential for the NAC activity [22]. The C2H2 motifs were first described in the *Xenopus*

laevis transcription factor IIIA (TFIIIA). 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 *Caenorhabditis 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 nucleic acid chaperone activity of the L1 elements [10]. However, this mechanism has not yet been completely understood in molecular terms. In the present paper we have analyzed 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 determined the implication of specific regions of C2-L1Tc in the nucleic acid binding properties of the protein and its relation to the NAC activity.

EXPERIMENTAL

Cloning and protein purification of the recombinant C2-L1Tc

The region of the L1Tc element between positions 3976 to 4851 (GenBank accession no. **AF208537**) [15] was cloned into the pCAS B vector (Active motif[®]) as described previously [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 solid phase peptide synthesis (SPPS) strategy on a p-methylbenzhydrylamide (MBHA) resin [26]. Purity was checked by high performance liquid chromatography (HPLC). Peptide sequences are shown in table I. Peptides were dissolved in sterile 1X PBS containing 30 μ M zinc chloride, at a 500 μ M final concentration.

RNA and DNA synthesis

144nt-RNA and 130nt-RNA were generated using a *Hind* III-digested pGR77 plasmid that contains 77 bp corresponding to the internal promoter of L1Tc [22,27] and a *Hind* III-digested TcKMP11n clone whose sequence is not related to L1Tc, respectively, (GenBank accession no. **AJ000077**) [28]. *In vitro* transcription was carried out using as template 2 μ g of linearized DNA and T7 RNA polymerase as described by Barroso-delJesus et al [29]. Thirty μ Ci of [α -³²P]UTP (3000C_i · mmol⁻¹) 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 gel, precipitated and resuspended in diethyl pyrocarbonate-treated water. A 2100nt-RNA containing a fragment of L1Tc mRNA from nucleotides 232 to 1468 was generated *in vitro* using the T7 polymerase from *Xba*I-linearized pCMV4NL1Tc as described above. Briefly, an 1234 bp *Kpn*I-*Xba*I PCR fragment from L1Tc (GenBank accession no. **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). 130 nt and 2100 nt single stranded RNAs (130nt-denatured-RNA and 2100nt-denatured-RNA) were obtained by heating for 2 minutes at 65 °C followed by cooling in ice for 2 minutes.

2.01kb-dsDNA and 135bp-dsDNA were obtained by PCR using a T7 oligonucleotide (5'GTAATACGACTCACTATAGGG3') as sense primer. To amplify 2.01kb-dsDNA, pCMV4NL1Tc was used as template and a 1446 L1Tc oligonucleotide (5' GCTGATGCGGCGTAGATA) as antisense primer. A 135bp-dsDNA was amplified using kmp2 (5'TTCCTCAAGAGTGGTGGC3') as antisense primers and TcKMP11n clone as template [25]. The PCR products were purified by gel filtration (Sephadex G-50). A 135nt single stranded DNA (135bp-ssDNA) fragment was generated by PCR reaction 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-135pb-dsDNA. Following the PCR amplification, the phosphorylated minus-strand of the PCR product was removed by digestion with lambda exonuclease (Fermentas). After the inactivation of the enzyme by heating at 80°C for 10 minutes, the plus strand was purified by gel filtration (sephadex-G25) and precipitated with ethanol. Both 135pb-dsDNA and 135nt-ssDNA products were 5' end labeled using [γ - 32 P]ATP and T4 polynucleotide kinase (Roche). The unincorporated isotope was removed by gel filtration chromatography (Sephadex G25).

Electrophoretic mobility shift assays (EMSAs)

In the dsDNA- and ssDNA- binding experiments, 32 P-labeled 135bp-dsDNA or 32 P-labeled 135nt-ssDNA (0.5 nM) and increasing amounts of C2-L1Tc protein (0.02- 2.8 μ M) or the indicated concentration of each peptide, were incubated in 20 μ l of binding buffer (20 mM HEPES (pH 7.5), 100 mM NaCl, 2 mM MgCl₂, 2 mM DTT, 5% glycerol and 100 μ g/ml bovine serum albumin), for 30 min at 37°C. For the RNA-binding experiments, 32 P-labeled 130nt-RNA (0.72 nM) was incubated with increasing concentrations of the C2-L1Tc (0.015-0.733 μ M) protein or the synthetic peptides (1-30 μ M) in 16 μ l of binding buffer containing 5 U RNasin (Ambion), for 30 min at 37°C. To compare the affinity of C2-L1Tc for 130nt-RNA and 130nt-denatured-RNA (Figure 3), the reactions containing the native or denatured *in vitro* transcripts and the indicated amount of C2-L1Tc protein were incubated for 5 minutes to avoid the formation of any secondary structure in the denatured 130nt-RNA. All these reactions were incubated in ice and stopped by addition of 8 μ l of dye solution (50% glycerol, 0.1% bromophenol blue, 0.1% xylenecyanol). Nucleic acid-protein complexes were resolved by electrophoresis through 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 analyzed with ImageQuant 5.2 (Amersham Biosciences).

Competition assays were performed by incubation of C2-L1Tc protein (0.67 μ M) with radiolabeled 130nt-RNA (0.72 nM) and increasing amount of the cold 130nt-denatured-RNA and cold 2100 nt-denatured-RNA in the binding buffer at 37°C for 5 minutes to avoid the formation of any secondary structure in the competitors. In a similar way, the binding affinity of the 130pb-dsDNA and 2.1kb-dsDNA fragments was calculated mixing increasing concentrations of these molecules with the radiolabeled 144nt-RNA transcript (0.65 nM). These reactions were incubated also in binding buffer with C2-L1Tc protein

(0.67 μM) at 37°C for 5 minutes. The reaction was stopped as described above. Electrophoretic analysis of the generated products was done as described above.

