

# A sensitive procedure for mapping the boundaries of RNA elements binding *in vitro* translated proteins defines a minimal hepatitis B virus encapsidation signal

Jürgen Beck and Michael Nassal\*

Zentrum für Molekulare Biologie, University of Heidelberg, Im Neuenheimer Feld 282, D-69120 Heidelberg, Germany

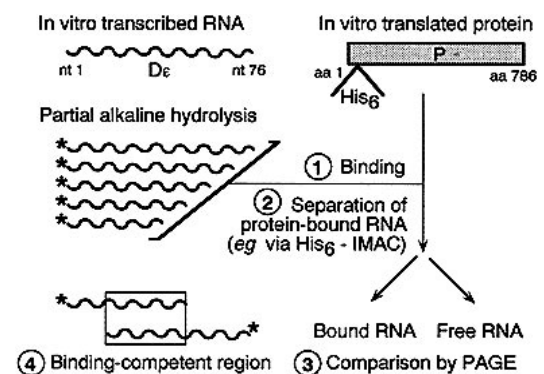
Received July 21, 1996; Revised and Accepted September 6, 1996

## ABSTRACT

Using the structured RNA encapsidation signal ( $D_E$ ) and the reverse transcriptase (P protein) of duck hepatitis B virus (DHBV) as an example, we devised a sensitive mapping procedure that yields accurate information on the minimal RNA sequence required for interaction with a few nanograms of an RNA-binding protein. RNAs from pools of end-labeled, partially hydrolyzed transcripts that bound to *in vitro* translated His-tagged P protein were isolated using immobilized  $Ni^{2+}$ -ions. Size analysis by PAGE is consistent with a gradual gain in binding-competence from a minimum of 5 to a maximum of 8 base pairs in the basal stem of  $D_E$ . The procedure should be generally applicable to the convenient and precise fine mapping of RNA–protein interactions.

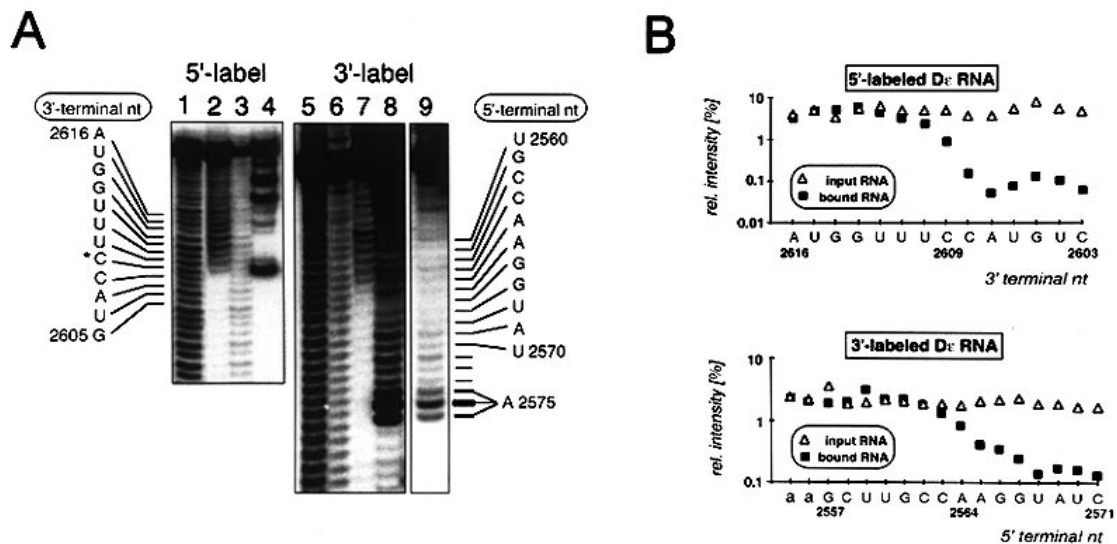
The fate of cellular and viral RNAs is regulated by the interaction of multiple *cis* elements with a variety of *trans*-acting partners, often proteins. Conventionally, the borders of a *cis* element are mapped by individually testing a limited number of differently sized transcripts, obtained from 5'-terminal deletion variants of an RNA expression plasmid linearized at varying 3'-proximal restriction sites, for binding to the isolated protein of interest. However, a complete set of shortened molecules can easily be obtained from the RNA itself by partial alkaline hydrolysis. Using RNA labeled at its 5'- or 3'-end, two pools are generated encompassing all possible deletions from one or the other terminus. Incubation with the protein, separation of protein bound from free RNA, and sizing by PAGE should reveal an overlapping set of nucleotides representing the minimal binding-competent sequence (Fig. 1). Using purified bacteriophage proteins, this principle has recently been used to determine the minimal size of binding-competent RNAs after their separation from free RNA on nitrocellulose filters (1).

