The EMBO Journal vol.13 no.7 pp.1525-1533, 1994

Specific binding of HIV-1 nucleocapsid protein to PSI RNA *in vitro* requires N-terminal zinc finger and flanking basic amino acid residues

Jens Dannull, Andrej Surovoy[1], Gunther Jung[1] and Karin Moelling[2,3]

Max-Planck-Institut für Molekulare Genetik, Abteilung Schuster, Ihnestrasse 73, D-14195 Berlin and [1]Institut für Organische Chemie, Universität Tübingen, Auf der Morgenstelle 18, D-72076 Tübingen, Germany

[2]Present address: Institut für Medizinische Virologie, Universität Zürich, CH-8028 Zürich, Switzerland

[3]Corresponding author

Communicated by K.Moelling

The nucleocapsid (NC) protein of human immunodeficiency virus HIV-1 (NCp7) is responsible for packaging the viral RNA by recognizing a packaging site (PSI) on the viral RNA genome. NCp7 is a molecule of 55 amino acids containing two zinc fingers, with only the first one being highly conserved among retroviruses. The first zinc finger is flanked by two basic amino acid clusters. Here we demonstrate that chemically synthesized NCp7 specifically binds to viral RNA containing the PSI using competitive filter binding assays. Deletion of the PSI from the RNA abrogates this effect. The 35 N-terminal amino acids of NCp7, comprising the first zinc finger, are sufficient for specific RNA binding. Chemically synthesized mutants of the first zinc finger demonstrate that the amino acid residues C-C-C/H-C/H are required for specific RNA binding and zinc coordination. Amino acid residues F16 and T24, but not K20, E21 and G22, located within this zinc finger, are essential for specific RNA binding as well. The second zinc finger cannot replace the first one. Furthermore, mutations in the basic amino acid residues flanking the first zinc finger demonstrate that R3, 7, 10, 29 and 32 but not K11, 14, 33 and 34 are also essential for specific binding. Specific binding to viral RNA is also observed with recombinant NCp15 and Pr55Gag. The results demonstrate for the first time specific interaction of a retroviral NC protein with its PSI RNA *in vitro*.

Key words: HIV-1/NC protein/specific RNA binding/zinc finger

Introduction

Retroviruses code for the structural protein Gag which harbours the nucleocapsid (NC) protein as well as the matrix and capsid proteins. In HIV-1 it is located at the C-terminus of the Pr55Gag polyprotein precursor and is processed through the action of the retroviral protease to the NCp15 protein which is further cleaved off the NCp7 (the actual NC protein), to a proline-rich p6 and a p1 byproduct (Henderson *et al.*, 1990, 1992; Wondrak *et al.*, 1993).

NC proteins play an essential role during retroviral replication. They interact specifically with the viral RNA, select it from the pool of cellular RNAs and enable its encapsidation into newly budding virus particles. This packaging is an essential step in retroviral replication as evidenced by genetic studies which lead to non-infectious virus particles if the NC protein or the RNA is mutated (Aldovini and Young, 1990; Gorelick *et al.*, 1990). The site recognized on the viral RNA genome is the packaging site PSI (Ψ). If this is deleted, non-specific RNA accumulates in the virus (Méric and Goff, 1989). Furthermore, the NC proteins promote the encapsidation of two RNA molecules inside the virion by enabling an increase of dimerization (Prats *et al.*, 1988). Furthermore, the NC proteins contribute to the binding of the tRNA primer to the primer binding site (PBS) on the viral RNA (Barat *et al.*, 1989; Prats *et al.*, 1991). It also improves reverse transcription of viral RNA to DNA *in vitro* although only to a low extent (Volkmann *et al.*, 1993). The specificity of interaction between NC protein and PSI RNA has been exploited for the design of packaging cell lines which express viral structural proteins from a retroviral mRNA from which the PSI has been deleted. Transfection of DNA which contains sequence information for the PSI then allows transcribed RNA to be packaged into defective retroviral particles, an approach recently exploited for transfer of genes for gene therapy purposes (Culver *et al.*, 1991).

The PSI is located upstream of the gag translation initiation codon (Figure 1A, gag start). It includes the splice-donor site and a highly conserved purine-rich region. Its structure has been characterized previously and shown to harbour an A_5U_4 motif which is important for RNA packaging *in vivo* (Lever *et al.*, 1989; Aldovini and Young, 1990; Sakaguchi *et al.*, 1993) and characteristic of many HIV isolates but not conserved among retroviruses (AIDS database).

The NC proteins of retroviruses contain one or two conserved zinc finger motifs which are characterized by the sequence $CX_2CX_4HX_4C$ (CCHC), which functions physiologically as a zinc binding domain (Bess *et al.*, 1992; Summers *et al.*, 1992) (Figure 1A). This motif is also found in several cellular nucleic acid binding proteins such as CNBP (Rajavashisth *et al.*, 1989), Xpo (Sato and Sargent, 1991) and SLU7 (Frank and Guthrie, 1992) and is also present in retrovirus-like elements such as the retrotransposon *copia* from *Drosophila* (Mount and Rubin, 1985), and intracisternal A-type particle from Syrian hamster (Ono *et al.*, 1985), the coat protein of cauliflower mosaic virus (Covey, 1986) and the yeast Ty3-2 element (Hansen *et al.*, 1988). It occurs twice in HIV-1 NCp7 protein and is totally missing in only one retrovirus, the human spumaretrovirus (Maurer *et al.*, 1988). Point mutations in the conserved zinc finger motifs of several retroviral *gag* genes severely attenuate genomic RNA encapsidation (Méric and Goff, 1989; Aldovini and Young, 1990; Gorelick *et al.*, 1990; Dorfman *et al.*, 1993), although recent evidence suggests