DNA melting assays

Assays were performed as described [9,22]. Briefly, a preannealed mismatched duplex was made by mixing 200 nM ^{32}P -labeled 29mer 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. 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, 0.1% triton). The sample was incubated for 5 min at temperatures ranging from 25° 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, 0.2% xylene cyanol blue). Gels and analysis were performed as described. The melting effect was monitored on native 15% polyacrylamide gels. The dried gel was analyzed 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 (Figure 1a and b), a non-LTR retrotransposon from *T. cruzi*, exhibits nucleic acids chaperone activity (NAC) and that it is able to bind to several types of nucleic acids (NA) with different affinity [22]. To determine whether the different affinity for NAs binding are due to specific features of the NAs, EMSA experiments were carried out using increasing protein concentrations and several radiolabeled molecules, (RNA, a single stranded DNA, and a double stranded DNA), having the same sequence composition. As shown in Figure 2a and consistent with previous studies when a low protein concentration of C2-L1Tc was incubated with RNA (130nt-RNA) a discrete product was formed. If the protein concentration increases the amount of reduced-mobility products also increase (Figure 2a). When a single stranded DNA (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 double stranded DNA (135pb-dsDNA) was incubated with the protein a faint smear was observed together with a reduction of the amount of dsDNA free form. 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 $24 \pm 5 \times 10^{-3} \mu\text{M}$ for the RNA, $98 \pm 20 \times 10^{-3} \mu\text{M}$ for single stranded DNA, and $1.78 \pm 0.35 \mu\text{M}$ for double stranded DNA. Thus, we may conclude that C2-L1Tc has at least 16-fold higher affinity for single stranded than for double stranded DNA 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 ($K_d = 35.6 \pm 1.5 \times 10^{-3} \mu\text{M}$ for 130nt-denatured-RNA and $K_d = 24.2 \pm 0.1 \times 10^{-3} \mu\text{M}$ for 130nt-RNA) indicated that the RNA conformation influences the C2-L1Tc binding capability (Figure 3a and 3b). However, the affinity of C2-L1Tc for the denatured 130nt-RNA is still higher than the observed affinity for single stranded 135nt-DNA ($K_d = 98 \pm 20 \times 10^{-3} \mu\text{M}$) as an indication that the C2-L1Tc protein has 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). The result of a representative single assay in which a constant amount of protein and a radio-labeled 130nt-RNA and, as cold competitors, a denatured 130nt-RNA (130nt-denatured-RNA) or a denatured 2,100nt-RNA (2100nt-denatured-RNA) were used is shown in figure 4a. 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 radiolabeled RNA, and the Hill coefficient (α_H) which reflects cooperativity, were determined. The EC_{50} values were $0.43 \pm 0.17 \text{ ng}/\mu\text{l}$ for the 130nt-denatured-RNA and $0.12 \pm 0.03 \text{ ng}/\mu\text{l}$ for the 2100nt-denatured-RNA (Figure 4a). A similar assay was performed using a labeled 144nt-RNA and increasing concentrations of two unlabeled double stranded DNA molecules, 130bp-dsDNA and 2.1kb-dsDNA (Figure 4b). In this case, the EC_{50} was $24.69 \pm 0.17 \text{ ng}/\mu\text{l}$ and $10.34 \pm 5.7 \text{ ng}/\mu\text{l}$, respectively. These results revealed that although C2-L1Tc has lower affinity for double stranded than for single stranded nucleic acids the binding affinity of C2-L1Tc to both types of NAs increases as the length of the nucleic acids increases. Furthermore, the Hill coefficient obtained in all cases is close to (\sim) 1 corresponding to a non-cooperative 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, *versus* 1.06 ± 0.29 and 1.14 ± 0.12 , respectively, for NAs of shorter length.

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 nuclear localization signal (NLS) 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 labeled 130nt-RNA. Figure 5a-c shows the band shift assays. In order to determine the dissociation constant, K_d , the peptide-bound RNA fraction was quantified and plotted *versus* peptide concentration and fitted to the Hill equation (Figure 5d). The analysis indicated that peptide 5015 (Figure 5a), that covers the NLS and RRR stretches has the strongest affinity for RNA with a K_d value of $0.43 \pm 0.11 \mu\text{M}$. Peptide 5033 containing the NLS motif and the zinc finger located immediately downstream of this domain (named upstream-finger) also showed to have high affinity for RNA molecule with a K_d value of $1.21 \pm 0.02 \mu\text{M}$ (Figure 5b). However, peptide 10987 containing the zinc finger situated towards the carboxyl terminal end of the C2-L1Tc protein (named downstream-finger) and the peptide covering the region located

downstream of this zinc finger, peptide 5020, has lower affinity ($K_d = 6.01 \pm 0.05 \mu\text{M}$ and $K_d = 19.79 \pm 3.18 \mu\text{M}$, 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 I) were EMSA assayed (Figure 5a and c). The results showed that the partial or complete deletion of NLS (peptides 5016 and 5032) results in a decrease ($K_d = 1.86 \pm 0.02 \mu\text{M}$) and a complete loss of affinity for RNA, respectively, in spite of the 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 NSL domain was present (Figure 5c). The substitution of the CCHH motif for SSHH in peptide 5016, peptide 5031, resulted in a two-fold decrease in binding affinity ($K_d = 3.08 \pm 0.11 \mu\text{M}$) 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 single stranded DNA

A similar approach to that described above was carried out in order to determine the implication of the C2-L1Tc motifs for single stranded DNA binding affinity. Thus, band-shift assays were performed using a constant amount of radiolabeled 135nt-ssDNA and increasing amount of each peptide (Figure 6a-d). To calculate the K_d , the peptide-bound DNA fraction was quantified and plotted *versus* 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 ($K_d = 0.21 \pm 0.06 \mu\text{M}$ and $K_d = 0.30 \pm 0.02 \mu\text{M}$, respectively) (Figure 6e). The presence of a single shifted band has been 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 (Figure 6b and e) 25% of the ssDNA remained unbound (B_{max} : 0.74 ± 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 high concentration of this peptide (Figure 6c).

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

Mapping of C2-L1Tc -binding domains to double stranded DNA

The role of the C2-L1Tc motifs in binding to double stranded DNA was also investigated by incubating 135pb-dsDNA and increasing concentrations of each peptide (Figure 7a-d). The data were analyzed using the Hill equation (Figure 7e). Peptide 10987 that contains the downstream-finger has only a slight affinity for double stranded DNA ($K_d = 25.68 \pm 4.22 \mu\text{M}$). However, the peptides containing the upstream-finger and the complete or partial NSL motif, peptide 5033 and 5016, have high affinity ($K_d = 1.34 \pm 0.08 \mu\text{M}$ and $3.12 \pm 1.01 \mu\text{M}$, respectively). The substitution of CCHH for SSHH in this zinc finger (peptide 5031) resulted in a five-fold decrease in binding affinity (K_d of $9.13 \pm 0.16 \mu\text{M}$), corroborating the important role of this motif in the binding of C2-L1Tc to double stranded DNA. Nevertheless, peptide 5032, which lacks the NSL but maintains the zinc finger, had no affinity for double stranded DNA. Thus, the basic stretches upstream of the zinc fingers (RRR and NLS) also participate in the binding of C2-L1Tc to double strand DNA. In fact, peptide 5015 that contains both regions shows among the assayed peptides high affinity (K_d of $1.51 \pm 0.09 \mu\text{M}$) for double stranded DNA. 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 the both C2H2 zinc-fingers (peptide 5020), bearing also a basic stretch (RRR), showed a low affinity for double stranded DNA. The K_d for this peptide 5020 was estimated to be $22.83 \pm 9.78 \mu\text{M}$.

Remarkably, the concentration dependent binding curves obtained for the binding of each peptide to dsDNA displayed significantly different shapes (Figure 7e). The sigmoidal or hyperbolic shapes of the curves indicate whether the binding is cooperative and non-cooperative, respectively. To obtain the cooperativity value a Hill transformation was applied to the dsDNA-binding data (Figure 7f). The Hill coefficient (α_H) for peptides 5015, 5033 and 5020 containing at least one of the complete basic stretches was 3.34 ± 0.21 , 3.35 ± 0.42 and 2.47 ± 0.48 respectively, indicating a high degree of dsDNA binding cooperativity. However, peptide 10987 containing the downstream-finger binds to dsDNA with low cooperativity ($\alpha_H = 1.29 \pm 0.35$). The binding cooperativity of peptide 5016 that contains the upstream-finger and a fraction of the NLS motif (RKEK) is also low ($\alpha_H = 1.63 \pm 0.16$). The relative low binding cooperativity 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) increases the degree of cooperativity more than two fold ($\alpha_H = 1.63 \pm 0.16$ and $\alpha_H = 3.75 \pm 0.19$, respectively).

Effect on duplex stability of C2-L1Tc protein's motifs

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 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 melting temperature (T_m) of mismatched duplexes, peptides bearing these motifs were tested in melting assays. As expected, peptides 5030 and 5032, lacking both single and double stranded DNA binding capability does not have any influence on the melting temperature of this imperfect DNA duplex (Figure 8 and table I). 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 cysteines by serines in the upstream-finger, peptide 5031, resulted in a reduction on the T_m (47°C and 42°C, respectively *versus* 52°C of the 5033 peptide) although both of them maintain 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) induce a slight decrease of the melting temperature of the duplex containing four internal mismatches (Figure 8a and b).

