© 1993 Oxford University Press

Nucleic Acids Research, 1993, Vol. 21, No. 4 823–829

Two short basic sequences surrounding the zinc finger of nucleocapsid protein NCp10 of Moloney murine leukemia virus are critical for RNA annealing activity

H.De Rocquigny, D.Ficheux, C.Gabus[1], B.Allain[1], M.C.Fournie-Zaluski, J.-L.Darlix[1] and B.P.Roques*

Unite de Pharmacochimie moléculaire et structurale, U266 INSERM, URA D1500 CNRS, Université René Descartes, 4, Avenue de l'Observatoire, 75270 Paris cedex 06 and [1]Laboretro INSERM, Ecole Normale Supérieure de Lyon, 46, allée d'Italie, 69364 Lyon, France

Received December 17, 1992; Revised and Accepted January 22, 1993

ABSTRACT

The 56 amino acid nucleocapsid protein (NCp10) of Moloney Murine Leukemia Virus, contains a $CysX_2CysX_4HisX_4Cys$ zinc finger flanked by basic residues. *In vitro* NCp10 promotes genomic RNA dimerization, a process most probably linked to genomic RNA packaging, and replication primer $tRNA^{Pro}$ annealing to the initiation site of reverse transcription. To characterize the amino-acid sequences involved in the various functions of NCp10, we have synthesized by solid phase method the native protein and a series of derived peptides shortened at the N- or C-terminus with or without the zinc finger domain. In the latter case, the two parts of the protein were linked by a Glycine – Glycine spacer. The *in vitro* studies of these peptides show that nucleic acid annealing activities of NCp10 do not require a zinc finger but are critically dependent on the presence of specific sequences located on each side of the CCHC domain and containing proline and basic residues. Thus, deletion of ^{11}R or ^{49}PRPQT, of the fully active 29 residue peptide ^{11}RQGGERRRSQLDRDGGKKPRGPRG-PRPQT^{53} leads to a complete loss of NCp10 activity. Therefore it is proposed that in NCp10, the zinc finger directs the spatial recognition of the target RNAs by the basic domains surrounding the zinc finger.

INTRODUCTION

A unique feature of retroviruses is that their genome is diploid and formed of two identical RNA molecules linked together by the Dimer Linkage Structure (DLS)(1–4). In the virion capsid the diploid genome is present as a 60S complex tightly associated with nucleocapsid (NC) protein molecules (2, 5–7). NC proteins such as the 56 amino acids NCp10 from Moloney Murine Leukemia Virus (MoMuLV) or NCp7 from Human Immunodeficiency Virus (HIV) are small basic proteins containing one (MoMuLV) or two (HIV) copies of the sequence $CysX_2CysX_4HisX_4Cys$ highly conserved among retroviruses and designated the CCHC box or finger domain (2, 8). The zinc finger has been shown to complex metallic ions with a stoechiometry of one cation per motif, the highest affinity being for Zn^{2+} ions with a Kapp of 10^{13} M^{-1} (9–11). In addition, structural analyses of NC protein fragments containing the zinc finger indicated that Zn^{2+} complexation causes a folding of the peptide backbone (12–14). Mutations of the cysteine and histine residues implicated in Zn^{2+} complexation result in a drastic decrease in Zn^{2+} binding affinity and an unfolding of the zinc finger (10–11).

The retroviral NC protein was shown to activate viral RNA dimerization leading to formation of the DLS (2, 7, 15–17), which in the case of Moloney Murine Leukemia Virus (MoMuLV) is probably located around position 300–350 (4, 7). Moreover viral RNA dimerization requires *cis* elements located between positions 200 to 400 from the RNA 5′ end (7). *In vivo* viral RNA dimerization is most probably associated with genomic RNA packaging since both processes require the same *cis* elements and NC protein (2, 7, 17). Furthermore NC proteins were shown to direct the annealing of replication primer tRNAs ($tRNA^{Pro}$ in MoMuLV) to the primer binding site (PBS) which is necessary for the initiation of reverse transcription (15, 16). Site directed mutations of cysteine or histidine implicated in zinc coordination, or of amino-acids within the zinc finger (18–22), were shown to induce a defect in genomic RNA packaging. However in synthetic peptides such mutations did not affect the ability of NCp10 to generate dimeric RNA or to promote the annealing of primer $tRNA^{Pro}$ to the PBS *in vitro* (17). These findings strongly suggest that the zinc finger is not directly involved in the RNA hybridization activity of NCp10 *in vitro*.

Therefore it was of interest to investigate the functions of the basic domains surrounding the zinc finger of NCp10 since all retroviral NC proteins, like many other nucleic acid binding proteins (23), contain rather well conserved sequences rich in basic and proline residues. For this purpose, we have synthesized MoMuLV NCp10 and a series of derived peptides shortened at

* To whom correspondence should be addressed

the N- and/or C-terminus and with or without the zinc finger motif. In the latter case the two fragments of NCp10 were linked by a glycine-glycine spacer. Here we report that the RNA annealing activity of NCp10 *in vitro* does not require the zinc finger but is critically dependent upon the domains rich in basic and proline residues surrounding the finger motif.

MATERIALS AND METHODS

Materials

Fmoc protected amino acids, 4-hydroxymethylphenoxymethylpolystyrene (HMP) resin, piperidine, N-methylpyrrolidone (NMP), dichloromethane, 4-dimethylaminopyridine (DMAP), dicyclohexylcarbodiimide (DCC) and 1-hydroxybenzotriazole (HOBt) were purchased from Applied Biosystems. Trifluoroacetic acid (TFA) was from Neosystem Laboratory, and 1.2 ethanedithiol (EDT), dithiothreitol (DTT) and phenol from Aldrich.