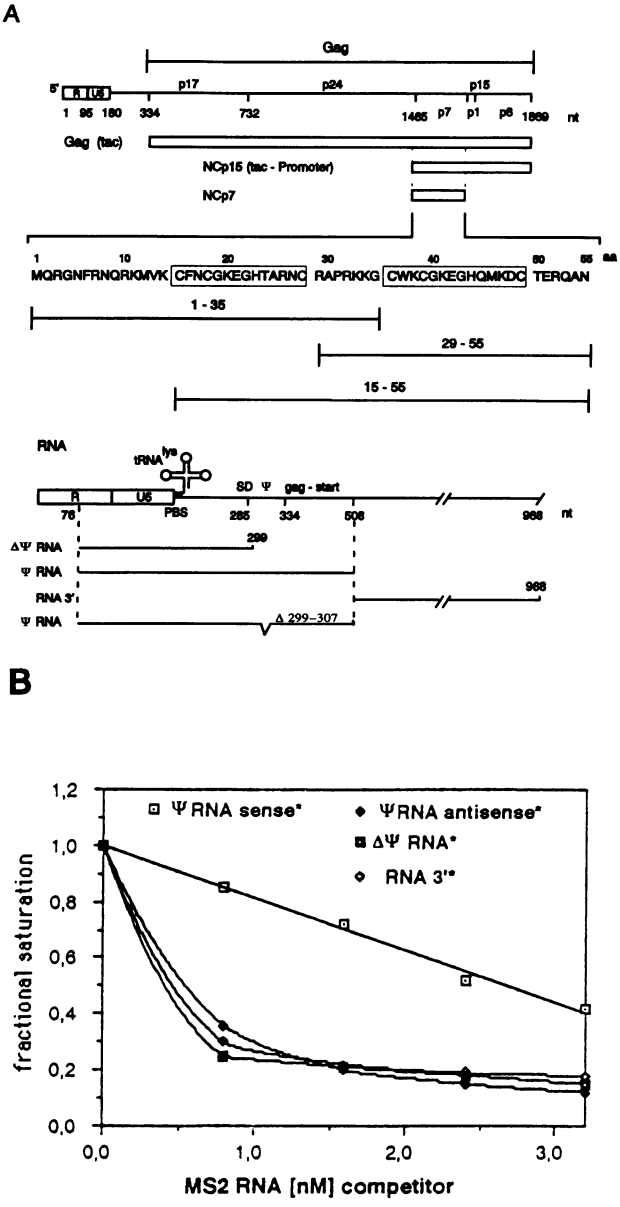


Fig. 1. (**A**) Schematic illustration of the 5′-HIV-1 RNA. The coding region of Gag is shown consisting of the structural proteins p17, p24 and p15 or NCp15 which is processed to the NCp7, p1 and p6. Numbers indicate the nucleotides (nt); R, repeat; U5, unique 5′ region. Two recombinant proteins expressed in *E.coli* under control of the *tac* promoter are indicated as NCp15 and NCp7. Three polypeptide fragments used in this study are shown schematically as lines with amino acid (aa) numbers. RNAs transcribed *in vitro* are illustrated underneath the 5′-leader of the HIV-1 RNA which is characterized by the $tRNA^{Lys}$, the PBS, the SD, the packaging site Ψ and the start codon of *gag*. A deletion is indicated by Δ, whereby the numbers refer to nucleotides deleted. (**B**) Filter binding assays of NCp7 to HIV-1 RNAs. Fractional saturation corresponds to the ratio of bound RNA to RNA bound at saturation. Binding of NCp7 to radioactively labelled (*) Ψ RNA sense, antisense, Δ Ψ RNA and RNA 3′ in the presence of bacteriophage MS2 RNA as competitor at the concentrations indicated. The concentrations of the radioactive RNAs were calculated to be $\sim 1.4 \times 10^{-11}$ M. The concentration of NCp7 was 7.25×10^{-8} M.

that this may be due more to instability of the RNA after encapsidation (Aronoff *et al.*, 1993). Furthermore, the amino acids of the first zinc finger are more highly conserved among retroviruses (Summers, 1991), indicating that the two zinc fingers are not under equal evolutionary selection pressure. Sequence comparison suggests a more detailed consensus sequence for the first zinc finger, $CaroXCX_3$-$GHhyd/aroX_3C$ (aro, aromatic; hyd, hydrophobic).

The first zinc finger is flanked by two clusters of basic amino acid residues, which have been shown by mutagenesis in *in vivo* studies to be involved in packaging of viral RNA (Fu *et al.*, 1988; Housset *et al.*, 1993). Basic amino acid sequences have been found to be characteristic of many RNA binding proteins (Lazinski *et al.*, 1989).

The NC proteins of several virus isolates have been isolated and characterized as single-stranded nucleic acid binding proteins (Sykora *et al.*, 1981; Karpel *et al.*, 1987). Efforts to characterize the PSI of HIV RNA have been made by several groups who have shown that deletion of sequences of the PSI region result in an RNA packaging deficiency of the virus particles (Lever *et al.*, 1989; Aldovini and Young, 1990; Poznansky *et al.*, 1991). Furthermore, stem–loop RNA structures have been proposed to play a role in packaging (Hayashi *et al.*, 1992; Sakaguchi *et al.*, 1993). They did not, however, analyse specific interaction with the NC protein *in vitro*. Some evidence for specific RNA binding of recombinant HIV-1 Gag protein in Northwestern blot analyses has been described (Luban and Goff, 1991).

Here we show for the first time that specific interaction of NCp7 with PSI-containing RNA can be determined *in vitro*. The results presented demonstrate that the N-terminal zinc finger, in particular two of the amino acids located within this zinc finger, is essential for specific RNA binding. The second zinc finger cannot replace the first one. Furthermore, the arginine residues flanking the first zinc finger are essential for specific PSI RNA recognition as well.

Results

The NCp7 protein used previously to investigate interaction with viral RNA has been synthesized chemically and has been shown to bind to RNA and single- as well as double-stranded DNA in filter binding assays (Surovoy *et al.*, 1993). The presence of zinc is required for the formation of the zinc finger type structures of the NCp7 protein. The sequence of NCp7 is indicated in Figure 1A. The NCp7 protein adopts α-helical structures at its N- and C-termini in organic solvents, as evidenced by circular dichroism (Surovoy *et al.*, 1993). The binding constant of NCp7 to RNA *in vitro* is $\sim 10^{-9}$ M and is 10-fold lower than that to DNA (Dannull *et al.*, unpublished observations). Binding to single-stranded DNA is slightly stronger than to double-stranded DNA. Modifications of the NCp7 by acetamidomethylation did not affect RNA binding. It did, however, slightly reduce the binding affinity to single-stranded DNA as shown previously (Surovoy *et al.*, 1993).

In order to investigate whether specific interaction between NCp7 and viral RNA could be detected *in vitro* we applied competitive filter binding assays using non-specific bacteriophage MS2 RNA as competitor in the initial studies with increasing concentrations in reactions containing radioactively labelled viral RNAs of various specificities. Figure 1 shows a scheme of various viral RNAs which were transcribed from DNA constructs obtained from HIV-1 D10.3 cDNA provirus (kindly supplied by Dr J.Karn) by

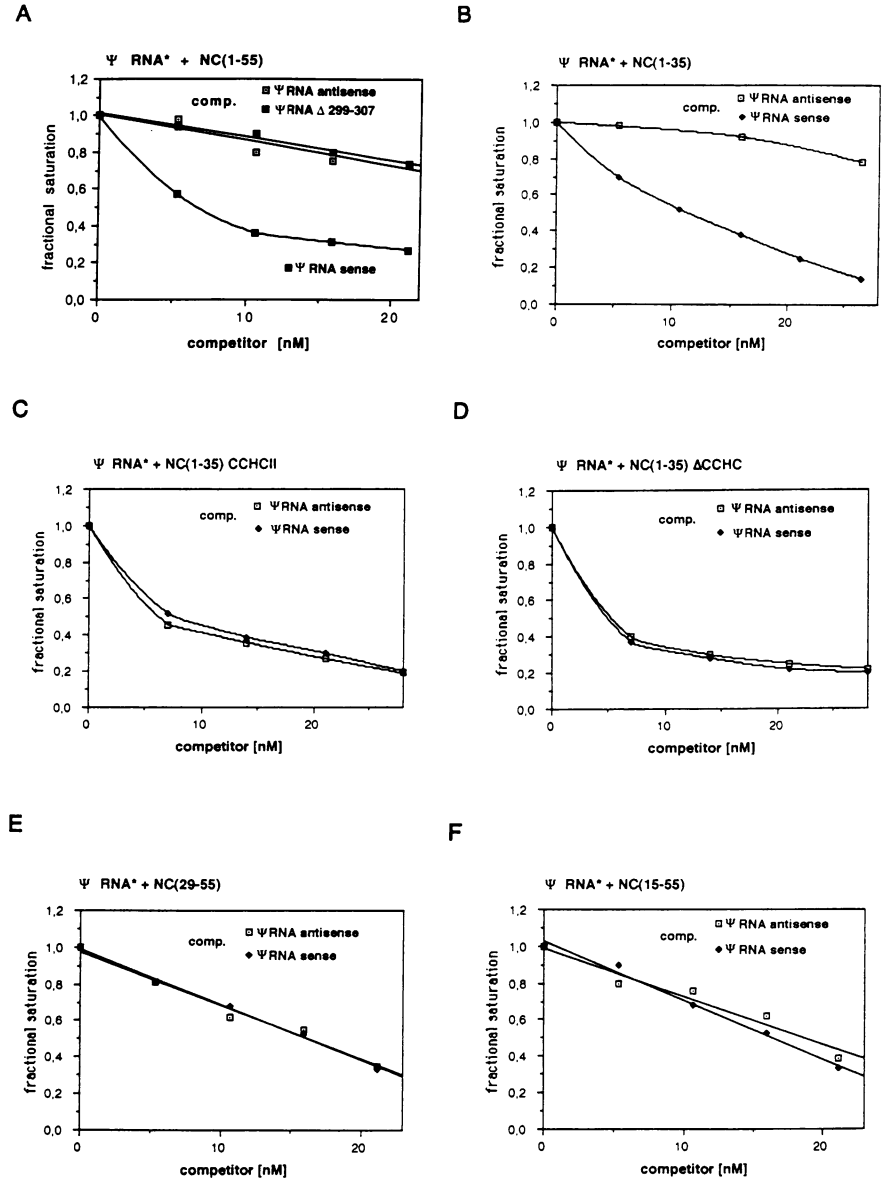
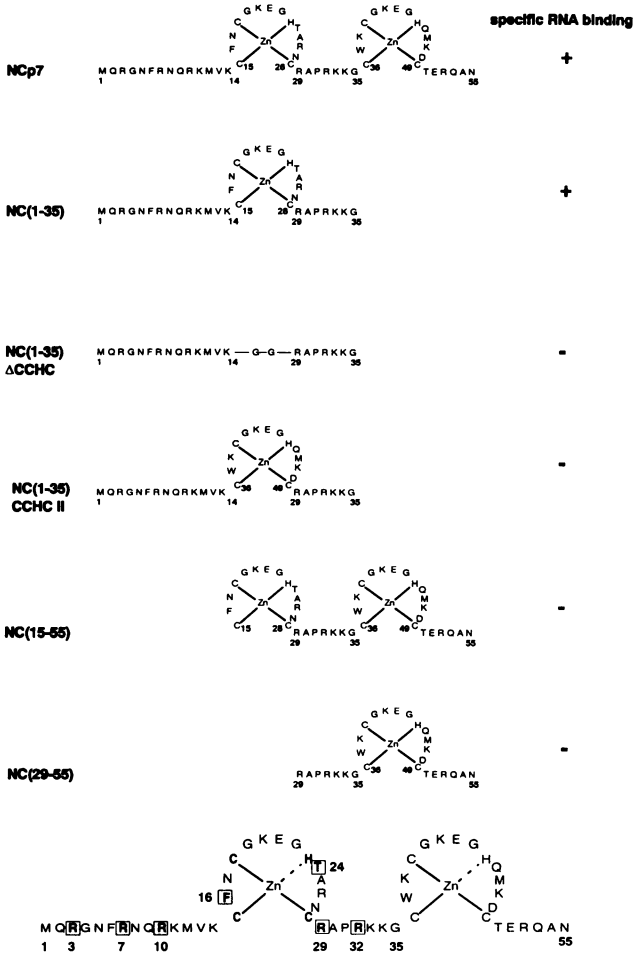