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 ribonucleoparticle (RNP) 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 retrotransposable 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 nucleic acid chaperone activity (NAC) [22]. The results presented in this paper show that C2-L1Tc exhibits 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 an altered RNPs and to a severe reduction of the retrotransposition efficiency [32].

C2-L1Tc binding profile to RNA corresponds to a non-cooperative binding model. However, the affinity of the C2-L1Tc protein for this type of nucleic acid molecules increases with the RNA size showing a clear non-specific sequence affinity. Consequently, since de L1Tc messenger RNA has a large size (5 kb), it is expected that the C2-L1Tc protein may bind to several positions on the RNA. The 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 C2L1Tc for the RNA molecule is reduced when the RNA molecule is in a denatured state. All together the data presented suggest that the C2-L1Tc protein plays an important role in the binding to the L1Tc transcript and consequently in the ribonucleoparticle formation. Furthermore, the data also suggest that the C2-L1Tc protein-RNA association may prevent L1Tc transcript from degradation.

We believe that the RRR and RRRKEK (NLS) domains as well as the zinc finger located immediately downstream of this 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 cooperativity of these motifs and the relative position of them into 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 indicated that the C2-L1Tc protein endowed with NAC activity stabilizes complementary DNA duplexes and does not modify the

melting temperature (T_m) of mismatched duplexes [22]. In this paper we provide some insights into the molecular mechanisms involved in this activity. In fact, C2-L1Tc has 16-fold higher affinity for single stranded than for double stranded DNA. 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 ribonucleoprotein 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 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 this basic stretches (upstream-finger) also shows a high affinity for both nucleic acids molecules. The substitution of cysteines in that finger for serines 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 carboxyl terminal end of C2-L1Tc protein (downstream-finger) has the ability to bind mainly single stranded DNA. These results indicate the two zinc fingers present into C2-L1Tc have a differential behaviour relative to nucleic acids 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 showed that the two C2H2 zinc fingers and the basic domains located upstream of the first C2H2 cysteine motif are involved in the nucleic acid chaperone (NAC) activity [22]. Our results show that the protein regions responsible for the nucleic acid binding are the same than those previously described to be implicated in NAC activity. However, the data presented suggest (Table I) 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 for 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]. Recent 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 single and double strand DNA 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 single strand DNA, have a destabilization effect. We suggest, therefore, that in order to properly function as a nucleic acid chaperone (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 .

ACKNOWLEDGEMENTS

We thank M. Caro for technical assistance in the purification of C2-L1Tc recombinant protein.

FUNDING

This work was supported by Grants BFU2006-07972 and BFU2007-64999 from Plan Nacional I+D+I (MICINN); P05-CVI-01227 from PAI (Junta de Andalucía) and RD06/0021/0014 from ISCIII-RETIC, Spain. F. Macias was supported by a PAI Predoctoral Fellowship (Grant P05-CVI-01227) (Junta de Andalucía). MEP was supported by a contract from Colciencias (RC-2007).

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, H. H., Jr. (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., Ghedin, E., Peacock, C., Bartholomeu, D. C., Haas, B. J., Tran, A. N., Wortman, J. R., Alsmark, U. C., Angiuoli, S., Anupama, A., Badger, J., Bringaud, F., Cadag, E., Carlton, J. M., Cerqueira, G. C., Creasy, T., Delcher, A. L., Djikeng, A., Embley, T. M., Hauser, C., Ivens, A. C., Kummerfeld, S. K., Pereira-Leal, J. B., Nilsson, D., Peterson, J., Salzberg, S. L., Shallom, J., Silva, J. C., Sundaram, J., Westenberger, S., White, O., Melville, S. E., Donelson, J. E., Andersson, B., Stuart, K. D. and Hall, N. (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* nonlong 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., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D. A., Carrington, M., Crabtree, J., Darban, H., da Silveira, J. F., de Jong, P., Edwards, K., Englund, P. T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A. C., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M. J., Lorenzi, H., Louie, T., Machado, C. R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J. C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J. L., Rinta, J., Robertson, L., Salzberg, S. L., Sanchez, D. O., Seyler, A., Sharma, R., Shetty, J., Simpson, A. J., Sisk, E., Tammi, M. T., Tarleton, R., Teixeira, S., Van Aken, S., Vogt, C., Ward, P. N., Wickstead, B., Wortman, J., White, O., Fraser, C. M., Stuart, K. D. and Andersson, B. (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**, 45621-45626.
- 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., Cumiskey, J., Branciforte, D. and Williams, M. C. (2008) A single amino acid substitution in ORF1 dramatically decreases L1 retrotransposition 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 5 S 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 interaction is a major component of HIV-1 nucleocapsid protein's nucleic 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 Henriot, 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.

TABLES AND FIGURES

Figure 1: Structure of L1Tc and sequence of the recombinant C2-L1Tc protein. Purification of the recombinant protein. (a) The single open reading frame of the L1Tc element is represented by a rectangular box. The endonuclease (EN), reverse transcriptase (RT), RNase H (RH), and nucleic acid chaperone (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 [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 this study are underlined and numbered. The two CCHH zinc finger domains are in bold, the RRR basic region and the nuclear localization signal (NLS) are in bold and labeled with (*) and (=), respectively. (b) *Escherichia coli* TAP-F strain (active motif) transformed with pCAS-C2L1Tc vector was grown in SOB medium. Total protein extracts from non-induced (1) and induced (2) cultures were resolved in 12% SDS polyacrilamide gels. C2-L1Tc recombinant protein was purified to homogeneity by Ni²⁺ affinity chromatography and gel filtration using a Superdex 75 column (3).

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 ³²P-labeled *in vitro* transcript 130nt-RNA (0.5 ng) was incubated with increasing concentration of C2-L1Tc (15-730 x10⁻³ μM). (b) **Single-**

stranded DNA binding. 0.5 nM single-stranded DNA (~4 ng) (135nt-ssDNA) was incubated with increasing amounts of C2-L1Tc ($19\text{-}2810 \times 10^{-3} \mu\text{M}$). **(c) Double stranded DNA binding.** 1nM double-stranded DNA (~8 ng) (135bp-dsDNA) was incubated with increasing amounts of C2-L1Tc ($19\text{-}2810 \times 10^{-3} \mu\text{M}$). **(d) C2-L1Tc affinity for RNA, ssDNA and dsDNA.** In the figures a, b and c, a Hill transformation was applied to the data obtained from three independent electrophoresis mobility shift assays. The average of the three $\log \frac{Y}{1-Y}$ values were in all cases plotted *versus* the log of C2-L1Tc concentration, where Y is the bound 130nt-RNA (●), 135nt-ssDNA (○) and the 135bp-dsDNA (▼) fraction ($R^2=0.99$, $R^2=0.91$ 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 $\frac{Y}{1-Y}=1$ and calculated from the average of the three values in a, b, and c. The position of the “complexes” and “free” nucleic acids is indicated, respectively, on the right hand (arrow heads) and on the left hand (F) of panels.