Synthesis of MoMuLV NCp10 and derived peptides

Assembly of the protected peptide chains was carried out using the stepwise solid phase method of Merrifield (24) using 9-fluorenylmethyloxycarbonyl (Fmoc)-chemistry on an Applied Biosystems 431 A peptide synthesizer.

An unloaded HMP support was used with the 0.1 mmole small scale cycle. The C-terminal residue (leucyl, glycyl, lysyl or threonyl) was attached to the resin with DCC in the presence of 4-dimethylaminopyridine. The subsequent Fmoc amino acids (1 mmole, 10 eq/resin) were incorporated using DCC/HOBt reagents.

Side chain protection was N^G-(2,2,5,7,8-pentamethylchroman-6-sulfonyl) (Pmc) for Arg, N^ϵ-t-butyloxycarbonyl (Boc) for Lys, trityl (trt) for Gln, Asn, His and Cys, t-butyl (tBu) for Ser, Thr and Tyr and t-butyl ester (OtBu) for Glu and Asp.

For the large peptides, after the 25th coupling step, half of the resin was removed both to reduce the peptide-resin volume and to increase the amount of Fmoc amino acid. Moreover, the standard program of the synthesizer was modified so as to mix the reactor during the transfer of activated amino acids allowing a good swelling of the resin and to progressively increase the coupling time as a function of the size of the growing peptide.

Deprotection of the Fmoc group was obtained by two successive treatments of 3 min and 15 min with 20% piperidine in N-methylpyrrolidone.

TFA cleavage

The peptide (HMP) resin was treated for two hours with TFA (82.5%), phenol (5%), EDT (2.5%), thioanisole (5%) and H_2O (5%) to remove protecting groups and to cleave the peptide from the resin. At the end of the reaction, the TFA solution was filtered through a medium porosity filter and concentrated under vacuum on a rotavapor. The peptide was then precipitated with 15 ml of cold ether and centrifuged for ten minutes at 2000 g at room temperature. The supernatant containing the scavengers was removed and the precipitate resuspended in 0.1% TFA to be lyophilized in a Savant Speed vac concentration.

HPLC purification

For NCp10 and all the peptides containing unprotected cysteine residues, the crude peptides were dissolved in 2 ml of Tris−HCl buffer, pH 7.5, containing 50 eq of DTT and incubated for 1 h at 37°C. Then the subsequent purification was performed under an argon atmosphere. Analytical and semipreparative purifications were carried out by reverse phase liquid chromatography on an Applied Biosystems 151A apparatus with a Nucleosil RP 100 C_5 column 4.6×150 mm (5 μm, 100Å) and Vydac C_4 column 10×220 mm (5 μm, 300 Å), respectively.

The mobile phase consisted of solution A (0.1% TFA in H_2O) and solution B (70% CH_3CN, 0.09% TFA in H_2O). The pure fractions of preparative runs were pooled and lyophilized under vacuum using a Savant speed vac concentrator.

Amino acid analysis and mass spectroscopy

For amino acid analysis, peptides were hydrolyzed for 2 h at 100°C in 6N HCl in evacuated sealed tubes. For a better recovery of the tyrosine, 1% (w/v) phenol was added to the acid solution before hydrolysis. Cysteine residues were determined as cysteic acid after oxydation of a separate sample with performic acid. The analyses were performed on a system 6300 Beckman apparatus (Palo Alto, CA, USA). Molecular mass was obtained by 252 Cf-plasma desorption mass spectroscopy on a Bio-Ion-20 time-of-flight instrument.

Thiol titration

Cysteine containing peptides were dissolved in freshly degazed Hepes (N-(2-Hydroxyethyl)Piperazine-N′-2-Ethane sulfonic acid) 50 mM (pH 7.5), KCl 100 mM buffer and immediately put into anaerobic quartz cells which maintain an inert argon atmosphere. The SH content of NCp10 and the various peptides, checked by titration at 412 nm with 5.5′ dithiobis-(2-nitrobenzoic acid), was always greater than 2.7 suggesting that more than 90% of the SH groups were in a reduced state.

Analysis of NC protein activity *in vitro*

Plasmid construction. Standard procedures were used for restriction nuclease digestion and plasmid DNA construction (25). *Escherichia coli* HB 101 (1035) was used for plasmid DNA amplification.

Plasmid pMLVSI-4 was used for *in vitro* synthesis of MoMuLV RNA (1−725) by bacteriophage T7 RNA polymerase (see ref. 9 for details). Plasmid ptRNABA-1 allows the 'in vitro' synthesis of murine $tRNA^{Pro}$ (26). Restriction nuclease digestion of ptRNABA-1 with BstNI and subsequent transcription of the template allowed the synthesis of a complete $tRNA^{Pro}$.

Proteins and enzymes. Restriction nucleases and bacteriophage RNA polymerases were from Promega. MuLV reverse transcriptase was from Biolab.

In vitro generated RNA. 5 μg of linearized plasmid DNA were transcribed in 0.1 ml of 40 mM Tris−HCl (pH 7.5), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 10 mM dithiothreitol, 0.5 mM of each ribonucleotide triphosphate with 40 units T_7 RNA polymerase and 50 units of ribonuclease inhibitor RNasin for 2 h at 40°C. To generate ^{32}P-labelled RNA, the transcription was performed in the presence of 60 mM (^{32}P) UTP at 20 Ci/mmol. Following DNase treatment the RNA was phenol extracted twice, chromatographed over Sephadex G75, precipitated with ethanol and dissolved in sterile double distilled water. ^{32}P-$tRNA^{Pro}$ was further purified by urea/PAGE on a 8% gel. ^{32}P-$tRNA^{Pro}$ present in the gel slice was recovered by diffusion, precipitated with ethanol and dissolved in sterile double distilled water. Before use synthetic $tRNA^{Pro}$ was heat denatured and slowly cooled. 5′ ^{32}P labelling of bovine $tRNA^{Pro}$ was carried out as previously described (15).