Fig. 2. Filter binding assays of synthetic NC peptide fragments to radioactively labelled (*) Ψ RNA ($\sim 1 \times 10^{-11}$ M) in the presence of increasing concentrations of either homologous Ψ RNA sense or heterologous Ψ RNA antisense as competitors and also with Ψ RNA Δ299–307 in the case of NCp7(1–55). Peptides used were: (**A**) NCp7(1–55) peptide (7.25×10^{-8} M), (**B**) NC(1–35) (9.6×10^{-8} M), (**C**) NC(1–35) CCHCII (9.6×10^{-8} M), (**D**) NC(1–35) ΔCHHC (9.6×10^{-8} M), (**E**) NC(29–55) (9×10^{-7} M), (**F**) NC(15–55) (7.5×10^{-7} M).

restriction nuclease cleavage and cloning behind the T7 promoter at the 5′- and the SP6 promoter at the 3′-end. Transcription thereby can be initiated at either end and allows synthesis of RNA in the sense as well as antisense orientation.

Radioactively labelled RNA representing nucleotides 76–508 of the 5′ HIV-1 RNA comprising the PSI (designated as Ψ RNA) in sense and antisense orientation was mixed with NCp7 protein for a filter binding assay. The amounts of RNA and protein used were predetermined in saturation curves with radioactively labelled RNA and increasing concentrations of protein (not shown). The conditions leading to saturation were then applied for competition with increasing concentrations of bacteriophage MS2 RNA added to the reaction mixture as competitor. The addition of protein was always last to ensure that binding assays were performed at equilibrium. The result of such a filter binding assay is shown in Figure 1B. As can be seen, binding of Ψ RNA sense to NCp7 is much more resistant to competition with MS2 RNA than Ψ RNA antisense, indicating a higher affinity of NCp7 to specific RNA sequences. RNA lacking PSI (ΔΨ RNA) is also competed out rapidly by MS2 RNA to a similar degree as the non-specific Ψ RNA antisense. The results demonstrate that sequences important for specific viral RNA binding must be located between nucleotides 299 and 508 (Figure 1A). The putative PSI determined by genetic analyses has been localized between nucleotides 293 and 314 (Lever *et al.*, 1989; Aldovini and Young, 1990; Poznansky *et al.*, 1991). It is apparently destroyed in ΔΨ RNA (which ends at nucleotide 299). RNA from nucleotides 509 to 968 (RNA 3′) which does not harbour the PSI gave, however, no indications for specific binding since it did not resist competition by MS2 RNA.

Table I. Synthetic NC peptide fragments were tested in competitive filter binding assays (shown in Figure 3A–F) for specific RNA binding



Numbers indicate amino acid residues. Two glycines (G–G) bridge the deleted first zinc finger (designated as ΔCCHC). The mutant designated as CCHCII contains the second zinc finger in place of the first one. (+) specific RNA binding, (–) no specific RNA binding. Amino acid residues of HIV-1 NCp7 essential for specific RNA binding are boxed and shaded.

Importance of the first zinc finger

Binding of NCp7(1–55) to radioactively labelled Ψ RNA can be easily competed out with unlabelled homologous Ψ RNA sense whereas it resists competition with Ψ RNA antisense (Figure 2A), indicating specific binding of NCp7 to Ψ RNA. To prove further the specificity of recognition of the PSI competition was also performed using viral RNA from which the PSI, spanning nucleotides 299–307, had been deleted (see scheme in Figure 1A). This RNA behaves like a non-specific inhibitor similar to the Ψ RNA antisense (Figure 2A).

The synthesis of a NC peptide comprising the 35 N-terminal amino acids [NC(1–35)] has been undertaken. This truncated NC peptide was used for a filter binding assay with Ψ RNA as radioactively labelled RNA (Figure 2B). Increasing amounts of unlabelled Ψ RNA in sense or antisense orientation were used as homologous and heterologous competitors at concentrations indicated in the legend to Figure 2. The result shown in Figure 2B demonstrates that NC(1–35) specifically binds to Ψ RNA sense and resists competition with Ψ RNA antisense. The binding is of similar efficiency to the one of the intact NCp7(1–55) (compare with Figure 2A). Two mutants were synthesized, one in which the first zinc finger was replaced by the second one, NC(1–35) CCHCII, and a second one containing two glycine residues instead of the first zinc finger, NC(1–35) ΔCCHC. Binding of NC(1–35) CCHCII and of NC(1–35) ΔCCHC to radioactively labelled Ψ RNA is competed out with equal concentrations of Ψ RNA sense or antisense (Figure 2C and D), indicating non-specific binding of these peptides. These experiments prove the crucial role of the first zinc finger for specific RNA binding. The second zinc finger—which is less conserved among retroviruses—is dispensible for specific RNA binding and cannot replace the first one.

Binding of NC(29–55) and NC(15–55) to radioactively labelled Ψ RNA is competed out with equal concentrations of Ψ RNA sense or antisense (Figure 2E and F), indicating non-specific binding of these peptides. These peptides bind ~10-fold less efficiently to RNA. Therefore protein concentrations had to be increased to compensate for the reduced binding. The shape of the competition curves of NC(29–55) and NC(15–55) corresponds to the initial phase of the biphasic curves of the other peptides. In summary, the results demonstrate that NC(1–35) is sufficient for specific NC–RNA interaction. The results of the binding assays are listed in Table I.

