

The chromosomal protein HMG-D binds to the TAR and RBE RNA of HIV-1

Paola B. Arimondo^a, Nathalie Gelus^a, François Hamy^b, Dominique Payet^{c,1},
Andrew Travers^c, Christian Bailly^{a,*}

^a*Institut National de la Santé et la Recherche Médicale, Unité 524, Place de Verdun, 59045 Lille Cedex, France*

^b*Pharma Research, Novartis Pharma Inc., CH-4002 Basle, Switzerland*

^c*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK*

Received 7 September 2000; accepted 6 October 2000

First published online 2 November 2000

Edited by Lev Kisselev

Abstract The high mobility group protein HMG-D is known to bind preferentially to DNA of irregular structures with little or no sequence specificity. Upon binding to DNA, this HMG-box protein widens the minor groove of the double helix and induces a significant bending of the helix. We show here that HMG-D can strongly bind to double-stranded RNA. Electrophoretic mobility shift assays show that HMG-D100 interacts with the transactivation response region (TAR) RNA from HIV-1. Strong interaction with a high affinity Rev protein binding element (RBE) RNA was also characterized. Gel shift experiments performed with several TAR RNA constructs lacking the lateral pyrimidine bulge or with modified apical loop regions indicate that the protein does not recognize the single-strand domains of the RNA but apparently interacts directly with the double-stranded stem regions. No protein–RNA complexes could be detected when using single-stranded oligoribonucleotides. HMG-D protein could bind to the wide minor groove of the A-form TAR RNA. The comparison of the amino acid sequence of HMG-D with that of known RNA binding proteins suggests that the interaction of the protein with a double-stranded RNA implicates the basic region of HMG-D as well as its HMG-box domain. From the *in vitro* data reported here, we propose a novel functional role for proteins of the HMG-1 family. The results suggest that architectural HMG proteins can be recruited by double-stranded RNA for the development of HIV-1 in the host cell. © 2000 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.

Key words: RNA binding; High mobility group protein; Transactivation response region RNA; Rev protein binding element RNA; Human immunodeficiency virus-1

1. Introduction

The transactivation response region (TAR) RNA sequence is located at the 5'-end of all nascent HIV-1 transcripts [1,2]. It adopts a hairpin secondary structure consisting of a highly

conservative 6 nucleotide (nt) loop and a 3 nt pyrimidine bulge (UCU or UUU at positions 23–25) flanked by two double-stranded stems [3,4] (Fig. 1). The pyrimidine bulge is essential for high affinity and specific binding to the *trans*-activator protein Tat [5]. Currently, it is believed that the major role of TAR is to act as an RNA enhancer that controls HIV-1 transcription at initiation and elongation. The functional Tat–TAR interaction *in vivo* requires cellular co-factors [6]. Cellular proteins capable of interacting directly with TAR include RNA polymerase II, the protein kinase PKR, its antagonist TRBP [7], Pur- α , the nuclear protein p140 [8], and the translation initiation factor 2, eIF2 [9]. This later protein, as well as TRBP, recognizes the apical loop of TAR. It is likely that many other proteins are recruited by TAR as the virus must require different host cell factors to achieve its specific gene regulation [10]. Identification of these proteins is essential for a better understanding of the mechanism responsible for the development of the virus in the host cell.

Binding of Tat protein to TAR is mediated by the nine amino acid region RKKRRQRRR [11–13]. Because similar Arg/Lys-rich peptide sequences are found in all high mobility group (HMG) proteins, we postulated that the TAR RNA could also interact with chromosomal proteins containing a HMG domain. To test this hypothesis, we chose HMG-D which is one of the *Drosophila melanogaster* counterparts of the abundant chromosomal protein HMG-1 [14,15]. This protein with a single HMG domain binds to the minor groove of double-stranded DNA with minimal sequence specificity [16], preferring duplex DNA that is deformable, 'pre-bent' and/or underwound [17,18]. The structure of the free HMG-D protein in solution shows an L-shaped fold with three α helices stabilized by two hydrophobic cores [19]. Recently, the X-ray crystal structure of the HMG domain of HMG-D complexed with the linear palindromic decamer GCGATATCGC was solved [20]. Binding of the protein to the minor groove of the DNA duplex induces a bend of the double helix away from the protein towards the major groove. The structure also revealed multiple intercalations and water-mediated DNA contacts [21]. The structural basis of DNA binding by HMG proteins is now relatively well understood [22]. In sharp contrast, to our knowledge, there is as yet no information available on the potential interaction of HMG proteins with RNA.