Ultraviolet light (UV) cross-linking experiments with NC proteins and MoMuLV RNA or primer tRNA^{Pro}

UV (252 nm) cross-linking experiments were carried out as described before with HIV and RSV (2, 5). Preparation of ^{32}P-labelled MoMuLV RNA (1−725) and replication primer tRNAPro is described above. Assays were in 10 μl containing 25 mM Tris-Cl (pH 7.5), 60 mM NaCl, 0.2 mM $MgCl_2$, 5 mM DTT, either 300 ng of ^{32}P-MoMuLV RNA or 50 ng ^{32}P-tRNA^{Pro} and NCp10 or NC derived peptides (25 to 100 ng). After 5 min at 37°C the complexes were irradiated for 3 to 5 min at 20°C using conditions reported before (5). The NC-tRNA complexes were then incubated in 5 mM EDTA and 1% SDS for 2 min at 80°C and analysed by 13% polyacrylamide gel electrophoresis (SDS-PAGE). After UV irradiation the MoMuLV RNA-NC complexes were digested for 30 min at 20°C with 5 units of T_1 RNase, then incubated in 5 mM EDTA and 1% SDS for 2 min at 80°C and analysed by 15% SDS-PAGE. Following electrophoresis, gels were autoradiographed for 2 to 12 hours.

MoMuLV RNA dimerization and primer tRNA^{Pro} annealing by NC protein

Nucleocapsid protein assays were performed in 10 μl reactions comprising 20 mM Tris−HCl (pH 8), 30 mM NaCl, 0.2 mM $MgCl_2$, 10 mM $ZnCl_2$, 5 mM dithiothreitol, 5 units RNasin, 0.5 μg of *in vitro* generated MoMuLV RNA (denatured 1 min at 95°C prior to use), where indicated 10 to 20 ng ^{32}P-tRNA^{Pro}, and 20 to 200 ng NCp10 for 10 min at 37°C. Reactions were terminated by the addition of SDS (0.5%, final concentration), the samples extracted first with phenol (8 μl of phenol saturated with 50 mM Tris−HCl pH 7.0, 5 mM EDTA) and then with 8 μl of chloroform and the RNA analysed by agarose gel electrophoresis. To analyse the extent of viral RNA dimerization, viral RNA was analysed by electrophoresis on a 1% (w/v) agarose gel in 50 mM Tris-borate (pH 8.3), 0.1 mM EDTA at 5 V/cm. Gels were washed with water and the RNA was visualized by ethidium bromide staining (1 μg/ml for 5 min). For the analysis of primer ^{32}P-tRNA^{Pro} annealing to the PBS by autoradiographic analysis, the gels were fixed with 5% (w/v) trichloroacetic acid and dried.

Reverse transcription

In vitro reverse transcription of MoMuLV RNA (1−725) was performed using primer ^{32}P-tRNA^{Pro} synthesized *in vitro*, or 5′ ^{32}P-tRNA^{Pro} from beef liver. Reactions were performed in 10 μl comprising 40 mM Tris−HCl (pH 8.3), 60 mM NaCl, 2 mM $MgCl_2$, 5 mM dithiothreitol, 0.5 μg *in vitro* generated RNA, 10 ng primer ^{32}P-tRNA^{Pro} (first annealed to the PBS using NCp10 or derived peptides), 0.2 mM of each deoxyribonucleotide triphosphate, and 0.2 mg MuLV reverse transcriptase. Samples were incubated for 15 min at 37°C and terminated by the addition of 1% SDS 10 mM EDTA. The samples were then treated with phenol, extracted with chloroform and heat denatured for 2 min at 95°C before analysis. The ^{32}P-cDNAs were analysed on an 8% PAGE containing 7 M urea and 50 mM Tris-borate (pH8.3). After electrophoresis, the gel was autoradiographed for 2 to 6 h at −80°C.

RESULTS

Synthesis of NCp10 and related peptides

In an attempt to investigate the role of the zinc fingers and flanking residues on RNA binding and annealing activities of MoMuLV NC protein, we have synthesized the native NCp10 and four peptides containing the zinc finger (Figure 1): (11−48) NCp10, **NCp10A**; (14−48)NCp10, **NCp10B**; (1−42)NCp10, **NCp10C** and (24−56)NCp10, **NCp10D**. Furthermore six zinc finger deleted peptides containing a Gly-Gly spacer linking the D^{24} and K^{41} residues were prepared: (1−56)NCp10, **NCp10E**; (6−53)NCp10, **NCp10F**, (11−53)NCp10, **NCp10G**; (11−48) NCp10, **NCp10H**, (12−48)NCp10, **NCp10I** and (14−53) NCp10, **NCp10J**.

All peptides were prepared by solid phase synthesis using Fmoc chemistry. The synthesis was performed in a single step with a coupling time of 20 min except for the large peptides (>30 amino acids) where the coupling time was increased to 30 min as a function of the size of the growing peptide (27).