Specific RNA binding of recombinant Pr55Gag and NCp15

Prior to budding out of the host cell the virus assembles and selects the viral RNA from the pool of cellular RNAs by the intact Pr55Gag polyprotein precursor, which harbours the NCp15 at its C-terminus, before it is processed by the HIV-1 specific protease, a reaction which occurs mainly after budding of the virus out of the cell. Futhermore, it is not clear at the present time whether cleavage of NCp15 to NCp7, p6 and a p1 byproduct is an essential process. When this processing occurs and its biological effect on viral replication is still unclear. Therefore it was of interest to determine whether or not Pr55Gag or NCp15 would also be capable of specifically binding viral RNA. To determine this *in vitro*, NCp15 and Pr55Gag were expressed as recombinant proteins (for details see Materials and methods). Lysates from bacteria expressing NCp15 and Pr55Gag were prepared and analysed on SDS–polyacrylamide gels. As the NCp15 and Pr55Gag proteins are degraded by bacterial proteases, they were isolated according to the method described by Hagar and Burgess (1980) and applied to an SDS–polyacrylamide gel for staining with Coomassie brilliant blue and were also identified in an immunoblot using a C-terminal-specific NCp15 peptide antibody (Figure 3A and C). The NCp15 was cut out of a preparative SDS–polyacrylamide gel, electroeluted and verified by protein staining and immunoblotting (Figure 3C). As a control the region from which NCp15 had been eluted was cut out of a control gel loaded with bacterial lysates of bacteria not expressing NCp15 protein (Figure 3C, control). NCp15 and Pr55Gag were analysed in filter binding assays with radioactively labelled Ψ RNA using increasing concentrations of Ψ RNA sense, antisense or Ψ RNA Δ299–307 as competitors. The results shown in Figure 3B and D indicate that Pr55Gag as

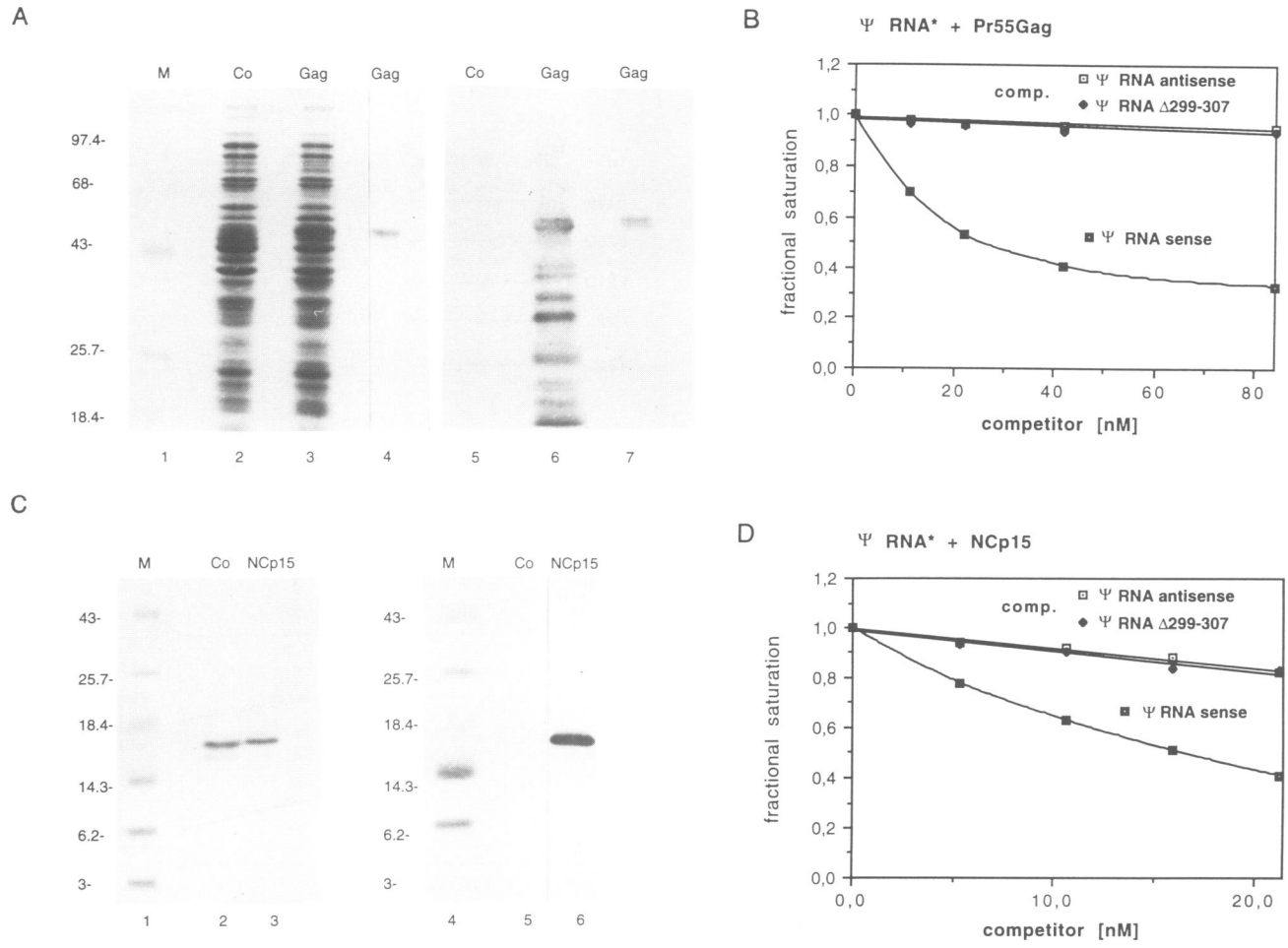


Fig. 3. Expression, purification and binding of recombinant NCp15 and Pr55Gag to Ψ RNA. (**A**) Bacterial lysates of bacteria expressing Pr55Gag and the parental plasmid (Co) were applied to SDS–polyacrylamide gels and stained with Coomassie brilliant blue (lanes 2 and 3). The Pr55Gag protein was eluted from a gel slice, an aliquot of which is shown in lane 4. The same material as shown in lanes 2–4 was analysed in immunoblot using an NC peptide-specific antibody (lanes 5, 6 and 7). M (lane 1) is a molecular weight marker, numbers correspond to kDa. (**B**) Filter binding assay using Pr55Gag (1.8×10^{-7} M) shown in (A) with radioactively labelled (*) Ψ RNA (1.4×10^{-11} M) in the presence of increasing concentrations of Ψ RNA sense, antisense or Ψ RNA Δ299–307 as competitors. (**C**) NCp15 (1.7×10^{-7} M) and a control (Co) protein were eluted from SDS–polyacrylamide gels to which lysates of bacteria expressing NCp15 or the parental plasmid had been applied. The NCp15 protein and a background control protein (Co) were stained with Coomassie brilliant blue (lanes 2 and 3). The same material is shown in immunoblot using an NC peptide-specific antibody (lanes 5 and 6). M (lanes 1 and 4) is a molecular weight marker, numbers correspond to kDa. (**D**) Filter binding assays using NCp15 shown in (C) with radioactively labelled Ψ RNA (*) using increasing concentrations of Ψ RNA sense, antisense or Ψ RNA Δ299–307 as competitors.

well as NCp15 specifically bind to viral Ψ RNA, even though ~3-fold higher protein concentrations were required. This may be due to partial inactivation of the proteins due to their isolation procedure. We noted that expression and purification of NCp15 as a glutathione-*S*-transferase (GST) fusion protein did not allow detection of specific RNA binding, suggesting that the fused part might affect the structure of the N-terminus and the zinc finger (Dannull *et al.*, unpublished observations).

Analysis of zinc finger mutants of NC(1–35)

In an attempt to understand the role of amino acids located within the first zinc finger we analysed several synthetic peptide mutants located in this region. Exchange of cysteine and histidine residues does not affect zinc co-ordination but allows identification of the importance of the cysteine and histidine residues.