Figure 3. Influence of the secondary structure in the binding capacity of C2-L1Tc to RNA. (a) EMSA to measure the binding capacity of C2-L1Tc to native and denatured 130nt-RNA. 0.72 nM native 130nt-RNA (panel 130nt-RNA) or denaturalized 130nt-RNA (panel 130nt-denatured-RNA) fragments were incubated with increasing concentrations of C2-L1Tc (95 to $560 \times 10^{-3} \mu\text{M}$). 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.** They were obtained from the average of three independent band-shift assays as those shown in Figure 3a. The average values corresponding to the fraction of radiolabeled 130nt-RNA (○) and 130nt-denatured-RNA (●) bound to C2-L1Tc are plotted versus the protein concentration. The curves correspond to the best fit of the Hill equation to the experimental data ($R^2=0.99$ and $R^2=0.99$, respectively). The equation used is as follows:

$$y = \frac{B_{\max} \cdot x^{\alpha_H}}{K_d^{\alpha_H} + x^{\alpha_H}}, \text{ where } x \text{ is the C2-L1Tc concentration and } y \text{ the fraction of the bound}$$

radiolabeled RNA. B_{\max} , the maximum value for y , is roughly 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 acids free forms is also indicated (F).

Figure 4. Effect of the nucleic acid length in the C2-L1Tc nucleic acid binding capacity. Titration curves were obtained by EMSA assays. 0.5 ng of ^{32}P -labeled *in vitro* transcribed 130nt-RNA (a) was incubated with increasing amounts of unlabeled competitors: 130nt-denatured-RNA (●) and 2,100nt-denatured-RNA (○). 0.5 ng of ^{32}P -labeled *in vitro* transcribed 144nt-RNA (b) was incubated with increasing amounts of unlabeled competitors: 130bp-dsDNA (●) and 2.1kb-dsDNA (○). The samples were further incubated with $0.67 \mu\text{M}$ of C2-L1Tc. The average values corresponding to the fraction of bound radiolabeled 130nt-RNA (a) or 144nt-RNA (b) from three independent experiments are plotted versus the concentration of the nucleic acid employed as competitor. The curves correspond to the best fit of a four-parameter logistic equation of the experimental data ($R^2=0.99$, $R^2=0.99$, $R^2=0.99$ and $R^2=0.97$, respectively). The equation is as follows:

$$y = B_{\min} + \frac{B_{\max} - B_{\min}}{1 + 10^{(\log EC_{50-x}) \cdot \alpha_H}}, \text{ where } x \text{ is the competitor concentration and } y \text{ the fraction of}$$

bound-radiolabeled RNA. The 50% effective concentration (EC_{50}) defined as the concentration of unlabeled 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 cooperativity, are indicated.

Figure 5. Binding analysis of C2-L1Tc-derived peptides to RNA by EMSA. 0.72 nM of *in vitro* transcribed ^{32}P -labeled 130nt-RNA were preincubated 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 I 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 panels of figure 5a, 5b and 5c. The average values corresponding to the bound RNA fraction were plotted *versus* the peptide concentration. The curves correspond to the best fit of the Hill equation to the experimental data ($R^2 \geq 0.99$). The equation runs as follows: $y = \frac{B_{\max} \cdot x^{\alpha_H}}{K_d^{\alpha_H} + x^{\alpha_H}}$, where the x is the

peptide concentration and y is the radiolabeled RNA bound fraction. 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 of panels as an arrow head and F, respectively,

Figure 6. Binding analysis of C2-L1Tc-derived peptides to single stranded DNA by EMSA. 0.5 nM of ^{32}P -labeled 135nt-ssDNA was preincubated with the indicated increasing concentrations of peptides 5015, 5033 (a); 5020, 10987 (b); 5032, 5030 (c); 5016 and 5031 (d); at 37°C for 30 min. (see Figure 1 and table I 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 labeled with an arrow head and 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 data were obtained from three independent experiments as those shown in panel a, b, c and d. The average values corresponding to the bound ssDNA fraction were plotted *versus* peptide concentration. The curves correspond to the best fit of the Hill equation to the experimental data ($R^2 \geq 0.98$).

The equation runs as follows: $y = \frac{B_{\max} \cdot x^{\alpha_H}}{K_d^{\alpha_H} + x^{\alpha_H}}$, where the x is the peptide concentration and y is the radiolabeled RNA bound fraction. 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 double stranded DNA by EMSA. 0.5 nM of ^{32}P -labeled 135bp-dsDNA were preincubated 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 the table I for sequence composition details), at 37°C for 30 min. Control reactions were performed without peptide addition (no peptide). Reactions were loaded on 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 data were obtained from three independent experiments as those shown in panels a, b, c and d. The average values corresponding to the bound dsDNA fraction were plotted *versus* the peptide concentration. The curves correspond to the best fit of the Hill equation to the experimental data ($R^2 \geq 0.97$). The equation runs as follows: $y = \frac{B_{\max} \cdot x^{\alpha_H}}{K_d^{\alpha_H} + x^{\alpha_H}}$, where the x is the peptide concentration and y is the radiolabeled dsDNA bound fraction. K_d , defined as the peptide concentration at which 50% of the dsDNA is bound, is indicated in the panel e. (f) A Hill transformation was applied to the data obtained from three independent electrophoresis mobility shift assays. The $\log \frac{Y}{1-Y}$ average values were plotted *versus* the log of C2-L1Tc concentration, where Y is the bound 135bp-dsDNA fraction. The thin lines correspond to the best fit determined by linear regression ($R^2 \geq 0.93$). The slope of the best-fit equation determines the Hill coefficient (α_H) and indicates the degree of cooperativity. The dotted line is the theoretical binding curve for a non-cooperative interaction. K_d of C2-L1Tc peptide 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. The vertical lines at the right hand of the gels indicate the shifted bands generated in each case. The mobility of the dsDNA free form is also indicated (F).

Figure 8. Effect of C2-L1Tc-derived peptides on the melting temperature of a preformed mismatched dsDNA duplex. (a) Phosphorimages of dried gels show the ^{32}P -labeled DNA products formed after incubation for five minutes of 1 mM 29 nt DNA duplex containing four mismatches (^{32}P -29mer/mm29c) with 1 μM C2-L1Tc-derived peptides or without 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 non-effect on the melting temperature. The ^{32}P -29mer/mm29c duplex (ds) and the ^{32}P -29mer oligonucleotide (ss) are indicated on the right hand of panels. (b) Fraction of unbound ^{32}P -29mer oligonucleotide *versus* temperature. The variation in melting temperatures 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 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 carboxyl terminal end of the C2-L1Tc protein (named downstream-finger); 5020, peptide covering the region located downstream of this zinc finger.

TABLE I. 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 labeled with an asterisk (*). Dissociation constants, K_d , 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 (Figures 5,6 and 7). The Hill formalism used is described by Henriët 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 (fig. 2d and 7f). The dependent variable “y” value is 0 when $\frac{Y}{1-Y} = 1$. ^(b)Increase (+) or decrease (-) in temperature relative to the melting temperature 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. *Standard deviation, mean \pm SD*, is indicated. *Conc50* represents the peptide concentration required to reach the formation of 50% of stable duplexes as a measure of NAC activity [22].