The deprotection and cleavage steps of the peptides from the resin were performed using the TFA-scavengers standard procedure and peptides were precipitated with cold Et_2O at the

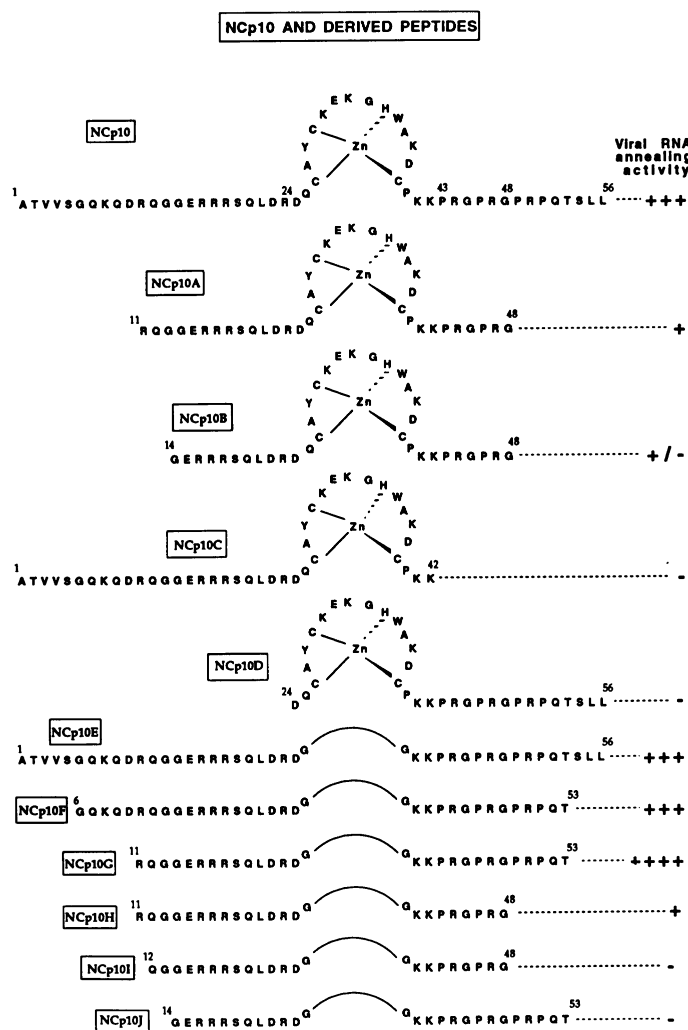

Figure 1. Primary sequence and RNA annealing activity of NCp10 and NCp10 derived peptides: MoMuLV RNA(1−725) dimerization and primer tRNA^{Pro} annealing to the PBS carried out by NCp10 and NCp10 derived peptides were examined using the conditions described in materials and methods (see also figures 3 and 4). Activity of each peptide is given on the right panel: −, no activity; +/−, 10% of RNA dimer and no primer tRNA annealed; +, 10−20% of RNA dimer and annealing of tRNA was about 5% of that with NCp10; +++, 80−100% RNA dimer and tRNA annealed.

Table 1. Amino acid content (uncorrected values) of purified peptides after 6N HCl hydrolysis (see Experimental Procedures)

Amino Acid	NCp10	A	B	C	D	E	F	G	H	I	J
Asx	3.9(4)	3.05(3)	3.03(3)	4.49(4)	2.00(2)	3.04(3)	2.89(3)	1.95(2)	1.91(2)	1.90(2)	2.00(2)
Thr	1.8(2)			0.80(1)	1.00(1)	1.67(2)	0.83(1)	1.00(1)			0.90 (1)
Ser	2.6(3)	0.75(1)	0.70(1)	1.48(2)	0.86(1)	2.33(3)	0.75(1)	0.95(1)	0.83(1)	0.82(1)	1.20(1)
Glx	8.1(8)	4.91(5)	3.78(4)	6.97(7)	2.88(3)	5.69(6)	5.86(6)	3.75(4)	3.00(3)	3.10(3)	3.40(3)
Pro	5.3(5)	3.20(3)	3.08(3)	1.3(1)	4.76(5)	3.71(4)	3.79(4)	4.05(4)	1.94(2)	2.00(2)	4.20(4)
Gly	5.6(6)	4.95(5)	3.85(4)	3.96(4)	2.83(3)	6.59(7)	6.95(7)	5.90(6)	5.82(6)	5.84(6)	5.20(5)
Ala	3.0(3)	2.00(2)	2.00(2)	3.00(3)	1.90(2)	1.00(1)					
Val	1.9(2)			1.52(2)		1.75(2)					
Leu	3.2(3)	1.00(1)	0.91(1)	1.05(1)	1.92(2)	2.80(3)	0.95(1)	0.98(1)	1.00(1)	1.00(1)	1.10(1)
Tyr	0.9(1)	0.95(1)	0.92(1)	1.02(1)	0.90(1)						
Lys	6.1(6)	4.92(5)	4.78(5)	6.04(6)	4.40(5)	2.77(3)	3.02(3)	1.90(2)	2.00(2)	2.00(2)	2.00(2)
His	1.1(1)	1.03(1)	1.02(1)	1.11(1)	1.00(1)						
Arg	8.2(8)	6.85(7)	5.44(6)	5.04(5)	2.81(3)	7.43(8)	8.20(8)	7.56(8)	6.70(7)	5.79(6)	5.75(7)
Cys	3.0(3)	2.96(3)	2.95(3)	2.90(3)	2.99(3)						

The theoritical amino acid composition of each peptide is given in parentheses. Presence of tryptophan was measured by U.V. spectroscopy.

end of the reaction time. For peptides containing cysteine residues, the crude peptide was treated with DTT, purified in an inert atmosphere and lyophilized in a Savant speed vacuum concentrator to reduce possible cystine formation. Purifications were performed by high pressure liquid chromatography and amino acid compositions (table I) and masses verified. All subsequent biochemical analyses were performed with NCp10 and NCp10 derivatives that were at least 98% pure.