First the CCHC motif was mutated as indicated schematically in Figure 4A. All mutants were tested in comparison with the wild-type [NC(1–35) CCHC] in competitive filter binding assays using radioactively labelled Ψ RNA and Ψ RNA sense or antisense as competitors (Figure 4B and C). Furthermore, zinc co-ordination was tested as described in Materials and methods. The results shown in Figure 4B and C indicate that the first two cysteine residues are essential for specific RNA recognition. Their mutations to histidine residues result in a reduction of specific RNA recognition indicated by +/–, whereas zinc co-ordination is not affected. On the other hand, mutations of the histidine to cysteine and the third cysteine to histidine do not influence specific RNA binding properties.

Furthermore, several synthetic mutants of amino acid residues located within the CCHC motif of the first zinc finger (Figure 4D) were tested in competitive filter binding assays using radioactively labelled Ψ RNA and Ψ RNA sense or antisense as competitors (Figure 4E and F). Interestingly, only mutants F16A and T24L are competed out rapidly by Ψ RNA antisense, indicating an important role of these amino

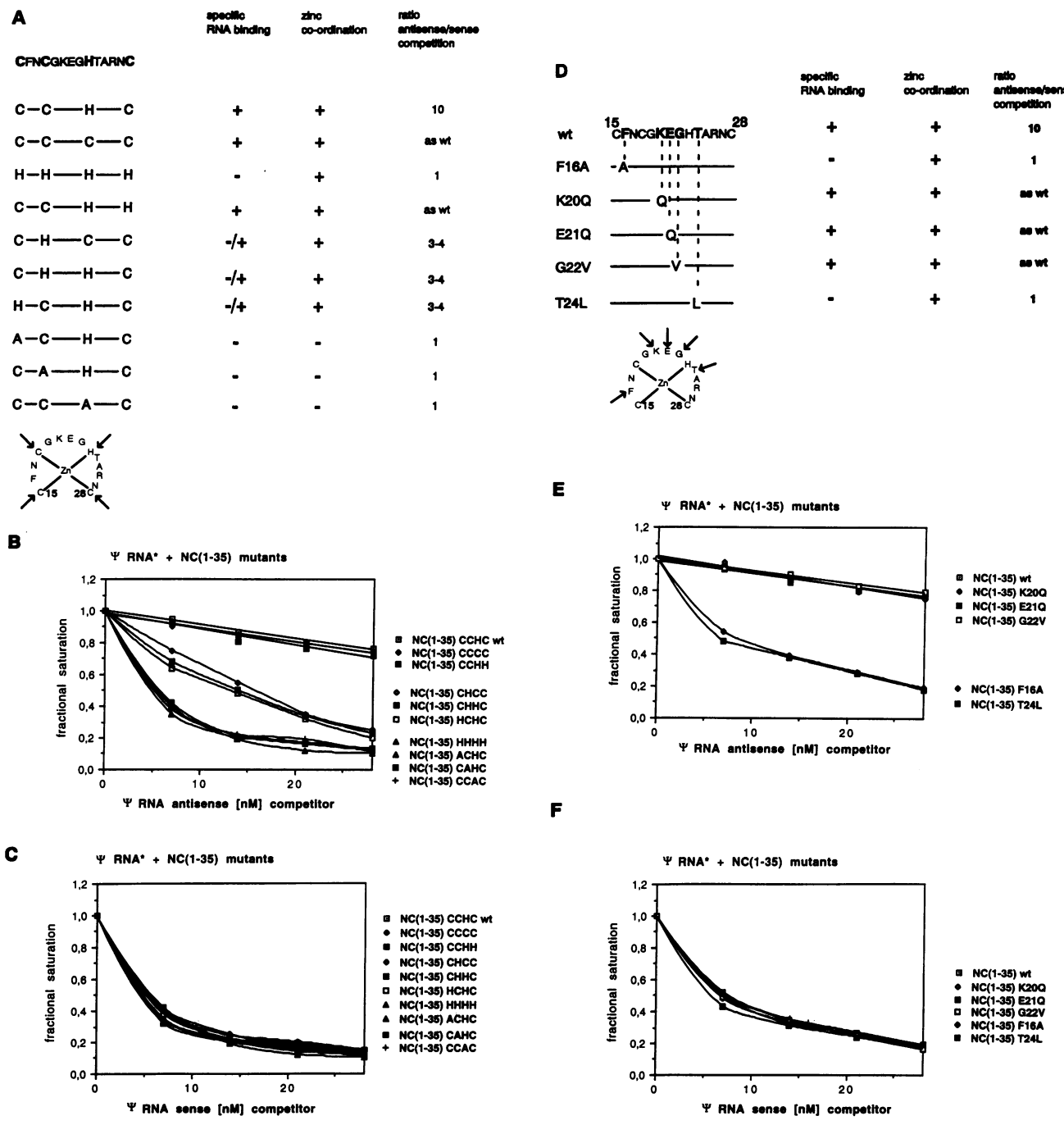


Fig. 4. Binding of NC(1–35) mutants to radioactively labelled Ψ RNA (*) (1.4×10^{-11} M) using increasing concentrations of heterologous Ψ RNA sense or antisense as competitors. Concentrations of peptides were 9.6×10^{-8} M. (**A**) Analyses of mutants of NC(1–35) CCHC with exchanges of C and H. (**B**) and (**C**) Filter binding studies of wild-type (wt) and mutant NC(1–35) peptides to Ψ RNA in the presence of Ψ RNA sense or antisense as competitors. Results of binding studies, of zinc determinations and ratios of Ψ RNA antisense to sense leading to the same fractional saturation values estimated from the curves are summarized in (A). (**D**) Analyses of NC(1–35) mutants located within the CCHC motif as indicated. (**E**) and (**F**) Filter binding assay of wild-type (wt) and mutant NC(1–35) peptides to radioactively labelled Ψ RNA in the presence of Ψ RNA sense or antisense as competitors. Results of binding studies, zinc determinations and ratios of Ψ RNA antisense to sense leading to the same fractional saturation values estimated from the curves are listed in (D).

acids in specific RNA recognition. All other mutants, K20 to Q, E21 to Q and G22 to V, resisted competition with Ψ RNA antisense (Figure 4E), indicating that their mutations were without effect on specific RNA binding. The results of competitions obtained with these mutants are presented as numbers (Figure 4A and D) estimated from the curves (Figure 4B, C, E and F) as ratios of concentrations of Ψ RNA antisense to sense leading to the same fractional saturation values.

Mutational analysis of basic amino acid residues flanking the first zinc finger

The basic amino acid residues flanking the first zinc finger correspond to the RNA binding sites of other RNA binding