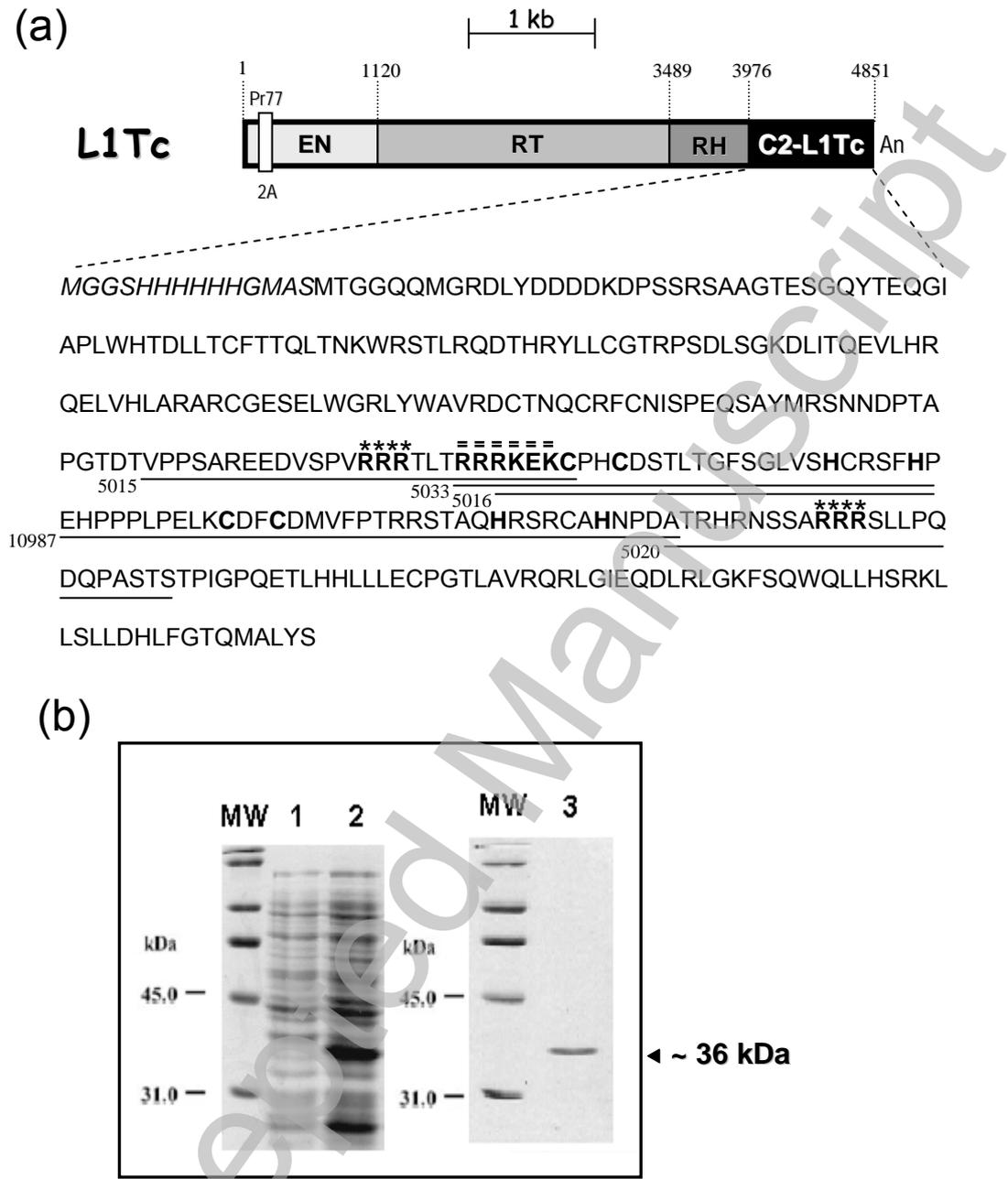


Figure 1

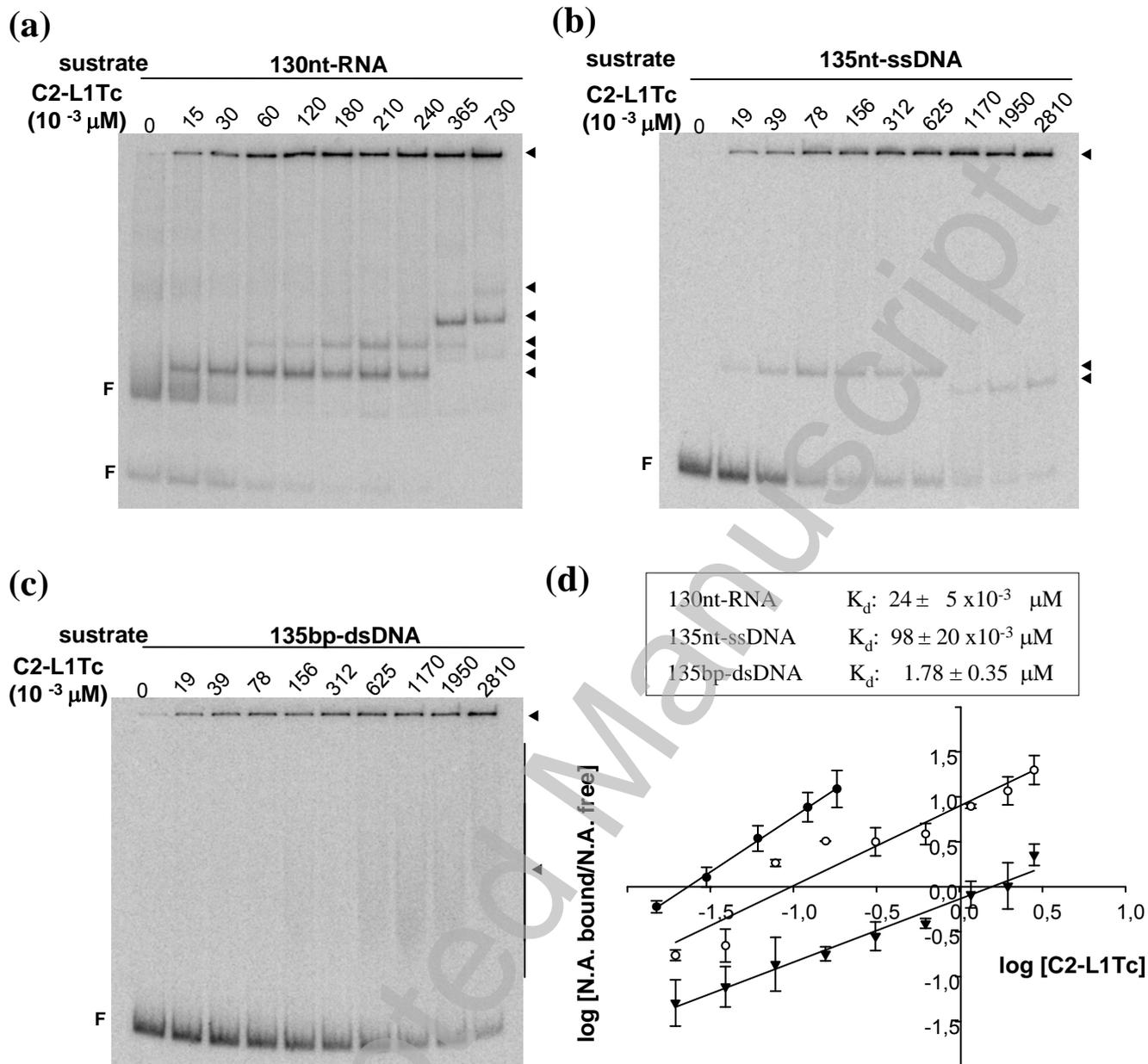
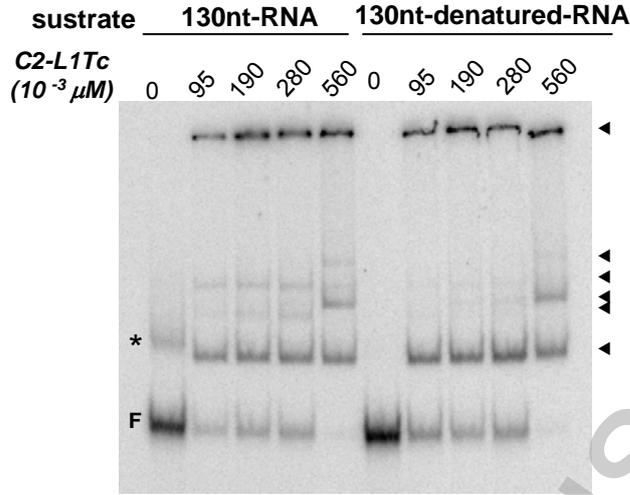