As an initial approach, we report here the results of gel shift experiments showing that HMG-D binds to the TAR RNA

*Corresponding author. Fax: (33)-320-16 92 29.
E-mail: bailly@lille.inserm.fr

¹ Present address: CIML, Parc de Luminy, 13288 Marseille Cedex 09, France.

Abbreviations: HMG, high mobility group; nt, nucleotide; RBE, Rev protein binding element; TAR RNA, transactivation response region RNA

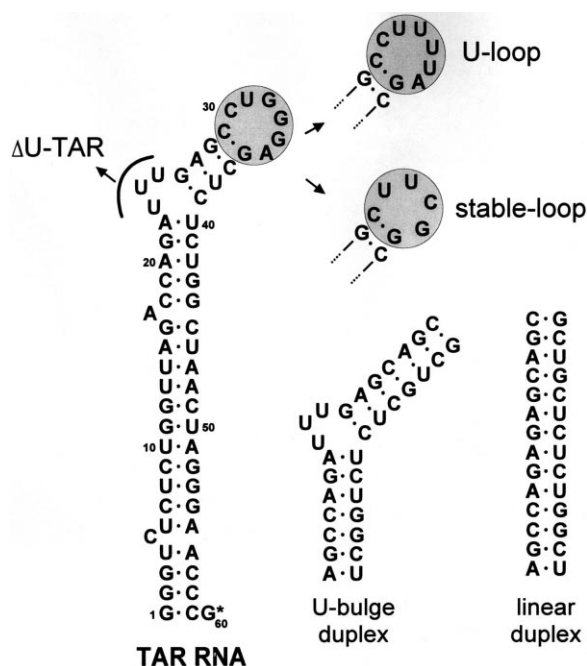


Fig. 1. Sequences and secondary structures of the wild type TAR RNA and the mutants used in this study. The uridine bulge is missing in the ΔU-TAR RNA.

hairpin structure. The use of several TAR RNA constructs and a 57 nt RNA corresponding to the high affinity Rev protein binding element (RBE), indicates that HMG-D interacts with the double-stranded region of the RNA. The results suggest that architectural HMG proteins are recruited by leader RNA for the development of the HIV-1 in the host cell.

2. Materials and methods

2.1. HMG-D protein

HMG-D contains a single HMG domain followed by a basic region and an acidic tail. In the experiments presented here, we used the HMG-D₁₀₀ construct which corresponds to the first 100 amino acids, that is the HMG domain plus the basic region but lacking the acidic tail [17]. The HMG-D₁₀₀ protein, overproduced in *Escherichia coli*, was purified according to a standard procedure [16]. It was stored at -20°C in 50 mM HEPES buffer pH 7.5 containing 1 mM EDTA and 50% glycerol. All other chemicals were analytical grade reagents, and solutions were prepared with doubly distilled sterile water to prevent from nuclease contamination. Tubes and tips were treated with 1% diethylpyrocarbonate (DEPC from Sigma).

2.2. In vitro transcription of RNA

Synthetic oligonucleotides corresponding to the wild type TAR and mutated TAR sequences were cloned between *Hind*III and *Eco*RI sites of the pUC19 plasmid. After digestion with *Eco*RI, the RNA was transcribed as a run-off product of 60 (wild type TAR and U loop TAR), 58 (UUCG stable loop TAR) and 57 (ΔU-TAR) nts from the T3 RNA polymerase promoter. In each case the transcript includes an additional G residue on the 3'-end derived from the *Eco*RI cleavage site. Transcription reaction was performed in buffer containing 40 mM Tris-HCl, pH 7.4, 25 mM NaCl, 16 mM MgCl₂, 10 mM DTT and 1 mM NTPs. The reaction was initiated by the addition of 10 μg linearized plasmid DNA template and 40 μg T3 RNA polymerase and incubated for 2 h at 37°C . Nucleic acids were then fractionated on a 10% (w/v) polyacrylamide gel containing 8 M urea in TBE buffer (89 mM Tris-borate pH 8.3, 10 mM EDTA). After electrophoresis, the RNA was eluted in water for 18 h at 4°C and precipitated with ethanol. The RNA was resuspended in water to give a 500 μM stock solution ($\epsilon^{260}/\text{phosphate} = 10688 \text{ M}^{-1} \text{ cm}^{-1}$). The RNA was 3'-end