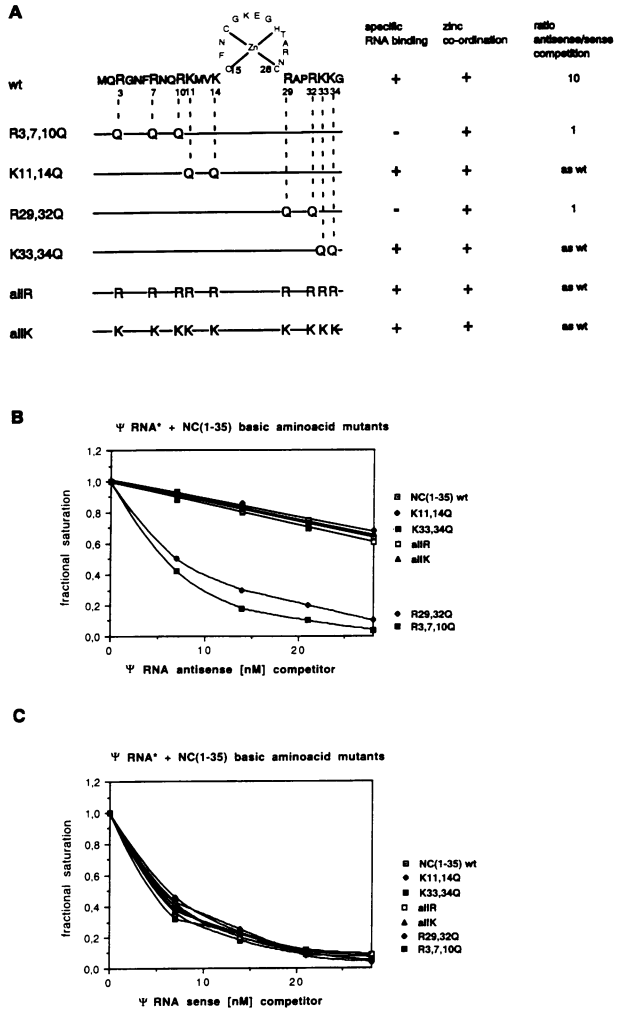


Fig. 5. (**A**) Binding of NC(1−35) peptides mutated in the basic amino acid residues as indicated to radioactively labelled Ψ RNA (*) (1.4×10^{-11} M) using Ψ RNA sense or antisense as competitors (**B**) and (**C**). Concentrations of peptides were 9.6×10^{-8} M but 1.92×10^{-7} M in the case of K33,34Q and R29,32Q due to reduced binding efficiencies. Results of filter binding studies, zinc determinations and the ratios of Ψ RNA antisense to sense (see legend to Figure 4) are summarized in (A).

proteins (Lazinski *et al.*, 1989) and have been shown to be important for RNA packaging *in vivo* (Fu *et al.*, 1988; Housset *et al.*, 1993), for tRNA annealing *in vitro* (deRocquigny *et al.*, 1992; Housset *et al.*, 1993) and for non-specific interaction of NC protein with viral RNA (Karpel *et al.*, 1987). Therefore the question arose, to what extent these basic amino acid residues contributed to the specific interaction of the NC protein with the PSI-containing RNA. For that purpose several mutants were designed (Figure 5A), whereby the first three arginine (R) residues were mutated to neutral glutamine (Q) residues designated as R3,7,10Q, the two lysine (K) residues to Q (K11,14Q), the two C-terminal R residues to Q (R29,32Q) and the two C-terminal K residues to Q (K33,34Q). Additionally, all the basic amino acid residues of one mutant were R, whilst the only basic amino acids another contained were K. These mutants were designated allR and allK, respectively. Competitive filter binding assays using these mutants and radioactively labelled Ψ RNA and Ψ RNA sense or antisense as competitors were performed as described above. The results of the binding assays and of zinc determinations are listed in Figure 5A. Numbers indicate the ratios of concentrations of Ψ RNA antisense to sense leading to same fractional mutants in saturation values. The results shown in Figures 5B and C demonstrate that mutants in R3, 7 and 10 and R29 and 32 are impaired in specific RNA binding whereas the others are not affected. Recent evidence indicates that R3 and R32 are more critical than the others (Schmalzbauer *et al.*, unpublished observation).

Discussion

The NC protein from HIV-1 contains two zinc fingers, only the first of which is highly conserved. Previous *in vivo* analyses have indicated that packaging of the viral RNA into the virion depends on the PSI region of the RNA and requires intact zinc finger structures (Méric and Goff, 1989; Aldovini and Young, 1990; Dupraz *et al.*, 1990). Mutations which destroy the structure of the first zinc finger have led to formation of defective virus particles and loss of selective recognition of the virus-specific RNA (Méric and Goff, 1989; Clavel and Orenstein, 1990). Furthermore, mutations of the basic amino acids flanking the first zinc finger have resulted in genomic RNA packaging defects (Fu *et al.*, 1988). These results have suggested that the basic amino acids flanking the first zinc finger and the zinc finger itself play a crucial role in recognition of PSIs on the viral RNA. *In vitro* and *in vivo* studies performed with NC proteins of various viral species have confirmed the biological data, that the basic amino acids flanking the zinc finger are critically important for binding of RNA (deRocquigny *et al.*, 1992; Housset *et al.*, 1993; Surovoy *et al.*, 1993). However, so far no specific binding to viral RNA or the PSI has been shown *in vitro*. Furthermore, deletion of the zinc fingers appeared not to affect various functions attributed to the NC protein such as increase of dimerization efficiency of viral RNA or annealing of tRNA primer to the viral RNA (deRocquigny *et al.*, 1992). Therefore an intact zinc finger structure is apparently not required for these virus-specific reactions.

Here we demonstrate that the structure of the first zinc finger of the HIV-1 NC protein is essential for recognition of specific viral RNA sequences *in vitro*, whereas the second zinc finger is dispensable for this interaction. This finger is much less conserved among various retroviruses and is of minor importance for RNA packaging in the case of Rous sarcoma virus and HIV-1 (Dupraz *et al.*, 1990; Dorfman *et al.*, 1993; Gorelick *et al.*, 1993). Inhibition of zinc co-ordination by site-directed mutagenesis of the cysteine and histidine residues of the first zinc finger or mutations located within this region destroy the specificity of recognition of the viral RNA. In particular amino acids F16 and T24 are essential for specific RNA recognition, which is in agreement with previous observations where hydrophobic amino acids of the zinc finger have been suggested to form a groove directly interacting with nucleotides in the RNA bound state of NCp7 (Summers *et al.*, 1992). Our results further indicate that the second zinc finger cannot replace the first one. It is likely that this may not only be due to different amino acid residues located within the second zinc finger but also to its conformation, since the structures of the first and second zinc finger have been shown to differ from each other with the first one exhibiting a more

pronounced β-hairpin-like structure and the second one an α-helical region at the C-terminus (South *et al.*, 1990; Surovoy *et al.*, 1993). Whether F16 and T24 are involved directly in RNA–protein interaction needs to be shown.

Our results suggest that the basic amino acids flanking the zinc finger need to be oriented by the intact zinc finger structure in such a way as to recognize the PSI of the viral RNA. The basic amino acid residues are essential for non-specific RNA binding, while the zinc finger induces specific recognition of the PSI of the viral RNA. However, even with an intact first zinc finger, specific RNA recognition depends on some of the basic amino acid residues. This is shown by mutations of the basic amino acid residues. Especially the amino acid residues R3, 7 and 10 located N-terminally and R32 and 33 located C-terminally of the first zinc finger are essential for specific PSI RNA interaction. The specificity does not simply correlate with the total number of basic amino acid residues. It is interesting to note that the arginine residues can be replaced by lysines, indicating that arginines may not necessarily be required. The basic amino acid stretches cluster in two regions adjacent to the zinc finger, while other RNA binding proteins contain only one such cluster. In the case of Rev and Tat the specificity for RNA recognition has been ascribed to a consensus sequence of basic amino acid residues present in various RNA binding proteins, BXB**RBXRR**X_2B (Lazinski *et al.*, 1989), which bears some resemblance to the two basic amino acid clusters in NCp7 [RX_2RK (N-terminally) and RX_2RK (C-terminally)]. The amino acid residues essential for specific PSI RNA–NCp7 interaction as determined in this study are summarized at the bottom of Table I.

NC proteins preferentially bind to RNA but also bind to single-stranded and double-stranded DNA *in vitro* (Sykora and Moelling, 1981; Surovoy *et al.*, 1993). We noticed previously that the intact zinc finger structure allowed a slightly higher binding efficiency to single-stranded over double-stranded DNA. When the zinc co-ordination of the finger was prevented by acetamidomethyl modification, this difference disappeared (Surovoy *et al.*, 1993). The biological significance of DNA binding of viral NC proteins is presently not known but several events in the retroviral life cycle (such as enhancement of the activity of retroviral integrase and hybridization of RNA to ssDNA as well as ssDNA to ssDNA) have recently been suggested to depend on the presence of NC protein (Dib-Hajj *et al.*, 1993; Lapadat-Tapolski *et al.*, 1993).