Frequently, however, the protein of interest is neither available in pure form nor in sufficient amounts. *In vitro* translation, on the other hand, is a straightforward procedure to express almost any protein, although the yields are low both in absolute terms, and in



**Figure 1.** Outline of mapping procedure. *In vitro* transcribed RNA, here  $D_E$ , is end-labeled and partially hydrolyzed by alkali. The resulting pools of 5'- and 3'- (not shown) labeled fragments are allowed to bind to the *in vitro* translated protein of interest, here His-tagged P protein of DHBV (step 1). Protein-bound and free RNA pools are separated (step 2), e.g. via isolation by immobilized metal affinity chromatography (IMAC). Protein-bound and free RNAs are compared by denaturing PAGE (step 3). The overlap region between the shortest fragments present in the protein-bound pools of 5'- and 3'-labeled fragments defines the boundaries of the binding-competent RNA region (step 4). Specific methods:  $D_E$  RNA of nominally 76 nt was obtained by *in vitro* transcription (Megashortscript, Ambion, Austin, TX) from *Clal*-linearized plasmid pD $\epsilon$ 1 which contains DHBV 16 sequence from position 2557 to 2624 under control of the T7 promoter. For 5'-labeling, 5  $\mu$ g of dephosphorylated RNA were re-phosphorylated using 5 U T4 polynucleotide kinase (NEB) and 50  $\mu$ Ci [ $\gamma$ - $^{32}$ P]ATP (3000 Ci/mmol; Amersham-Buchler, Braunschweig, Germany). For 3'-labeling, 3  $\mu$ g  $D_E$ -RNA were treated with 20 U T4 RNA ligase (Gibco-BRL, Eggenstein, Germany) and 200  $\mu$ Ci [ $^{32}$ P]pCp (3000 Ci/mmol; Amersham-Buchler) as recommended by the enzyme manufacturer. For partial hydrolysis,  $5 \times 10^6$  c.p.m. of 5'-labeled (specific activity  $\sim 50\,000$  c.p.m./pmol), or  $1.5 \times 10^6$  c.p.m. of 3'-labeled (specific activity  $\sim 15\,000$  c.p.m./pmol) gel purified RNA were incubated for 15 min in 50  $\mu$ l 50 mM  $NaHCO_3$  at 90°C. The reaction was stopped by sodium acetate/ethanol precipitation. Pellets were washed with 70% ethanol and dissolved in 10  $\mu$ l TE. His<sub>6</sub>-tagged DHBV16 P protein was obtained by *in vitro* translation in 25  $\mu$ l rabbit reticulocyte lysate (Promega, Heidelberg, Germany) programmed with 2  $\mu$ g of a  $D_E$ -deficient P-mRNA (DHBV sequence from nt 170 to 2530) *in vitro* transcribed from *Afl*II-linearized plasmid pT7AMVpol16H6; this construct was derived from pT7AMVpol16 (15) by PCR-mediated insertion of six His codons between the second and third codon of the P ORF.