labeled with [^{32}P]cytidine biphosphate and T4 RNA ligase and then repurified from a 10% denaturing acrylamide gel [13]. The 14 and 17 base oligoribonucleotides were purchased from Genset (Paris) and purified on a 15% denaturing acrylamide gel prior to the labelling with P^{32} . A published procedure was followed to prepare the 57 nt RBE RNA [23].

2.3. Gel mobility shift assay

Binding reactions (20 μl) contained the ^{32}P -labeled RNA (2 μl, about 15 nM) and varying concentrations of the HMG-D protein in a TK buffer consisting of 10 mM Tris-HCl (pH 7.4), 50 mM NaCl, 1 mM EDTA, 5 mM DTT and 0.1% Triton X-100. Samples were incubated for at least 10 min on ice prior to the addition of 2 μl loading mix and electrophoresis on a 8% non-denaturing polyacryl-

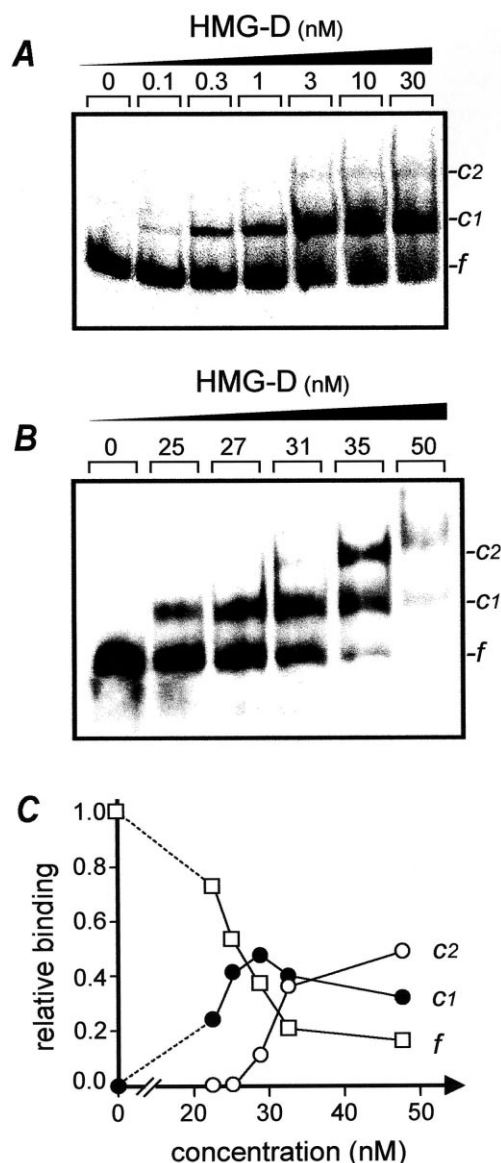


Fig. 2. Electrophoretic mobility shift assay of HMG-D binding to the TAR RNA at (A) 20°C or (B) 4°C . The RNA was 3'-end labeled with [^{32}P]cytidine biphosphate and T4 RNA ligase. An estimated 100 pM of ^{32}P -60-nt RNA was titrated with the indicated concentrations (nM) of the protein. The two HMG-TAR complexes and free RNA are identified as c_1 , c_2 and f , respectively. Complexes were separated from unbound RNA by electrophoresis in non-denaturing 8% polyacrylamide gels containing 0.1% Triton X-100 in $1 \times$ TBE buffer. C: The binding plots show the concentration dependence for the formation of the c_1 and c_2 complexes and the decrease of the unbound RNA form f .

amide gel containing 0.1% Triton X-100 and $1\times$ TBE buffer. After about 1.5 h electrophoresis at 200 V and at 4°C, the gel was dried and analyzed on a Phosphor-imager (Molecular Dynamics).