We have previously described the binding constant of the NCp7 protein from HIV-1 to RNA as 10^{-9} M (Surovoy *et al.*, 1993). The specific binding is ~10-fold higher than non-specific, which is still significantly lower than that observed for the Rev–RRE (Rev responsive element) interaction (Daly *et al.*, 1989; Heaphy *et al.*, 1990). The NC proteins play an important role in the viral life cycle. On one hand they specifically recognize the PSI of the viral RNA for virus assembly, and on the other hand they bind to the viral RNA in a non-specific manner protecting it against nucleases or enhancing condensation of the RNA for packaging. These two functions resemble those of the Rev–RRE interaction which exhibits a specific initiation event and subsequently non-specific binding of the RNA surrounding the RRE site, thereby preventing splicing of the viral RNA (Heaphy *et al.*, 1991). A similar scenario can be envisaged for the NC–PSI interaction which explains the relatively low difference between specific and non-specific binding.

The enhancement of RNA dimerization by NC proteins is not due to sequence-specific interaction with RNA because the zinc fingers are dispensible for this function (Dupraz *et al.*, 1991; deRocquigny *et al.*, 1992). The NC protein apparently increases annealing of nucleic acids in a non-specific way. This must also be the case for annealing of the tRNA to the primer binding site. Since both of these events are highly specific, the specificity must be due to other factors. For dimerization G-cage formation (Sundquist and Klug, 1989; Sundquist and Heaphy, 1993) could be the specific initial event and for tRNA–RNA hybridization pseudo-knot formation has been described for the avian sarcoma leukemia virus (Aiyar *et al.*, 1992) which may be responsible for the specificity of this event. NC proteins may improve RNA–RNA interaction by their ability to unwind or melt RNA as described previously (Sykora and Moelling, 1981; Khan and Giedroc, 1992; Dib-Hajj *et al.*, 1993). This melting effect is, however, only of minor importance for the reverse transcriptase *in vitro* (Volkmann *et al.*, 1993). Whether it is more important *in vivo* has not yet been shown.

Viral assembly takes place by means of the Gag polyprotein precursor and not the processed p15 or p7 proteins. Therefore it is important to note that the Gag polyprotein is also capable of specifically recognizing the viral RNA. The HIV-1 NCp15 is cleaved to NCp7, p1 and p6 during virus maturation or possibly also early during infection. Reduced binding affinities of NCp15 and Pr55Gag compared with NCp7 may be accounted for by differences in preparations and renaturation efficiencies. It needs to be analysed what the differences in biological function of NCp15 versus NCp7 might be. The C-terminal p6 is very proline-rich and acidic and may play a role in interaction with the basic portion of NCp7 or other proteins.

Our results demonstrate that the PSI is specifically recognized by the NCp7 *in vitro*. The PSI may not be sufficient for selective packaging of HIV-1 RNA, since *cis*-acting sequences located within the *gag* and *env* genes of HIV-1 have been shown to be important for RNA encapsidation or specific RNA binding as well (Luban and Goff, 1991; Richardson *et al.*, 1993). Also the matrix proteins of HIV-1 and other retroviruses exhibit RNA binding properties (Katoh *et al.*, 1991; Bukrinskaya *et al.*, 1992) and may contribute to specific packaging. The *in vitro* system described here will allow detailed analyses of these effects.

Materials and methods

Synthetic nucleocapsid proteins

The NCp7 protein of HIV-1 consists of 55 amino acids and was synthesized chemically on an automated peptide synthesizer ABI430A (Applied Biosystems) by Dr A.Surovoy as described recently (Surovoy *et al.*, 1993). Furthermore, NC(1–35) as well as the NC(1–35) mutants, NC(29–55) and NC(15–55) were synthesized analogously, whereby numbers indicate amino acids of the NCp7 peptides.

Recombinant proteins

NCp15 was cloned by means of PCR primers into the *Nco*I and *Hin*dIII sites of pKK233-2 which leads to addition of an N-terminal methionine and alanine. Expression of recombinant protein is under control of the *tac* promoter in *Escherichia coli* (JM109). Expression of recombinant proteins can be induced by addition of 1 mM IPTG to mid-log phase cultures. Induction was for 3 h, the bacteria were lysed and subjected to SDS–PAGE and stained with Coomassie brilliant blue or treated with an antibody directed against the C-terminal synthetic peptide of the NCp15 (Seromed, Germany)

for immunoblotting. The NCp15 and the Pr55Gag proteins were recovered from the gel by elution according to the procedure described by Hagar and Burgess (1980). To account for non-specific proteins the parental vector was transformed into bacteria which were lysed and applied to gel electrophoresis in parallel. The equivalent regions were sliced and electroeluted and the proteins used as controls. Preparative gels were electrophoresed for longer periods than shown in Figure 3C to improve the separation of NCp15 from a non-specific band. Protein contents were determined by optical density based on standard proteins (not shown). The NCp15 preparation may be contaminated with background protein which, however, does not contribute to binding reactions, since RNA binding properties of Pr55Gag and NCp15 did not differ significantly, as determined by RNA binding assays using increasing concentrations of proteins (not shown). The eluted proteins were stored in 6 M urea and renatured prior to use for 1 h at 4°C in 10 mM Tris–HCl pH 7.5, 50 mM NaCl, 10 μM $ZnCl_2$, 10 mM dithiothreitol supplemented with 500 ng/μl bovine serum albumin.

In vitro transcription of RNA

The DNA insert containing the *Hin*dIII–*Acc*I fragment of HIV-1 (439 nucleotides in length) was cloned into the multiple cloning site of the pGEM4Z plasmid which contains the SP6 T7 promoters in opposite directions (Promega, Madison, WI), allowing transcription in sense and antisense orientation. *In vitro* transcription by RNA polymerase results in an RNA transcript which initiates at nucleotide 76 within the R region and ends at nucleotide 508 within the *gag* coding region, covering the PBS, the splice donor (SD) and the PSI. The RNA was designated as Ψ RNA. An RNA fragment ΔΨ RNA with a deleted PSI region was transcribed from a DNA fragment (*Hin*dIII–*Hph*I) corresponding to nucleotides 76–299. A third RNA fragment RNA-3′ ranged from nucleotide 508 to 968, corresponding to the *Acc*I–*Pst*I HIV-1 DNA. A deletion mutant was derived from the DNA construct coding for the Ψ RNA. The Ψ deletion was obtained by recombinant PCR and resulted in deletion of nucleotides 299–307 (see Figure 1A). Radioactive labelling of Ψ RNA was performed by incorporation of [α-^{32}P]UTP during transcription of the DNA templates by T7 or SP6 RNA polymerase.

RNA binding assays

Filter binding reactions were performed using radioactively labelled RNAs, NC peptides and competitor RNAs as indicated in the figure legends. Details have been described before (Surovoy *et al.*, 1993). Each assay was performed in duplicate wherein variations did not exceed 10%.

Determination of zinc ions

Peptides were incubated in 50 mM DTT for 1 h at 37°C and dialysed against 50 mM Tris–HCl pH 8.0, 50 mM NaCl, 10 μM $ZnCl_2$. Concentrations of zinc ions were determined after dialysis of the peptides against zinc-free buffer using diphenylthiocarbazone (dithizone) as described by Holmquist (1988).

Acknowledgements

The authors are indebted to Dr J.Karn for kindly supplying the D10.3 provirus and to Dr K.von der Helm for the Pr55Gag expression vector. The RNA deletion mutant was kindly supplied by Edith Schmalzbauer. This work was supported by the Deutsche Forschungsgemeinschaft SFB 344 and the Dr Mildred Scheel Stiftung F 66.

References

Aiyar,A., Cobrinik,D., Ge,Z., Kung,H.J. and Leis,J. (1992) *J. Virol.*, **66**, 2464–2474.