Binding of NCp10 and derived peptides to primer $tRNA^{Pro}$

Since ultraviolet light (UV) links proteins to nucleic acids when reactive groups are no more than 0.1 nm apart (28) we used UV cross-linking to analyse the interactions between NC protein and either primer $tRNA^{Pro}$ or MoMuLV RNA (1–725). NCp10 or a derived peptide was incubated with ^{32}P labelled RNA and the complexes formed were UV cross-linked and analysed by polyacrylamide gel electrophoresis in SDS (PAGE-SDS). As reported in figure 2, NCp10 and NCp10 derivatives without the finger (NCp10F and NCp10G) interacted with primer $tRNA^{Pro}$ to form NC–tRNA complexes (lanes 2,3 ; lanes 5, 6 and 8, 9). Although the yield of cross linking does not necessarly reflect the stability of the complex between NC peptides and MoMuLV RNA, it is interesting to observe that there is an apparent correlation between the annealing activity of the studied peptides and their ability (see figs. 3 and 4 above) to form ribonucleoprotein complex. The apparent molecular weight of about 29 Kda for the major NCp10-tRNA complexes indicates that they probably correspond to one NCp10 molecule per tRNA. NCp10B (lanes 11, 12) appears to interact less tightly with primer $tRNA^{Pro}$, while NCp10J, NCp10C (lanes 13–16) and other peptides did not show significant tight RNA interactions under these *in vitro* conditions. UV cross-linking of NCp10 and derived

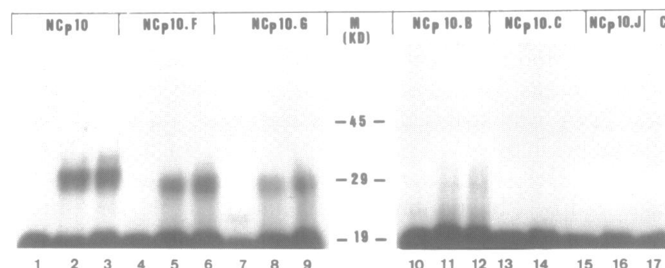

Figure 2. UV cross-linking analysis of the interactions of NCp10 and derivatives with primer ^{32}P $tRNA^{Pro}$. Lanes 1 and 17: primer tRNA + UV; lanes 2 and 3: 25 and 50 ng NCp10 with UV irradiation; lane 4: 100 ng NCp10F minus UV irradiation; lanes 5 and 6: 50 and 100 ng NCp10F with UV irradiation; lane 7: 100 ng NCp10G minus UV irradiation; lanes 8 and 9: 50 and 100 ng NCp10G with UV irradiation; lane 10: 200 ng NCp10B minus UV irradiation; lanes 11 and 12: 100 and 200 ng NCp10B with UV irradiation; lanes 13 and 14: 100 and 200 ng NCp10C with UV irradiation; lanes 15 and 16: 100 and 200 ng NCp10J with UV irradiation. Autoradiography was for 2 h. Note that the major NC protein-tRNA complex migrated with an apparent molecular weight of about 29 kDa and ^{32}P-$tRNA^{Pro}$ with an apparent molecular weight of 19 kDa.

peptides to MoMuLV RNA (1–725) was also carried out, and the results obtained were very close to these obtained with primer $tRNA^{Pro}$ (not shown).

MoMuLV RNA dimerization by NCp10 and derived peptides

MoMuLV RNA dimerization was examined upon incubation of the viral RNA with NCp10 and derived peptides under optimal conditions for NCp10 activity (7, 29). After 20 min at 37°C, reactions were stopped with SDS and the RNAs extracted with phenol to remove NC protein and analysed by agarose gel electrophoresis under native conditions. Figure 3A reports that

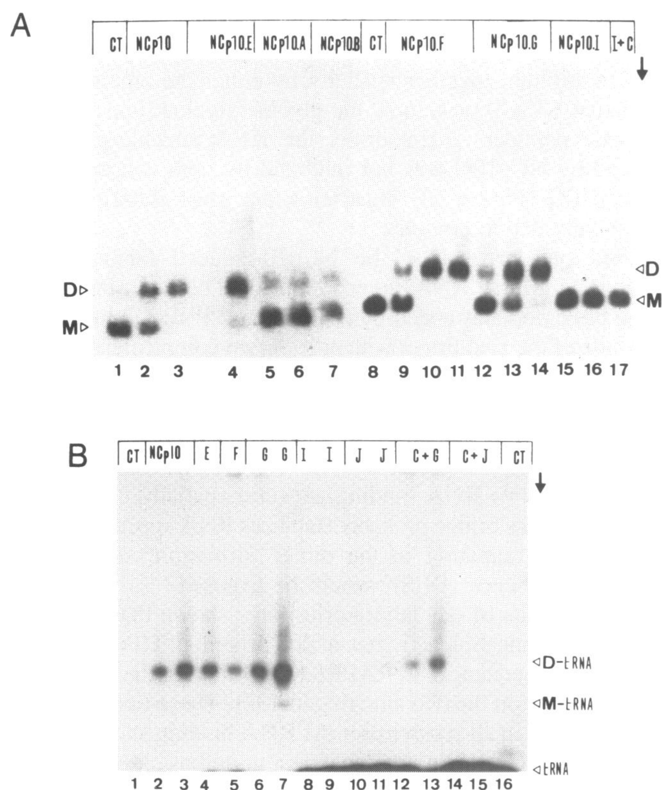

Figure 3. Viral RNA annealing activity of NCp10 and derived peptides: **A/** Dimerization of MoMuLV RNA(1−725) was carried out using the conditions described in materials and methods. Lane 1: no protein; lanes 2−3: 40 and 80 ng NCp10; lane 4: 80 ng NCp10E; lanes 5, 6: 100, and 250 ng NCp10A; lane 7: 500 ng NCp10B; lane 8: no protein; lanes 9, 10 and 11: 25, 50 and 100 ng NCp10F; lanes 12, 13 and 14: 25, 50 and 100 ng NCp10G; lanes 15, 16: 200 and 100 ng NCp10I; lane 17: 100 ng NCp10I and then 100 ng NCp10C. **B/** Annealing of primer ^{32}P-tRNA^{Pro} to the PBS. Lane 1: no protein; lanes 2, 3: 40 and 80 ng NCp10; lane 4: 80 ng NCp10E; lane 5: 50 ng NCp10F; lanes 6, 7: 50 and 100 ng NCp10G; lanes 8, 9: 100 and 200 ng NCp10I; lanes 10, 11: 100 and 200 ng NCp10J; lanes 12, 13: 100 ng NCp10C and then 25 and 50 ng NCp10G; lanes 14, 15: 100 ng NCp10C and then 100 and 200 ng NCp10J; lane 16: no protein. M and D correspond to MoMuLV monomeric and dimeric RNA, and M-tRNA and D-tRNA to MoMuLV monomeric and dimeric RNA with primer ^{32}P-tRNAPro annealed to the PBS. Molecular weight markers were rat 18S and 28S rRNAs (not shown). Samples 1 to 7, and 8 to 17 in figure 3A were run on two different gels. Samples 1 to 16 in figure 3B were run on the same gel.