2.4. RNase A footprinting, gel electrophoresis and data processing

The procedure for the footprinting experiments was adapted from published protocols [24]. Briefly, samples of the labelled RNA fragment were incubated with a buffered solution containing the desired protein concentration. After 20 min incubation at 4°C to ensure equilibration, the digestion was initiated by addition of the RNase A solution. After 1 min incubation at room temperature, the reaction was stopped by freeze drying and samples were lyophilized. The RNA in each tube was resuspended in 5 μ l of formamide–TBE loading buffer, denatured at 90°C for 4 min then chilled in ice for 4 min prior to loading on to a 0.3 mm thick, 10% polyacrylamide gel containing 8 M urea and TBE buffer (89 mM Tris base, 89 mM boric acid, 2.5 mM Na₂ EDTA, pH 8.3). After 2 h electrophoresis at 1500 V, the gel was soaked in 10% acetic acid for 10 min, transferred to Whatman 3MM paper, dried under vacuum at 80°C and then analyzed on the phosphorimager (Molecular Dynamics). Each resolved band on the autoradiograph was assigned to a particular bond within the RNA fragment by comparison of its position relative to sequencing standards generated by treatment of the RNA with diethylpyrocarbonate followed by aniline-induced cleavage at the modified bases (A track).

3. Results

A simple electrophoretic mobility shift assay was employed to investigate the interaction of the HMG-D protein with the wild type TAR RNA prepared by *in vitro* transcription. Different amounts of purified protein were incubated with a fixed amount of radiolabeled RNA in a standard binding buffer and the mixture of free and protein-bound RNA was resolved on a non-denaturing polyacrylamide gel (Fig. 2). We found that HMG-D forms two complexes with the RNA suggesting the presence of at least two binding sites within the 60 nt sequence. The complex containing at least two proteins per RNA molecule was difficult to visualize when the binding reaction and electrophoresis were performed at room temperature (Fig. 2A). Lowering the temperature greatly helped to distinguish the c2 complex but in this case much of the protein–DNA complexes did not enter the gel for [HMG-D] = 50 nM (Fig. 2B). This first set of experiments provides direct evidence that the HMG protein interacts strongly with the TAR RNA as judged from the formation of protein–DNA complexes at concentrations as low as 0.3 nM. Several gels as in Fig. 2B were quantitated by densitometry (Fig. 2C) and a dissociation constant, K_d of 29 ± 4 nM was determined from the binding plots.

To identify a potential HMG-D binding site on the TAR RNA, we probed the RNA and RNA–drug complexes with RNase A. Since the susceptibility of RNA to attack by RNase A is dependent on secondary structure rather than on primary sequence, this enzyme is not the most appropriate nuclease to monitor protein binding to the two double-stranded stem regions of the RNA. However, this enzyme is particularly well suited for investigating drug binding to the single strand parts, including the bulge and loop regions. The cleavage patterns are shown in Fig. 3. In the presence of 50 nM HMG-D, the strong cleavage at positions 37–38 is largely reduced whereas the cutting at position 29 is slightly enhanced. These footprinting experiments indicate that the reactivity of the nt residues near the apical loop (CCU at positions 29–31) is apparently not affected by the protein binding but in contrast, the HMG protein may interact with the CUC 37–39 sequence opposite the UUU bulge. RNase A detects a concerted change