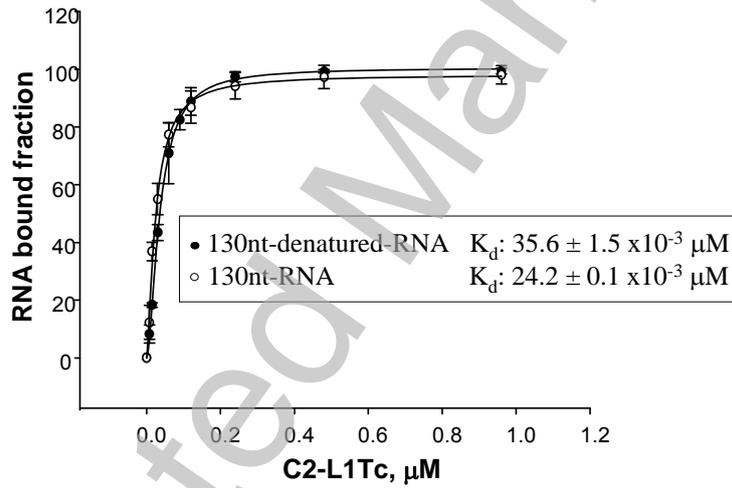
Figure 2

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20090766

(a)



(b)



THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20090766

Figure 3

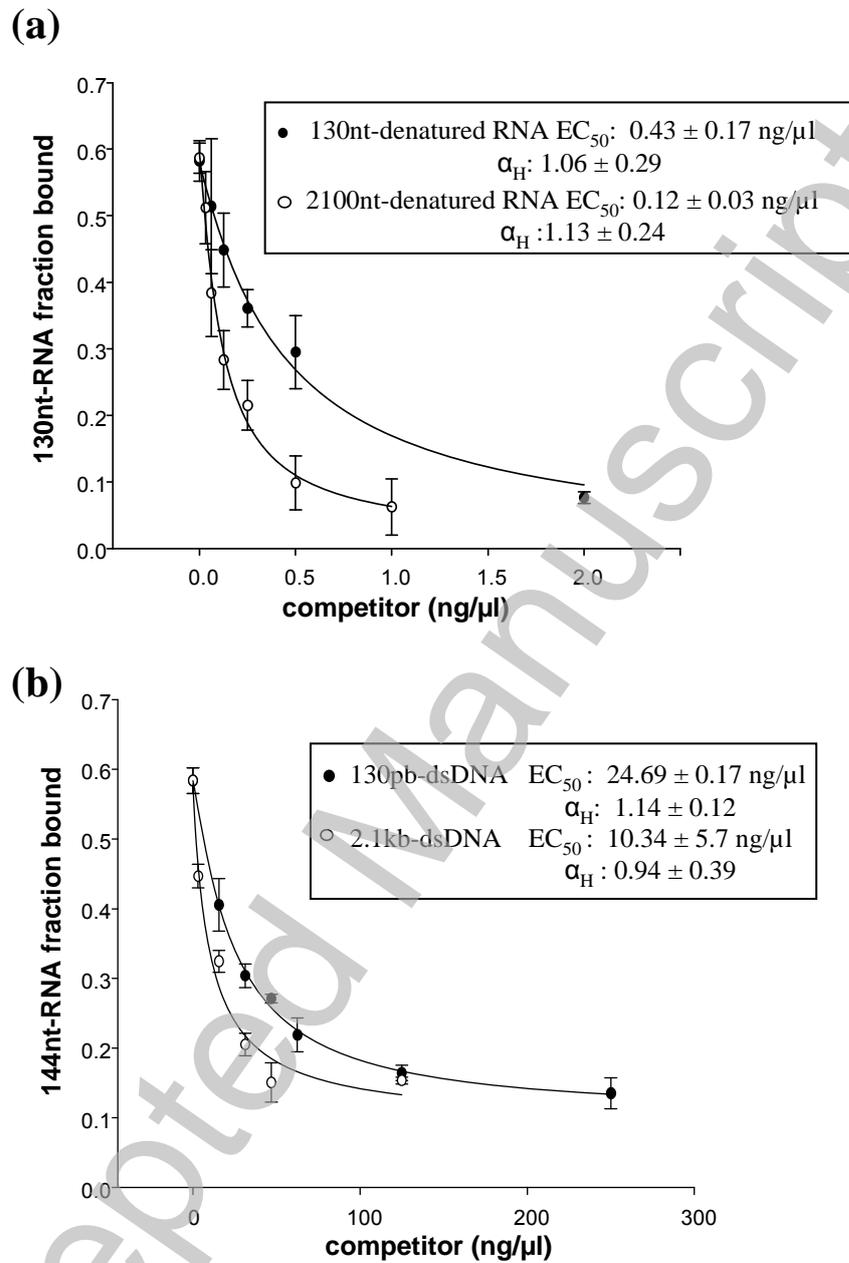


Figure 4

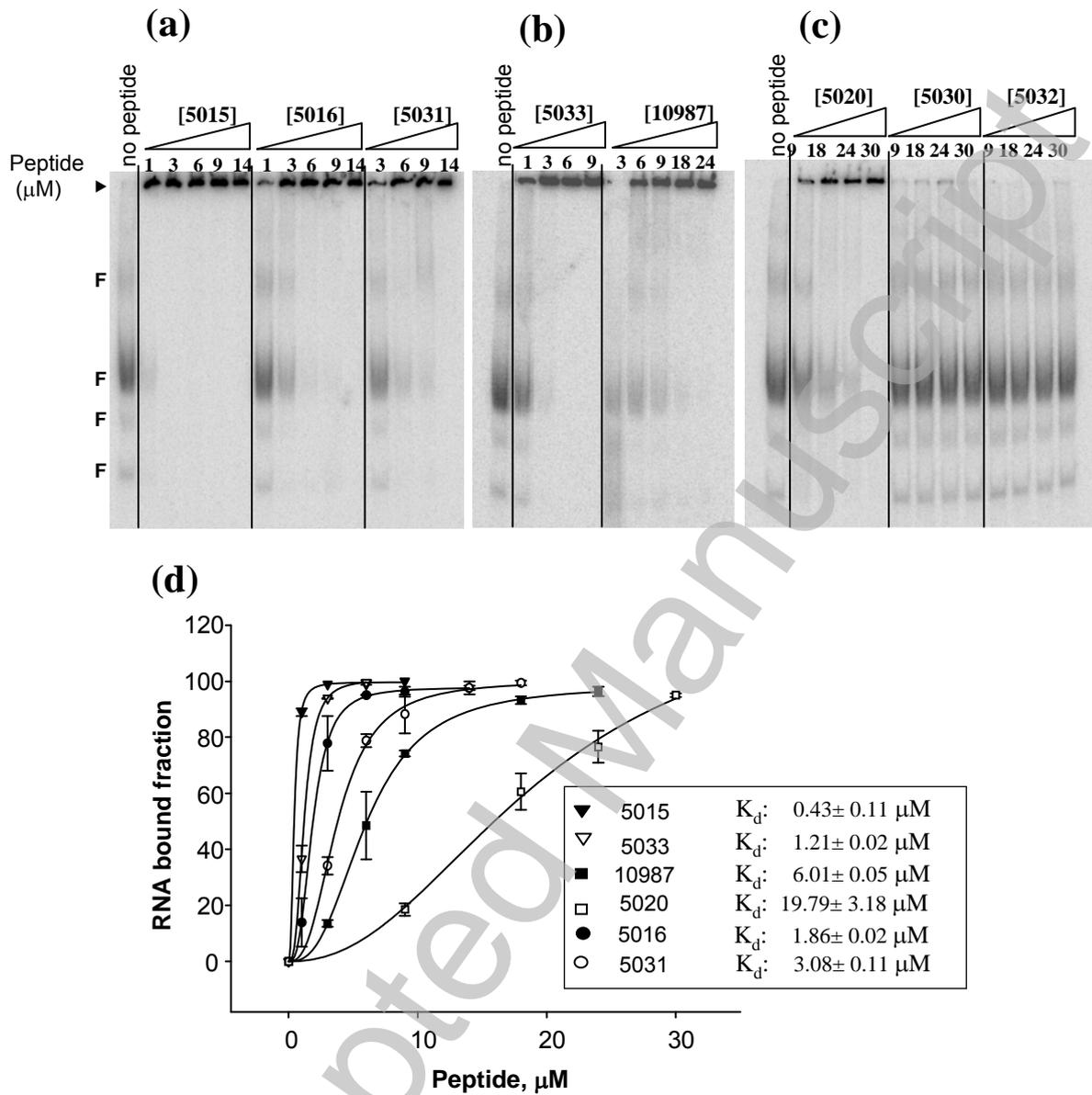


Figure 5

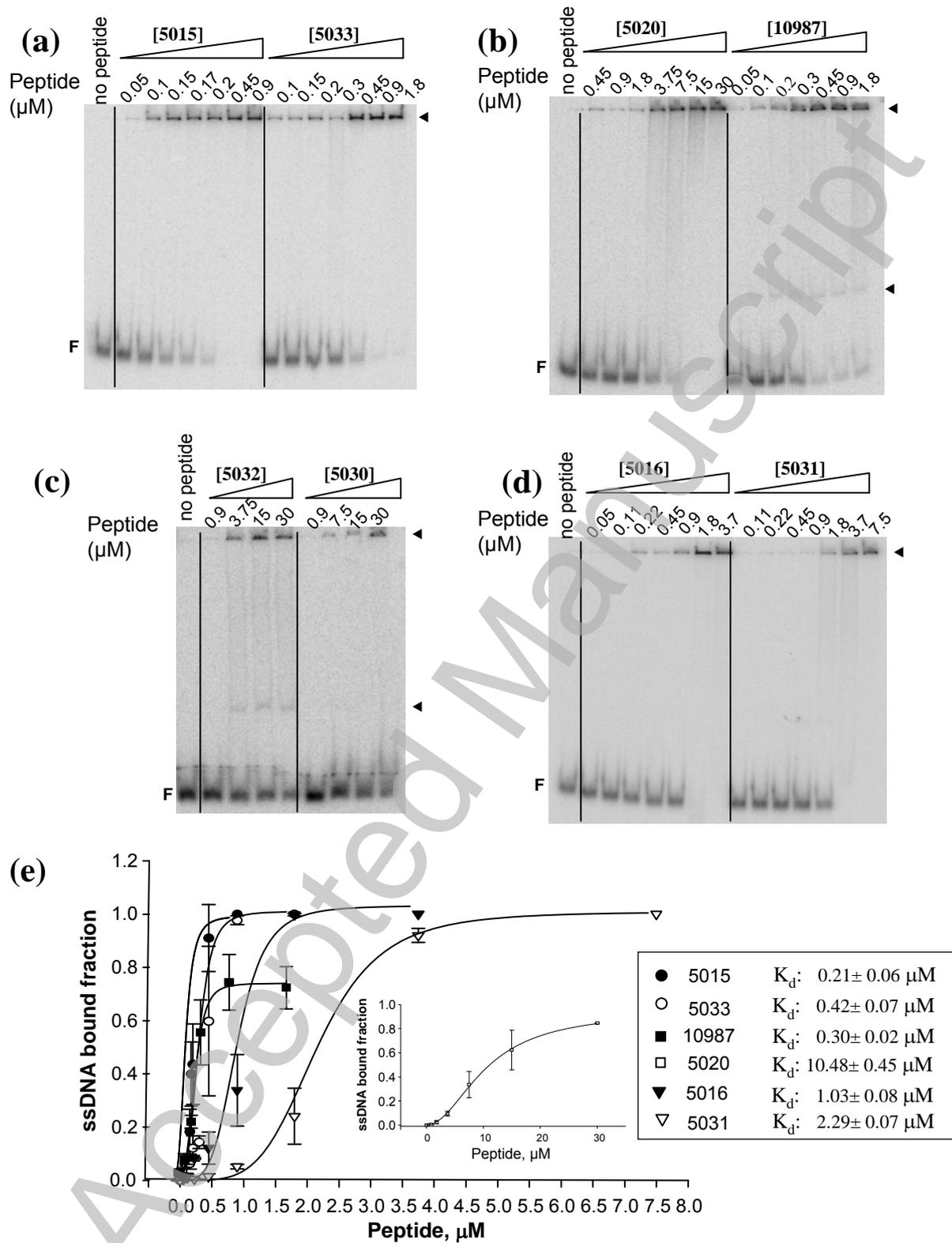


Figure 6

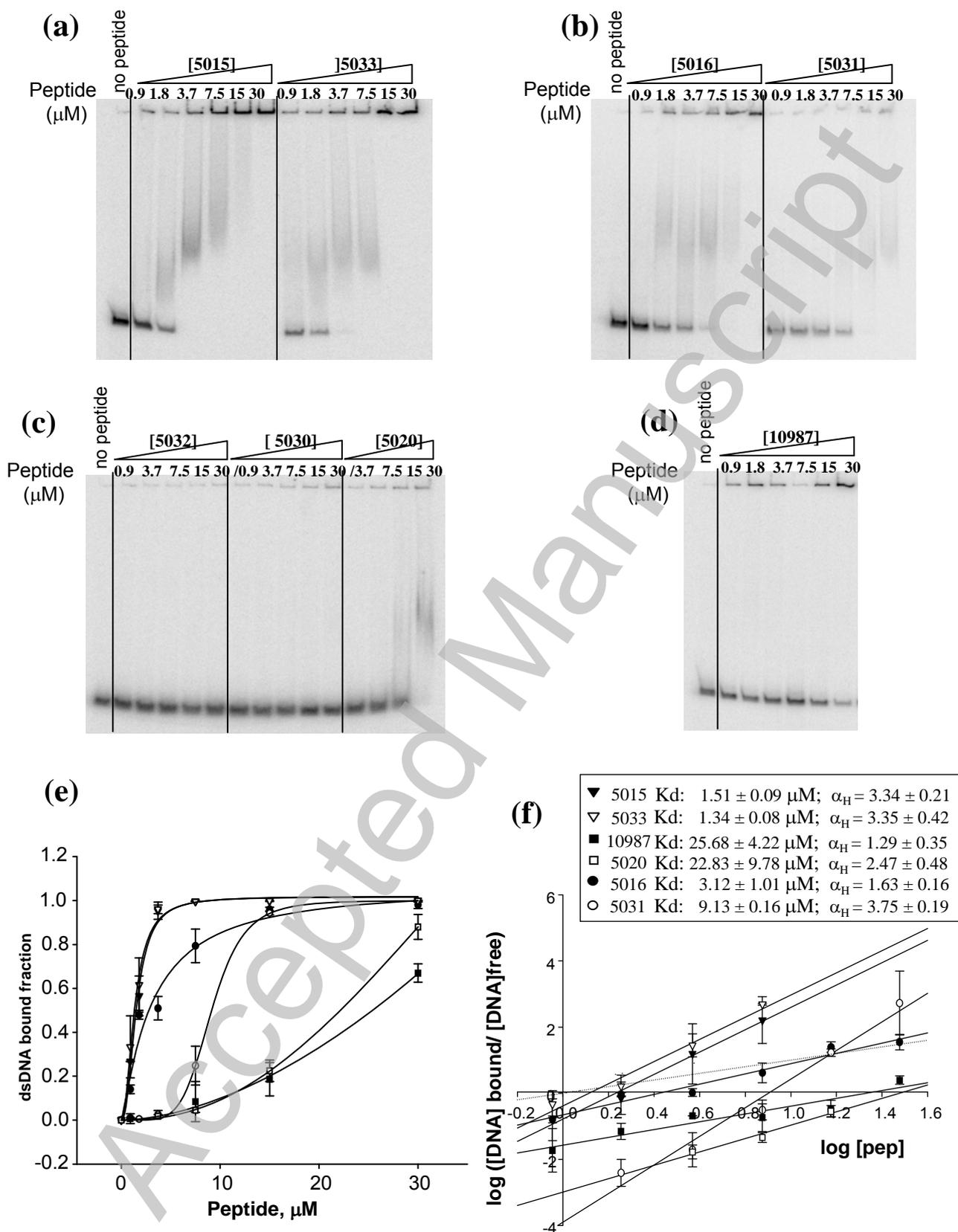


Figure 7

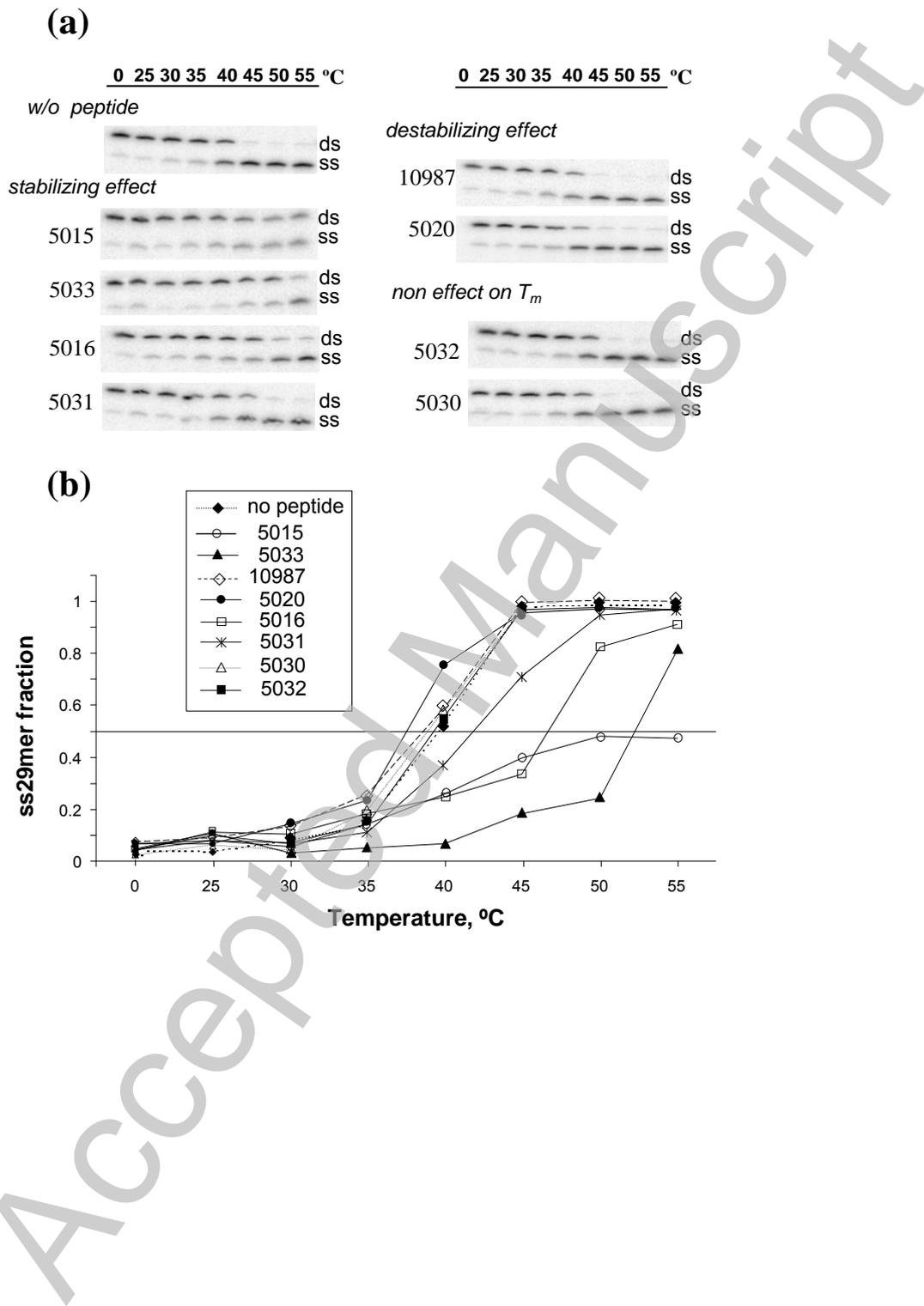