\*To whom correspondence should be addressed. Tel: +49 6221 54 68 15; Fax: +49 6221 54 58 93; Email: mnassal@sun0.urz.uni-heidelberg.de



**Figure 2.** Experimental application of mapping procedure. (A) PAGE analysis of P protein-bound RNAs. Dε RNA aliquots corresponding to  $3 \times 10^6$  c.p.m. (5'-label) or  $1.4 \times 10^6$  c.p.m. (3'-label) were added to 25  $\mu$ l of His<sub>6</sub>-P mRNA programmed reticulocyte lysate (see legend of Fig. 1 for details). After 90 min at 30°C, 400  $\mu$ l binding buffer (0.1 M sodium phosphate pH 7.5, 150 mM NaCl, 0.1% Nonidet-P40, 20 mM imidazole, 100  $\mu$ g/ml yeast total RNA) containing 25  $\mu$ l Ni<sup>2+</sup>-NTA-agarose beads (Qiagen, Hilden, Germany) were added. The mixture was shaken for 2 h at 4°C, then the beads were washed twice each with 1 ml binding buffer and 1 ml TMK buffer (50 mM Tris-HCl pH 7.5, 40 mM KCl, 10 mM MgCl<sub>2</sub>, 100  $\mu$ g/ml yeast total RNA). Phenol extracted, ethanol precipitated RNAs were analyzed on denaturing gels containing 8% polyacrylamide. RNase treatments for assigning band positions were performed with 50 000 c.p.m. end-labeled RNA in 50  $\mu$ l TMK buffer at 4°C for 30 min using 2.5 U Benzonase (Merck, Darmstadt, Germany) for the 5'-labeled, and 2 ng RNase A (Boehringer, Mannheim, Germany) for the 3'-labeled RNA. Lanes 1, 3, 5 and 6, input RNA pools at two different concentrations; lanes 2 and 7, protein-bound RNAs; lane 4, 5-labeled RNA treated with Benzonase; lane 8, 3'-labeled RNA treated with RNase A; lane 9, shorter exposure of lane 8. Sequence assignments are shown at the left and right side. \*C indicates the position of a major Benzonase cut at C2609; due to the 3'-heterogeneity of the *in vitro* transcripts, three signals rather than one were generated from the 3' labeled RNA by RNase A (e.g. A2575 in lane 9). (B) Semi-quantitative determination of protein binding-competence for 5' and 3'-labeled Dε RNAs. Band intensities of input and protein-bound RNAs were quantified using a PhosphorImager (Molecular Dynamics, Krefeld, Germany), and plotted, on a logarithmic scale, as percent intensity of the intact full-length RNA from the same lane versus the positions of the corresponding 3'- or 5'-terminal nucleotides. Note the gradual drop in binding-competence over a stretch of 4–5 nt.

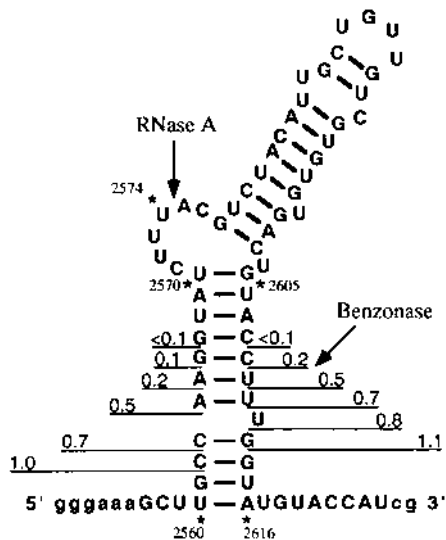
comparison with the total protein content of the cell lysate-based translation systems. For the mapping experiments described below, we used the reverse transcriptase, P protein, of duck hepatitis B virus (DHBV) and its cognate RNA, the structured encapsidation signal, Dε (Fig. 3). Here, in fact, *in vitro* translation is the only source for active, authentic protein (2), probably because the enzyme requires cellular chaperones for activity (3). DHBV is an avian representative of the hepadnavirus family which includes the medically important human hepatitis B virus (HBV). These small enveloped DNA-containing viruses replicate by reverse transcription of an RNA intermediate, the RNA pregenome (for review see 4,5). As shown by transfection experiments, encapsidation of the HBV pregenome into nucleocapsids (6) and initiation of reverse transcription (7) are mediated by the interaction of P protein (8) with the 5'-proximal ε-signal. ε forms a characteristic secondary structure with a lower and an upper stem, separated by a bulge of probably 6 nt (9–11). DHBV provides the distinct advantage that the interaction of its P protein with Dε, which is functionally similar to ε (12,13), can be studied using the *in vitro* translation system (2); the apparent  $K_d$  is in the range of 10–100 nM (14; Beck and Nassal, submitted). The experimentally determined secondary structure of Dε resembles that of ε (Beck and Nassal, submitted; Fig. 3); it was not clear, however, to what extent the entire structure is essential for forming an interaction-competent signal.