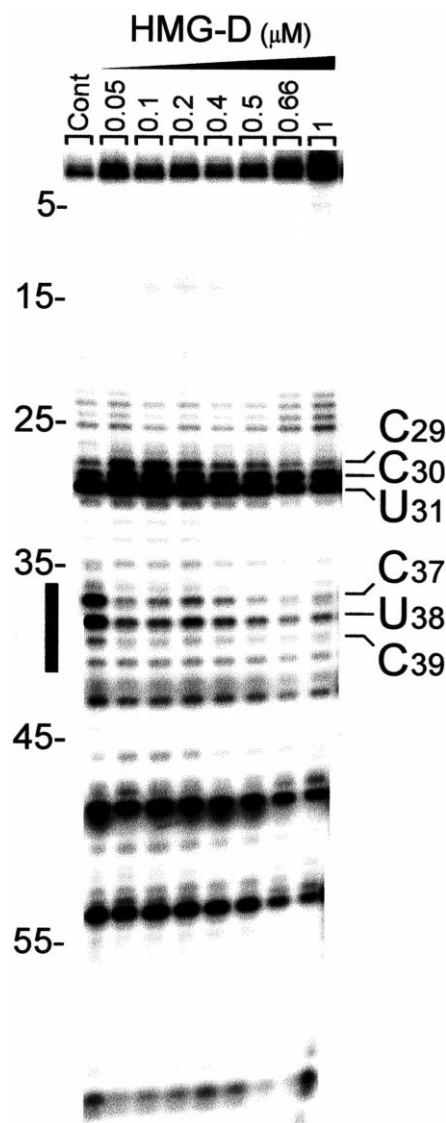


Fig. 3. RNase A footprinting of HMG-D on wild type TAR. The cleavage products of RNase digestion were resolved on a 10% polyacrylamide gel containing 8 M urea. The concentration (μ M) of the protein is shown at the top of the appropriate gel lanes. Control track labeled 'Cont' contained no protein. The vertical black bar refers to the site of reduced RNase A cleavage induced by HMG-D.

in the helix conformation as a result of binding of HMG-D to the TAR RNA. Initially, we interpreted the base protection effect as being due to direct interaction with the protein but, on the basis of the subsequent experiments (see below), the protection could arise from protein-induced structural changes.

The uridine bulge is an important contributor to the structural stability of TAR and is absolutely essential for the interaction with the Tat protein [25]. According to the above footprinting experiments, the uridine bulge may also play an important role in the binding of HMG-D to the RNA. To verify this idea, we elaborated an analogue lacking the pyrimidine bulge and the resulting 57 nt Δ U-TAR RNA was used as a substrate for HMG-D. As shown in Fig. 4A, the deletion of the UUU bulge does not prevent HMG-D binding to the RNA. Both the c1 and c2 complexes seen with the wild type

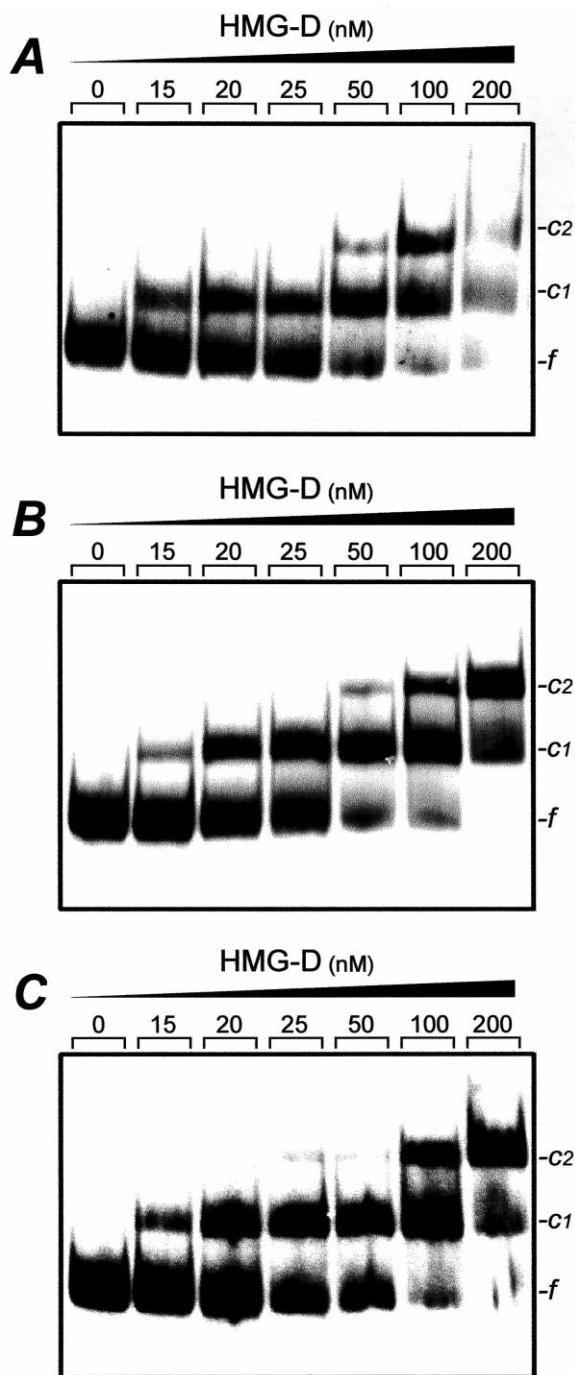


Fig. 4. Electrophoretic mobility shift assay of HMG-D binding to the mutants TAR RNA: (A) Δ U-TAR, (B) U loop TAR and (C) UUCG stable loop RNA. Complexes were separated from unbound RNA by electrophoresis in non-denaturing 8% polyacrylamide gels containing 0.1% Triton X-100 in $1\times$ TBE buffer at 4°C . Other details as for Fig. 2.