Figure 8

Table I.

<i>Peptide</i>	<i>Sequence</i>	<i>RNA-K_d</i> (μM)	<i>ssDNA-K_d</i> (μM)	<i>dsDNA-K_d</i> (μM)	ΔT_m ^(b) ($^{\circ}\text{C}$)	<i>Conc50</i> (μM)
C2-L1Tc		$24^{(a)} \pm 5 \times 10^{-3}$	$98 \pm 20 \times 10^{-3}$	1.78 ± 0.35	0	0.07
5033	RRRKEK C_PHC_DSTLTGFSGLVSH_CRSFHP	1.21 ± 0.02	0.42 ± 0.07	1.34 ± 0.08	+12	0.4
5015	TVPPSAREEDVSPV RRRTLLRRRKEK C	0.43 ± 0.11	0.21 ± 0.06	1.51 ± 0.09	<i>no melted</i>	0.51
5016	RKEK C_PHC_DSTLTGFSGLVSH_CRSFHP	1.86 ± 0.02	1.03 ± 0.08	3.12 ± 1.01	+7	1.5
5031*	RKEK S_PHS_DSTLTGFSGLVSH_CRSFHP	3.08 ± 0.11	2.29 ± 0.07	9.13 ± 0.16	+2	2.4
10987	EHPPPLPELK_CDF_CDMVFP TRR STAQHRSRCA_HNP	6.01 ± 0.05	0.30 ± 0.02	25.68 ± 4.22	-1	2.9
5020	ATRRNSSA RRR SLLPQDQPAST	19.79 ± 3.18	10.48 ± 0.45	22.83 ± 9.78	-2	NO
5030*	TVPPSAREEDVSPV.....TL TRRKEK C	NO	NO	NO	0	NO
5032*	LT.....C_PHC_DSTLTGFSGLVSH_CRSFHP	NO	NO	NO	0	NO

Accepted Manuscript