Aldovini,A. and Young,R.A. (1990) *J. Virol.*, **64**, 1920–1926.

Aronoff,R., Hajjar,A.M. and Linial,M.L. (1993) *J. Virol.*, **67**, 178–188.

Barat,C., Lullien,V., Schatz,O., Keith,G., Nugeyre,M.T., Grüninger-Leitch,F., Barré-Sinoussi,F., LeGrice,S.F.J. and Darlix,J.L. (1989) *EMBO J.*, **8**, 3279–3285.

Bess,J.W., Powell,P.J., Issaq,H.J., Schumack,L.J., Grimes,M.K., Henderson,L.E. and Arthur,L.O. (1992) *J. Virol.*, **66**, 840–847.

Bukrinskaya,A.G., Vorkunova,G.K. and Tentsov,Y.Y. (1992) *AIDS Res. Hum. Retrovir.*, **8**, 1795–1801.

Clavel,F. and Orenstein,J.M. (1990) *J. Virol.*, **64**, 5230–5234.

Covey,S.N. (1986) *Nucleic Acids Res.*, **14**, 623–633.

Culver,K., Cornetta,K., Morgan,R., Morecki,S., Aebersold,P., Kasid,A., Lotze,M., Rosenberg,S.A., Anderson,W.F. and Blaese,R.M. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 3155–3159.

Daly,T.J., Cook,K.S., Gray,G.S., Maione,T.E. and Rusche,J.R. (1989) *Nature*, **342**, 816–819.

Daly,T.J., Cook,K.S., Gray,G.S., Maione,T.E. and Rusche,J.R. (1989) *Nature*, **342**, 816–819.

deRocquigny,H., Gabus,C., Vincent,A., Fournié-Zaluski,M.C., Rocques,B. and Darlix,J.L. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 6472–6476.

Dib-Hajj,F., Khan,R. and Giedroc,D.P. (1993) *Protein Sci.*, **2**, 231–243.

Dorfman,T., Luban,J., Goff,S.P., Haseltine,W.A. and Gottlinger,H.G. (1993) *J. Virol.*, **67**, 6159–6169.

Dupraz,P., Oertle,S., Méric,C., Damay,P. and Spahr,P.F. (1990) *J. Virol.*, **64**, 4978–4987.

Frank,D. and Guthrie,C. (1992) *Genes Dev.*, **6**, 2112–2124.

Fu,X., Katz,R.A., Skalka,A.M. and Leis,J. (1988) *J. Biol. Chem.*, **263**, 2140–2145.

Gorelick,R.J., Nigida,S.M., Bess,J.W., Arthur,L.O., Henderson,L.E. and Rein,A. (1990) *J. Virol.*, **64**, 3207–3211.

Gorelick,R.J., Chabot,D.J., Rein,A., Henderson,L.E. and Arthur,L.O. (1993) *J. Virol.*, **67**, 4027–4036.

Hagar,D.A. and Burgess,R.R. (1980) *Anal. Biochem.*, **109**, 76–86.

Hansen,L.J., Chalder,D.L. and Sandmeyer,S.B. (1988) *Mol. Cell. Biol.*, **8**, 5245–5256.

Hayashi,T., Shioda,T., Iwakura,Y. and Shibuta,H. (1992) *Virology*, **188**, 590–599.

Heaphy,S., Dingwall,C., Ernberg,I., Gait,M.J., Green,S.M., Karn,J., Lowe,A.D., Singh,M. and Skinner,M.A. (1990) *Cell*, **60**, 685–693.

Heaphy,S., Finch,J.T., Gait,M.J., Karn,J. and Singh,M. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 7366–7370.

Henderson,L.E., Sowder,R.C., Copeland,T.D., Oroszlan,S. and Benveniste,R.E. (1990) *J. Med. Primatol.*, **19**, 411–419.

Henderson,L.E., Bowers,M.A., Sowder,R.C., Serabyn,S.A., Johnson,D.G., Bess,J.W., Arthur,L.O., Bryant,D.K. and Fenselau,C. (1992) *J. Virol.*, **66**, 1856–1865.

Holmquist,B. (1988) *Methods Enzymol.*, **158**, 6–10.

Housset,V., deRocquigny,H., Rocques,B.P. and Darlix,J.L. (1993) *J. Virol.*, **67**, 2537–2545.

Karpel,R.L., Henderson,L.E. and Oroszlan,S. (1987) *J. Biol. Chem.*, **262**, 4961–4967.

Katoh,I., Kyushiki,H., Sakamoto,Y., Ikawa,Y. and Yoshinaka,Y. (1991) *J. Virol.*, **65**, 6845–6855.

Khan,R. and Giedroc,D.P. (1992) *J. Biol. Chem.*, **267**, 6689–6695.

Lapadat-Tapolski,M., deRocquigny,H., van Gent,D., Rocques,B., Plasterk,R. and Darlix,J.L. (1993) *Nucleic Acids Res.*, **21**, 831–839.

Lazinski,D., Grzadzilska,E. and Das,A. (1989) *Cell*, **59**, 207–218.

Lever,A., Gottlinger,H., Haseltine,W. and Sodroski,J. (1989) *J. Virol.*, **63**, 4085–4087.

Luban,J. and Goff,S.P. (1991) *J. Virol.*, **65**, 3203–3212.

Maurer,B., Bannert,H., Darai,G. and Flügel,R.M. (1988) *J. Virol.*, **62**, 1590–1597.

Méric,C. and Goff,S.P. (1989) *J. Virol.*, **63**, 1558–1568.

Mount,S.M. and Rubin,G. (1985) *Mol. Cell. Biol.*, **5**, 1630–1638.

Ono,M., Toh,H., Miyata,T. and Awaya,T. (1985) *J. Virol.*, **55**, 387–394.

Poznansky,M., Lever,A., Bergeron,L., Haseltine,W. and Sodrosky,J. (1991) *J. Virol.*, **65**, 532–536.

Prats,A.C., Sarih,L., Gabus,C., Litvak,S., Keith,G. and Darlix,J.L. (1988) *EMBO J.*, **7**, 1777–1783.

Prats,A.C., Housset,V., deBilly,G., Cornille,F., Prats,H., Rocques,B. and Darlix,J.L. (1991) *Nucleic Acids Res.*, **19**, 3533–3541.

Rajavashisth,T.B., Taylor,A.K., Andalibi,A., Svenson,K.L. and Lusis,A.J. (1989) *Science*, **245**, 640–643.

Richardson,J.H., Child,L.A. and Lever,A.M. (1993) *J. Virol.*, **67**, 3997–4005.

Sakaguchi,K. *et al.* (1993) *Proc. Natl Acad. Sci. USA*, **90**, 5219–5223.

Sato,S.M. and Sargent,T.D. (1991) *Development*, **112**, 747–753.

South,T.L., Blake,P.R., Sowder,R.C., Arthur,L.O., Henderson,L.E. and Summers,M.F. (1990) *Biochemistry*, **29**, 7786–7789.

Summers,M.F. (1991) *J. Cell. Biochem.*, **45**, 41–48.

Summers,M.F. *et al.* (1992) *Protein Sci.*, **1**, 563–574.

Sundquist,W.I. and Klug,A. (1989) *Nature*, **342**, 825–829.

Sundquist,W.I. and Heaphy,S. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 3393–3397.

Surovoy,A., Dannull,J., Mölling,K. and Jung,G. (1993) *J. Mol. Biol.*, **229**, 94–104.

Sykora,K.W. and Mölling,K. (1981) *J. Gen. Virol.*, **55**, 379–391.

Volkmann,S., Dannull,J. and Mölling,K. (1993) *Biochimie*, **75**, 71–78.

Wondrak,E.M., Louis,J.M., deRocquigny,H., Cherman,J.C. and Rocques,B.P. (1993) *FEBS Lett.*, **333**, 21–24.

Received on October 11, 1993; revised on January 12, 1994