RNA were detected with Δ U-TAR and the measured dissociation constant was almost unchanged ($K_d = 36 \pm 3$ nM). Therefore, the pyrimidine bulge of the TAR RNA is not necessary for HMG-D binding.

Two additional TAR mutants were prepared by in vitro transcription. Firstly, we substituted the UGGG upper loop of TAR, which provides a binding site for certain proteins such as TRBP (7), for a UUUU loop. However, here again,

the interaction of HMG-D with the RNA was not significantly affected as the two protein–RNA complexes were detected with the same range of protein concentrations (Fig. 4B). Secondly, we replaced the 6 nt upper loop of TAR by a shorter 4 nt loop. The UUCG tetrad was substituted for the CUGGGA sequence of TAR. The loop sequence UUCG occurs exceptionally often in ribosomal and other RNAs, and is believed to serve as a nucleation site for RNA folding and as a protein recognition site [26]. A hairpin RNA containing this loop displays unusually high thermodynamic stability and the structure of the loop was solved by nuclear magnetic resonance [27]. We reasoned that a TAR RNA containing such a potentially highly stable loop may offer a stronger binding site for HMG-D. The corresponding 58 nt TAR RNA containing the UUCG loop (Fig. 1) was synthesized by in vitro transcription, radiolabeled and then used as a substrate for

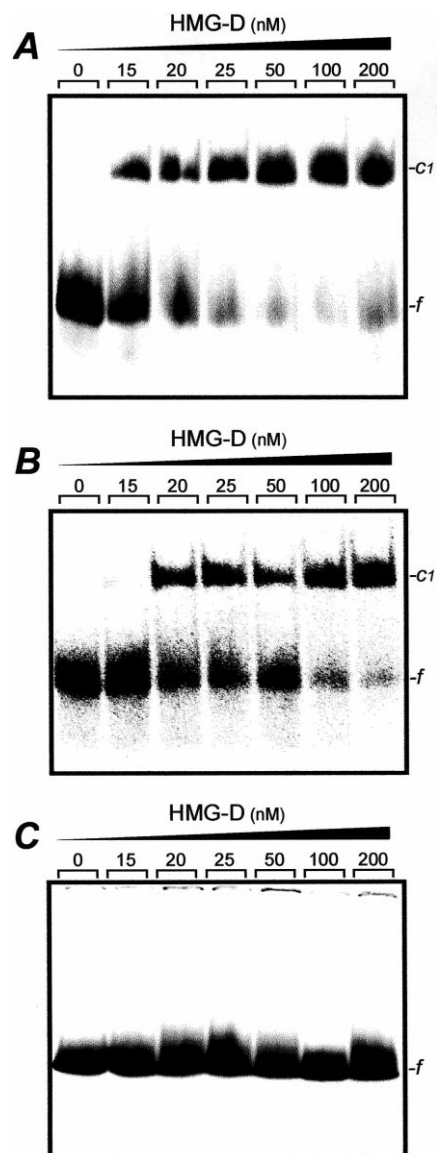


Fig. 5. Electrophoretic mobility shift assay of HMG-D binding to (A) the U bulge and (B) the linear oligoribonucleotide duplexes (sequences shown in Fig. 1). Complexes were separated from unbound RNA by electrophoresis in non-denaturing 10% polyacrylamide gels containing 0.1% Triton X-100 at 4°C . The lower panel (C) shows that HMG-D does not bind to the 14 bases single-stranded oligoribonucleotide. Other details as for Fig. 2.

limit precisely the mode of interaction between double-stranded RNA and HMG-D (and probably related HMG-1 type proteins), as well as the consequences of this interaction in terms of structures and functions. This study opens a completely new avenue for HMG proteins.