dimerization of MoMuLV RNA (1−725) did not occur in absence of NC protein (lane 1) under these *in vitro* conditions. Upon addition of NCp10, RNA dimer formation was observed in a dose dependent manner (figure 3A, lanes 2,3). The finger deleted peptides, NCp10 E, F and G were also able to promote viral RNA dimerization (fig. 3A, lanes 4,9−11 and 12−14). However, NCp10A and B are active only at a high protein concentration (100 ng) and did not dimerize more than 10% of MoMuLV RNA in the presence of 500 ng protein while NCp10I is completly inactive (lanes 15, 16). Additional experiments have been carried out with NCp10 derived peptides D, H and J (see figure 1) and results obtained were all negative with NCp10D and J, and with NCp10H very similar to those obtained with NCp10A. Moreover mixing two inactive peptides like NCp10 C and D, did not restore the RNA annealing activity *in vitro* (data not shown).

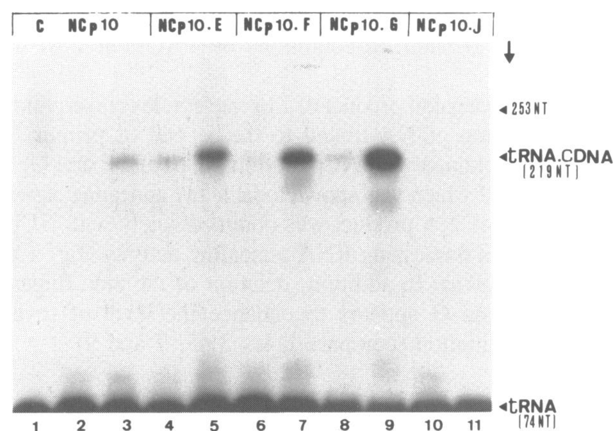

Figure 4. Synthesis of MuLV strong stop cDNA by MuLV RT in the presence of NCp10 and derived peptides. At the end of the reaction leading to the annealing of primer tRNA^{Pro} to the PBS of MoMuLV RNA, 0.2 mg of MoMuLV reverse transcriptase (RT) was added together with 0.25 mM of each deoxyribonucleotide triphosphate. Analysis of the extended cDNA product was as described in methods. Lane 1: control with MoMuLV RNA (1−725), ^{32}P-tRNA^{Pro} and MoMuLV reverse transcriptase; lanes 2 and 3: 25 and 50 ng NCp10; lanes 4 and 5: 25 and 50 ng NCp10E; lanes 6 and 7: 25 and 50 ng NCp10F; lanes 8 and 9: 25 and 50 ng NCp10G; lanes 10 and 11: 25 and 50 ng NCp10J.

Replication primer tRNA^{Pro} annealing to the initiation site of reverse transcription (PBS) of MoMuLV RNA by NCp10 and derived peptides

To investigate the annealing of primer tRNA^{Pro} to the PBS of MoMuLV RNA we used both 'in vitro' generated ^{32}P-tRNA^{Pro} and beef liver tRNA^{Pro} labelled at its 5′ end (15). Results obtained with synthetic tRNA^{Pro} and NCp10 were almost identical to those already reported with natural tRNA^{Pro} (9, 15 and data not shown). Thus, most of the experiments have been carried out with synthetic ^{32}P-tRNA^{Pro}. NCp10 strongly activated annealing of primer ^{32}P-tRNA^{Pro} to the dimeric viral RNA (Fig. 3B, lanes 2, 3). The zinc finger deleted peptides, NCp10E, F and G were also able to promote primer ^{32}P-tRNA^{Pro} annealing to the PBS (Fig. 3B, lanes 4, 5, 6, 7, for NCp10E, F and G, respectively). However, NCp10I and J alone or in combination with NCp10C were unable to promote the annealing of ^{32}P-tRNA^{Pro} to the PBS (lanes 8, 9, 10, 11 and 14, 15, for NCp10I, J and C plus J, respectively). Interestingly, NCp10C (100 ng) was unable to inhibit viral RNA dimerization and primer tRNA^{Pro} annealing to the PBS promoted by NCp10G (lanes 12, 25 ng and lane 13, 50 ng, compare with lane 6). All the other NCp10 peptides such as NCp10B, D and H were found to be very poorly active under these 'in vitro' conditions using either natural or synthetic tRNA^{Pro}.

Initiation of reverse transcription promoted by NCp10 and derived peptides

The ability of NCp10 and NCp10 derived peptides to promote the initiation of reverse transcription was investigated as follows: first, NCp10 or a derived peptide was incubated with MoMuLV RNA (1−725) and primer ^{32}P-tRNA^{Pro} to allow annealing of primer tRNA to the PBS and the formation of high molecular weight complexes. Second, deoxyribonucleotide triphosphates and MoMuLV reverse transcriptase were added and reverse transcription allowed to proceed for 15 min. The strong stop

cDNA (ss-cDNA) product linked to primer ^{32}P-tRNA^{Pro} was analysed under denaturing conditions by PAGE in 7 M urea (figure 4).