About 25 ng (~250 fmol) of the 87 kDa DHBV P protein can be produced per 25  $\mu$ l of reticulocyte lysate programmed with a recombinant P mRNA obtained from plasmid pT7AMVpol16

(15), as estimated by [<sup>35</sup>S]Met incorporation (16). Many smaller proteins are more efficiently expressed, but here the low amounts of P protein provided a rigorous test for the sensitivity of the method. For efficient separation of protein bound from free RNA, we modified the parental plasmid such that an N-proximal (His)<sub>6</sub>-tag was introduced (Fig. 1) that should exhibit a very high affinity to immobilized Ni<sup>2+</sup>-ions ( $K_d \sim 10^{-13}$  M; ref. 17), and hence allow for efficient protein immobilization even at very low concentrations. The tagged P protein has wild-type activity in an *in vitro* priming assay which measures the Dε-dependent covalent transfer to the protein of a dNMP from the corresponding [ $\alpha$ -<sup>32</sup>P]dNTP (data not shown).

*In vitro* transcribed Dε RNA (Fig. 3), either 5'- or 3'-terminally labeled with <sup>32</sup>P, was subjected to partial alkaline hydrolysis. The RNA pools were allowed to bind to *in vitro* translated P protein, and bound RNAs were separated via immobilization of the protein to Ni<sup>2+</sup>-NTA agarose (see legend of Fig. 2 for details). Size-comparison of free and bound transcripts by denaturing PAGE showed the expected ladders from the input material (Fig. 2A, lanes 1, 3, 5 and 6), while only the full-length products plus a limited number of shorter RNAs appeared in the bound fractions, with a cut-off extending over few positions (Fig. 2A, lanes 2 and 7).

For the 5'-labeled transcripts, of which a fraction of 4000 c.p.m. (~100 fmol) was bound specifically to the immobilized P protein, the cut-off position was determined using a previously mapped prominent Benzonase sensitive site at C2609 (Fig. 3). Quantification of the band intensities using a PhosphorImager (Fig. 2B, upper panel) showed a progressive decrease, from wt to zero levels, over



**Figure 3.** Minimal binding-competent Dε RNA. The sequence corresponds to that present in the *in vitro* transcribed Dε RNAs; lower case letters indicate vector-derived nucleotides; numbers are DHBV nucleotide positions. The secondary structure (Beck and Nassal, submitted) is predicted to be suboptimal by the m-fold program but is in good accord with the experimental data. The relative binding-competence of 5'- and 3'-labeled fragments is indicated by the bars and numbers (with the full-length RNA set at 1.0) at the left and right of the lower stem.

a stretch of 5 nt corresponding to the central part of the lower right half-stem. These data suggested that the three most basal base-pairs are not required for P protein binding.