The experiments presented in this study were all performed with purified RNA species prepared by *in vitro* transcription and use the truncated HMG-D100 protein lacking the 12 residue acidic tail of the full length HMG-D. This C-terminal domain is known to decrease the affinity of HMG-D for linear DNA and confers selectivity for deformed DNA structures [17]. The same trend may apply in the RNA context. Here we show that HMG-D100 has strong interactions with the duplex regions of TAR; it is possible that the full length protein rather prefers the more distorted regions of the RNA. This question is open to experimental probing.

Another key question concerns the relevance of this *in vitro* study. Do the binding of HMG-D to duplex RNA occur *in vivo*? This is of course an essential point to address but at present there are reasons to believe that this interaction can take place in cells. Bianchi et al. have reported that HMG-1 protein is not stably associated with the chromosomes of somatic cells [32]. This protein appears to be uniformly distributed in the cell nucleus. Apart from their well recognized architectural roles at the DNA level, HMG-1 proteins could act as copartner for polymerases and other proteins acting during transcription. There are several examples of proteins which interact with both double-stranded DNA and RNA, such as TFIIIA. This is also the case for histone proteins [33]. The eukaryotic Y-box proteins [34] also bind both RNA and DNA but in this case the single-stranded form is the preferred ligand. Double-stranded RNA binding proteins such as p68 kinase [35] usually contain scattered basic (both Arg and Lys) amino acids which are essential for RNA recognition [36]. HMG-D contains a stretch of basic amino acids commonly found in known RNA binding proteins including Tat and Rev. The basic region of HMG-D (amino acids 80–100) flanking the HMG domain presents a noticeable degree of similarity with the RNA binding domain of Tat. The sequence KKRAKPAKKVAKSKK (from amino acid positions 85–100) closely resembles the Arg/Lys sequence found in the 55–70 region of Tat (30.8% identity in a 13 amino acid overlap). Moreover, the underlined sequence in HMG-D is analogous to the basic motif KKLAKRN which represents the essential part of the TAR RNA binding domain of the TRBP protein [7]. Similar Arg/Lys peptides also form complexes with TAR [37]. *A posteriori*, it is therefore not entirely surprising to observe complex formation between HMG-D and the TAR RNA. The interaction of the protein with a double-stranded RNA may implicate the basic region of HMG-D as well as its HMG-box domain. Continued biochemical and structural analyses will tell us how the HMG-D protein interacts with duplex RNA and, hopefully, will allow us to increase our understanding of the functional role of this previously unexpected and newly discovered interaction. Finally, the results also suggest that HMG-1 type proteins may be recruited by the HIV-1 via its leader RNA. Binding of HMG proteins to the TAR element of HIV mRNAs may influence Tat-mediated activation. The virus likely needs architectural proteins such as HMG-1 in order to fully exploit the cell machinery to replicate into infected cells.

Acknowledgements: This work was supported by a research grant (to C.B.) from the SIDACTION. P.B.A. is currently supported by a fellowship from the Ligue Nationale Française Contre le Cancer.