The correct extension product of 219 nucleotides corresponding to the strong stop cDNA linked to the 3′ end of primer ^{32}P-tRNA^{Pro} was obtained with NCp10, and NCp10E, F and G, but not with NCp10J which was shown to lack any annealing activity. The ^{32}P-tRNA-cDNA product was obtained solely with NCp10 derived peptides possessing RNA annealing activity (fig. 1 and 3) (data not shown). In addition, deletion of the zinc finger in NCp10 E, F and G appears to enhance the level of reverse transcription initiation (compare lanes 3, 5, 7 and 9).

DISCUSSION

The aim of this study was to investigate the role of the zinc finger in the RNA annealing activity of MoMuLV NC protein *in vitro*, since it has been shown that the finger motif is required for the specific packaging of the diploid genome *in vivo* (18,20,22). Furthermore, a functional analysis of the basic domains surrounding the zinc finger was carried out since these regions are highly conserved among the retroviral NC proteins and little attention has been paid to them (2, 31 and references herein). To determine the smallest NC peptide retaining the full activity of NCp10 *in vitro*, peptides with deletions in the N- and/or C-terminal ends were chemically synthesized. To evaluate the role of the zinc finger, peptides with a glycine-glycine linker replacing the finger motif and with N- and/or C-terminal deletions were prepared. The glycine-glycine spacer does not exactly mimic the distance between D^{24} and K^{41} found in the structure of the native protein determined by NMR spectroscopy (data to be published), but was selected in order to link in a flexible way the domains flanking the zinc finger (see figure 1).

All the NCp10 peptides synthesized were assayed for their biological activity *in vitro* which includes their tight interactions with primer tRNA^{Pro} (figure 2), and their ability to dimerize MoMuLV RNA containing the Psi/dimer sequence (figure 3A) and to anneal primer tRNA^{Pro} to the PBS (figure 3B), resulting in the initiation of reverse transcription upon addition of reverse transcriptase and nucleotides (figure 4).

The complementary role of the amino-acids surrounding the zinc finger is illustrated by the fact that NCp10C and NCp10D are completely inactive. The partial activity of NCp10A and NCp10B indicates that a minimal number of basic amino acids located on both sides of the zinc finger is required.

The four NCp10 derived peptides (NCp10E to NCp10G) lacking the finger motif were found to be as active as the native protein *in vitro*. This agrees with the activity of modified NCp10 where cysteines involved in zinc coordination have been replaced by serines or chemically modified (17). In addition these data indicate that amino acids of the finger motif do not directly participate in the RNA annealing activity of NCp10. The minimal peptide retaining full activity contains the N-terminal RQGGERRRSQLDRD and the C-terminal KKPRGPRGPRPQT of NCp10, linked by a G-G spacer (NCp10G). The most interesting findings are that arginine at position 11 and the proline rich sequence ^{49}PRPQT are critically involved in the biologically relevant RNA annealing activities of NC protein *in vitro*. In addition the two minimal basic domains surrounding the finger motif must be linked for NC protein activity since a mixture of peptides corresponding to the N and C domains was still inactive (compare NCp10 I and C, NCp10C and G, figure 3A, 3B). This underlines the fact that the two basic domains, which are far away from each other in the primary sequence, must be brought together spatially by either the zinc finger in the native NCp10 protein or the glycine-glycine linkage in the synthetic peptides. Interestingly the RNA annealing activity induced by NCp10G was not inhibited by high concentrations of NCp10C (Figure 3B) illustrating the great stability of the biologically active complex.

Based on our results, the NCp10-induced facilitation of genomic MoMuLV RNA dimerization could occur by interaction of the basic domains ended by (^{11}RQG and ^{49}PRPQT) located on each side of the zinc finger with at least two complementary non-continuous RNA sequences. The lack of annealing activity of NCp10I and of a mixture of terminal ends (NCp10C and NCp10D) suggests that the native protein or the active zinc-finger deleted peptides recognizes and stabilizes a particular structure in which the two RNA binding sites are spatially close . The formation of this bridge probably stabilizes RNA species in which sequences participating in the dimer formation, such as the 283−300 sequence (DLS), would be exposed.

Recent results of our laboratories have shown that the RNA binding and annealing activities of NC protein of HIV-1 (NCp7) rely on basic sequences (^{29}RAPRKKG^{35}) linking the two finger motifs and not on the two zinc fingers (32). These findings favor the notion that in all retroviruses the RNA binding and annealing activities of NC protein are dependent upon basic domains rich in proline and glycine residues which flank the zinc finger(s) (e.g. ^{29}RAPRKKG^{35} in HIV-1, ^{35}PKKRKSG^{41} in RSV, ^{41}KKPRGP-RG^{48} in MoMuLV, ^{47}KRPRKKP^{53} in FeLV and PRPSRGRGR in the human spumaretrovirus HSRV) (30−33 and references herein). Site-directed mutagenesis of these NC protein basic domains are presently being performed and results indicate that basic residues flanking the zinc finger of MoMuLV genomic RNA packaging and virus infectivity *in vivo* (Housset, De Rocquigny, Roques and Darlix, submitted for publication). This is in agreement with a previous report on RSV demonstrating that mutation of the two lysines next to the first zinc finger of RSV NC protein (^{36}KK) leads to a severe defect in genomic RNA packaging and infectivity (33).

In the present study the question of how the genomic RNA is recognized amongst the cellular RNAs has not been addressed. Both the encapsidation element Psi and the NC protein are required for this recognition *in vivo*. Specific interactions between NC protein and the viral Psi/dimer element could be the first step towards packaging of the genome. Genetic and biochemical data favor the idea that the NC protein zinc finger and the flanking basic residues cooperate to direct the selective recognition of the genomic Psi$^+$ RNA in the infected cell.

Finally, the observation that short basic domains are essential for the RNA binding and annealing activities of NC proteins should facilitate the rational design of compounds aimed at inhibiting retrovirus replication.