This conclusion was confirmed with the 3'-labeled Dε transcripts of which ~1500 c.p.m. (again ~100 fmol) were specifically immobilized. Owing to the known 3'-heterogeneity of *in vitro* transcripts (18), three bands rather than one were observed in the control reaction with RNase A used for calibration (Fig. 2A, lanes 8 and 9). Binding-competence was gradually established (Fig. 2B, lower panel) with fragments having from 9 to 13 more nt at their 5'-end than the most prominent RNase A product starting with A2575 (Fig. 3). This corresponds to the left side of essentially the same lower stem region as defined by the 5'-labeled RNAs. Both datasets are hence consistent with a minimal requirement for 5 bp in the lower stem to produce detectable binding; provision of further base pairs enhances binding up to a plateau once the stem contains 8 bp (Fig. 3). This gradual increase probably reflects the increasing stability of the lower stem necessary to maintain the bulge-and-loop shape of the apical part. Secondary structure calculations using the program m-fold (19) support this view: terminal deletions such that <5 bp could form the lower stem strongly favor alternative structures. In turn, these data strongly support the importance of the bulge-and-loop structure in Dε.

That both Dε pools indicate essentially the same boundary suggests that the flanking 5'- and 3'-sequences have no major impact on P protein binding.

In summary, the method described here allows the simple, sensitive and accurate determination of the functional borders of protein-binding RNA elements using tagged, *in vitro* translated protein. Even nanogram amounts of the relatively poorly expressed hepadnaviral P protein proved sufficient for rapid data acquisition approaching single nucleotide resolution; a few thousand counts per minute of protein-bound RNA allowed for a clear distinction between binders and non-binders. Hence the procedure should work well with a whole variety of RNA binding proteins, many of which will conveniently, and in higher yield than P protein, be available by *in vitro* translation. The only requirement is an efficient method for immobilizing the protein, e.g. by specific antibodies, or any artificially introduced tag to which a high affinity capture reagent is available. With DHBV P protein, we have not observed major truncated *in vitro* translation products which could obscure the results. However, N-terminal tags, as in our case, discriminate against shortened proteins arising from internal translation initiation; potential problems due to C-terminal truncations will be circumvented by placing the tag at the C-terminus.

## ACKNOWLEDGEMENTS

We thank H. Schaller for critical discussions and continuous support, and H. Bartos for providing plasmids pT7AMVpol16 and pDε1. This work was supported by the Deutsche Forschungsgemeinschaft through SFB229-D3.

## REFERENCES

- Brown, D. and Gold, L. (1995) *Biochemistry*, **34**, 14765–14774.
- Wang, G.H. and Seeger, C. (1992) *Cell*, **71**, 663–670.
- Hu, J. and Seeger, C. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 1060–1064.
- Nassal, M. and Schaller, H. (1993) *Trends Microbiol.*, **1**, 221–226.
- Nassal, M. (1996) *Curr. Top. Microbiol. Immunol.*, **214**, 297–337.
- Junker-Niepmann, M., Bartenschlager, R. and Schaller, H. (1990) *EMBO J.*, **9**, 3389–3396.
- Nassal, M. and Rieger, A. (1996) *J. Virol.*, **70**, 2764–2773.
- Bartenschlager, R. and Schaller, H. (1992) *EMBO J.*, **11**, 3413–3420.
- Knaus, T. and Nassal, M. (1993) *Nucleic Acids Res.*, **21**, 3967–3975.
- Pollack, J.R. and Ganem, D. (1993) *J. Virol.*, **67**, 3254–3263.
- Rieger, A. and Nassal, M. (1995) *Nucleic Acids Res.*, **23**, 3309–3315.
- Wang, G.H. and Seeger, C. (1993) *J. Virol.*, **67**, 6507–6512.
- Tavis, J.E., Perri, S. and Ganem, D. (1994) *J. Virol.*, **68**, 3536–3543.
- Pollack, J.R. and Ganem, D. (1994) *J. Virol.*, **68**, 5579–5587.
- Weber, M., Bronsema, V., Bartos, H., Bosserhoff, A., Bartenschlager, R. and Schaller, H. (1994) *J. Virol.*, **68**, 2994–2999.
- Bartos, H. (1995) PhD thesis, University of Heidelberg.
- Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R. and Stüber, D. (1988) *BioTechnology*, **6**, 1321–1325.
- Ferré-D'Amaré, A.R. and Doudna, J.A. (1996) *Nucleic Acids Res.*, **24**, 977–978.
- Zuker, M. (1989) *Science*, **244**, 48–52.