References

- [1] Jones, K.A. and Peterlin, B.M. (1994) *Annu. Rev. Biochem.* 63, 717–743.
- [2] Berkhout, B. (1996) *Prog. Nucleic Acid Res. Mol. Biol.* 54, 1–34.
- [3] Aboul-ela, F., Karn, J. and Varani, G. (1995) *J. Mol. Biol.* 253, 313–332.
- [4] Aboul-ela, F., Karn, J. and Varani, G. (1996) *Nucleic Acids Res.* 24, 3974–3981.
- [5] Watson, K. and Edwards, R.J. (1999) *Biochem. Pharmacol.* 58, 1521–1528.
- [6] Gagnon, A., Duarte, M., Daviet, L., Chang, Y.N. and Jeang, K.T. (1996) *Gene Express.* 5, 217–228.
- [7] Erard, M., Barker, D.G., Amalric, F., Jeang, K.-T. and Gagnon, A. (1998) *J. Mol. Biol.* 279, 1085–1099.
- [8] Rothblum, C.J., Jackman, J., Mikovits, J., Shukla, R.R. and Kumar, A. (1995) *J. Virol.* 69, 5156–5163.
- [9] Ben-Asouli, Y., Banai, Y., Hauser, H. and Kaempfer, R. (2000) *Nucleic Acids Res.* 28, 1011–1018.
- [10] Rosen, C.A. (1991) *Trends Genet.* 7, 9–14.
- [11] Weeks, K.M., Ampe, C., Schultz, S.C., Steitz, T.A. and Crothers, D.M. (1990) *Science* 249, 1281–1285.
- [12] Hamy, F., Asselyne, U., Grasby, J., Iwai, S., Pritchard, C., Slim, G., Butler, P.J.G., Karn, J. and Gait, M.J. (1993) *J. Mol. Biol.* 230, 111–123.
- [13] Churcher, M.J., Lamont, C., Hamy, F., Dingwall, C., Green, S.M., Lowe, A.D., Butler, P.J.G., Gait, M.J. and Karn, J. (1993) *J. Mol. Biol.* 230, 90–110.
- [14] Wagner, C.R., Hamana, K. and Elgin, S.C.R. (1992) *Mol. Cell. Biol.* 12, 1915–1923.
- [15] Ner, S.S., Churchill, M.E.A., Searles, M.A. and Travers, A.A. (1993) *Nucleic Acids Res.* 21, 4369–4371.
- [16] Churchill, M.E.A., Jones, D.N.M., Glaser, T., Hefner, H., Searles, M.A. and Travers, A.A. (1995) *EMBO J.* 14, 1264–1275.
- [17] Payet, D. and Travers, A.A. (1997) *J. Mol. Biol.* 266, 66–75.
- [18] Churchill, M.E.A., Changela, A., Dow, L.K. and Krieg, A.J. (1999) *Methods Enzymol.* 304, 99–133.
- [19] Jones, N.M., Searles, M.A., Shaw, G.L., Churchill, M.E.A., Ner, S.S., Keeler, J., Travers, A.A. and Neuhaus, D. (1994) *Structure* 2, 609–627.
- [20] Murphy IV, F.V., Sweet, R.M. and Churchill, M.E.A. (1999) *EMBO J.* 18, 6610–6618.
- [21] Murphy IV, F.V. and Churchill, M.E.A. (2000) *Curr. Opin. Struct. Biol.* 8, R83–R89.
- [22] Travers, A. (2000) *Curr. Opin. Struct. Biol.* 10, 102–109.
- [23] Heaphy, S., Finch, J.T., Gait, M.J., Karn, J. and Singh, M. (1991) *Proc. Natl. Acad. Sci. USA* 88, 7366–7370.
- [24] Weeks, K.M. and Crothers, D.M. (1991) *Cell* 66, 577–588.
- [25] Gait, M.J. and Karn, J. (1993) *Trends Biochem. Sci.* 18, 255–259.
- [26] Cheong, C. and Varani, G. (1990) *Nature* 346, 680–682.
- [27] Varani, G. and Cheong, C. (1991) *Biochemistry* 30, 3280–3289.
- [28] Hung, L.-W., Holbrook, E.L. and Holbrook, S.R. (2000) *Proc. Natl. Acad. Sci. USA* 97, 5107–5112.
- [29] Ner, S.S. and Travers, A.A. (1994) *EMBO J.* 13, 1817–1822.
- [30] Payet, D., Hillisch, A., Lowe, N., Diekmann, S. and Travers, A. (1999) *J. Mol. Biol.* 294, 79–91.
- [31] Lorenz, M., Hillisch, A., Payet, D., Buttinelli, M., Travers, A. and Diekmann, S. (1999) *Biochemistry* 38, 12150–12158.
- [32] Falcioni, L., Spada, F., Calogero, S., Langst, G., Voit, R., Grummt, I. and Bianchi, M.E. (1997) *J. Cell Biol.* 137, 19–26.
- [33] Jacobs, F. and Imani, B.L. (1988) *J. Interferon Cytokine Res.* 8, 821–830.
- [34] Hayman, M.L. and Read, L.K. (1999) *J. Biol. Chem.* 274, 12067–12074.
- [35] Feng, G.-S., Chong, K., Kumar, A. and Williams, B.R.G. (1992) *Proc. Natl. Acad. Sci. USA* 89, 5447–5451.
- [36] Burd, C.G. and Dreyfuss, G. (1994) *Science* 265, 615–621.
- [37] Calnan, B.J., Biancalana, S., Hudson, D. and Frankel, A.D. (1991) *Genes Dev.* 5, 201–210.