ACKNOWLEDGEMENTS

Thanks are due to G.Keith (Strasbourg) for a gift of beef liver tRNA^{Pro}, Y.Mely for thiol titration, M.Lapadat-Tapolsky for a critical reading of the manuscript, A.Beaumont for helpful discussions and stylistic revision of the paper and C.Dupuis for her assistance in preparing it. This work was supported by grants from the french association against AIDS (ANRS).

REFERENCES

1. Coffin, J.M. (1984) in RNA Tumor Viruses (Weiss, R., Teich, N., Varmus, H., Coffin, J.M., eds) vol I, pp 261–368, Cold Spring Harbor Laboratory press, Cold Spring Harbor, NY.
2. Darlix, J.-L., Gabus, C., Nugeyre, M.-T., Clavel, F. & Barré-Sinoussi, F. (1990) *J. Mol. Biol.* **216**, 689–699.
3. Bender, W. ,Chien, Y.H., Chattopadhyay, S., Vogt, P.K., Gardner, M.B. & Davidson, N. (1978). J. of Virology, **25**, 888–896.
4. Murti, K.G., Bondurant, M. & Tereba, A. (1981) *J. Virol.* **37**, 411–419.
5. Darlix, J.L. & Spahr, P.F. (1982) *J. Mol. Biol.* **160**, 147–161.
6. Méric, C., Darlix, J.L. & Spahr, P.F. (1984) *J. Mol. Biol.* **173**, 531–538.
7. Prats, A.C., Roy, C., Wang, P., Erard, M., Housset, V., Gabus, C., Paoletti, C. & Darlix, J.L. (1990) *J. Virol.* **64**, 774–783.
8. Berg, J.M. (1986) *Science* **232**, 485–487.
9. Cornille, F., Mely, Y., Ficheux, D., Salvignol, I., Gerard, D., Darlix, J.-L., Fournié-Zaluski, M.-C. & Roques, B.P. (1990) *Int. J. Peptide Protein Res.* **36**, 551–558.
10. Green, L.M. & Berg, J.M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6403–6407.
11. Mély, Y., Cornille, F., Fournié-Zaluski, M.C., Darlix, J.L., Roques, B.P. & Gérard, D. (1991) *Biopolymers* **31**, 899–906.
12. South, T.L., Blake, P.R., Have, D.R. & Summers, M.F. (1991) *Biochemistry* **30**, 6342–6349.
13. Morellet, N., Jullian, N., De Rocquigny, H., Maigret, B., Darlix, J.L. & Roques, B.P. (1992) *EMBO J.*, **11**(8), 3059–3065.
14. Roberts, W.J., Pan, T., Elliot, J.I., Coleman, J.E., Williams, K.R., (1989), *Biochemistry*, **28**, 10043–10047.
15. Prats, A.C., Sarih, L., Gabus, C., Litvak, S., Keith, G. & Darlix, J.L. (1988) *EMBO J.* **7**, 1777–1783.
16. Barat, C., Lullien, V., Keith, G., Nugeyre, M.T., Grüninger-Leitch, Barré-Sinoussi, F., LeGrice, S.F.J. & Darlix, J.L.(1989) *EMBO J.* **8**, 3279–3285.
17. Prats, A.C., Housset, V. , de Billy, G., Cornille, F., Prats, H., Roques, B. & Darlix, J.L. (1991) *Nucleic Acids Res.* **13**, 3533–3541.
18. Gorelick, R. J., Henderson, L. E. , Hanser J. P. & Rein A. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8420–8424.
19. Gorelick, R.J., Stephen, M.N., Bess, J.W.Jr., Arthur, L.O., Henderson, L.E. & Rein, A. (1990) *J. of Virol.* **64**, 3207–3211.
20. Aldovini, A. & Young, R.A. (1990) *J. Virol.* **64**, 1920–1926.
21. Méric, C. & Goff, S. J.(1989) *J. Virol.* **63**(4), 1558–1568.
22. Méric, C. & Spahr, P.-F.(1986) *J. Virol.* **60**, 450–459.
23. Reeves, R. & Nissen, M.S. (1990) *J. Biol. Chem.* **265**, 8573–8582.
24. Barany, G. and Merrifield, R.B., (1979), in The Peptides (Gross, E. and Meienhofer, J., eds), Vol 2, pp. 1–284, Academic Press, New York.
25. Maniatis, T., Fritsch, E. & Sambrook, I. (1982) in Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
26. Darlix, J.L., Gabus, C. and Allain, B. 1992, *J. Virol.*, in press.
27. De Rocquigny, H., Ficheux, D., Gabus, C., Fournié-Zaluski, M.C., Darlix, J.L. and Roques, B.P., (1991), *Biochem. Biophys. Res. Commun.* **180**, 1010–1018.
28. Havron, A. and Sperling, J. (1977). *Biochemistry* **16**, 5631–5635.
29. Roy, C., Tounekti, N., Mougel, M., Darlix, J.L., Paoletti, C., Ereshmann, C., Ereshmann, B. & Paoletti, J. (1990) *Nucleic Acid. Res.* **18**, 7287–7292.
30. Bieth, E., Gabus, C. and Darlix, J.L. (1990) *Nucleic Acids Res.* **18**, 119–127.
31. Maurer, B., Bannert, H., Darai, G. and Flügel, R. (1988) *J. Virol.* **62**, 1590–1597.
32. De Rocquigny, H., Gabus, C., Vincent, A., Fournié-Zaluski, M.C., Roques, B.P. & Darlix, J.L. (1992) *Proc. Natl. Acad. Sci.* USA **89**, 6472–6476.
33. Fu, X., Katz, R.A., Skalka, A.M. and Leis, J. (1988) *J. Biol. Chem.* **263**, 2140–2145.