The extended AT-hook is a novel RNA binding motif

Michael Filarsky[1,2,†], Karina Zillner[1,†], Ingrid Araya[1,†], Ana Villar-Garea[3,4], Rainer Merkl[5], Gernot Längst[1,*], and Attila Németh[1,*]

[1]Biochemistry Center Regensburg; University of Regensburg; Regensburg, Germany; [2]Medical Parasitology & Infectious Biology; Swiss Tropical and Public Health Institute; Basel, Switzerland; [3]Munich Center of Integrated Protein Science and Adolf-Butenandt Institute; Ludwig Maximilians University of Munich; Munich, Germany; [4]Biochemistry Center Regensburg; University of Regensburg; Regensburg, Germany; [5]Institute of Biophysics and Physical Biochemistry; University of Regensburg; Regensburg, Germany

[†]These authors equally contributed to this work.

The AT-hook has been defined as a DNA binding peptide motif that contains a glycine-arginine-proline (G-R-P) tripeptide core flanked by basic amino acids. Recent reports documented variations in the sequence of AT-hooks and revealed RNA binding activity of some canonical AT-hooks, suggesting a higher structural and functional variability of this protein domain than previously anticipated. Here we describe the discovery and characterization of the extended AT-hook peptide motif (eAT-hook), in which basic amino acids appear symmetrical mainly at a distance of 12–15 amino acids from the G-R-P core. We identified 80 human and 60 mouse eAT-hook proteins and biochemically characterized the eAT-hooks of Tip5/BAZ2A, PTOV1 and GPBP1. Microscale thermophoresis and electrophoretic mobility shift assays reveal the nucleic acid binding features of this peptide motif, and show that eAT-hooks bind RNA with one order of magnitude higher affinity than DNA. In addition, cellular localization studies suggest a role for the N-terminal eAT-hook of PTOV1 in nucleocytoplasmic shuttling. In summary, our findings classify the eAT-hook as a novel nucleic acid binding motif, which potentially mediates various RNA-dependent cellular processes.

Introduction

Nucleic acid binding domains are essential in protein-mediated regulation of DNA- and RNA-dependent biological processes. The AT-hook belongs to this largely diverse group of protein domains and was originally described as a short, 3 times repeated peptide motif in the HMGA1 (High mobility group AT-hook 1) protein.[1] It binds to the minor groove and alters the architecture of DNA, which facilitates binding of other proteins in the opposite major groove.[2] The AT-hook motif is a characteristic sequence feature of HMGA proteins and distinguishes them from the HMGB (HMG box) and HMGN (HMG nucleosome-binding domain) proteins.[3] Based on the biochemical and structural characterization of HMGA1s AT-hooks,[1,2] the G-R-P tripeptide motif has been considered as the core sequence of this short protein domain. Almost two decades ago a systematic search for the G-R-P motif in protein sequence databases led to the discovery of a wide variety of AT-hook proteins, including numerous non-HMG proteins.[4] Multiple sequence alignments of all AT-hooks identified a 10–15 amino acids long peptide motif, in which the G-R-P core motif is mainly surrounded by K and R residues. Notably, most of the eukaryotic AT-hook proteins are associated with chromatin regulatory functions, such as chromatin remodelling, histone modifications or chromatin insulator function. Therefore the AT-hooks were proposed to anchor chromatin-modifying proteins to AT-rich DNA sequences.

Importantly, several reports suggest a more relaxed definition for the AT-hook sequence motif than originally proposed. Most strikingly, 2 conserved AT-hook-like motifs have been identified in the β′ subunits of DNA-dependent RNA polymerases (DNGRRGAPVTNPGSDRPLRSL in *Thermus thermophilus*; the GRP-like motifs are underlined). Based on the ubiquity of this sequence motif in the β' subunits of a wide variety of species, it has been suggested that AT-hook-like domains functioned as nucleic acid recognition peptides already in the last universal common ancestor of the present life forms, or even earlier.[5] Moreover, a careful investigation of Drosophila TAF1 (TATA-binding protein-associated factor 1), the only protein in this species with detectable canonical AT-hook, revealed that the replacement of glycine and proline residues of the G-R-P core motif does not interfere markedly with the DNA binding activity of this protein domain. This led to the assumption that AT-hook-like motifs, which do not perfectly match to the consensus AT-hook sequence, might exist in the fruit fly proteome. Indeed, 5 AT-hook-like motifs (KRGRGRARK, EKKPGRKRK, VKRMRRSVR, PKKRGRKAK, and PKRSKRRIK) have been finally identified in the Drosophila homolog of the G9a histone methyltransferase enzyme.[6] Additional examples for non-canonical AT-hook-like domain proteins include the AKNA (AT-hook-containing transcription factor), PSR/JMJD6 (Phosphatidylserine receptor, Jmj-domain containing protein D6), CaATL1 (Capsicum annuum AT-hook-Like gene 1), TdIF1 (Terminal deoxynucleotidyltransferase-interacting factor 1), LEDGF/p75 (Lens epithelium-derived growth factor 75) and MeCP2 (Methyl-CpG-binding protein 2) proteins.[7-12] Remarkably, LEDGF/p75 and MeCP2 contain also canonical AT-hooks besides their AT-hook-like domains. These studies revealed largely different non-canonical AT-hooks displaying alternative amino acid sequences in the close proximity of the G-R-P core motif, or even within the motif. Notably, these AT-hook-like sequence motifs are similar in size to the canonical AT-hook, i.e. 10–15 amino acids long, and they point to the existence of a large number of so far unrecognized DNA binding domains, which likely perform similar functions to the AT-hooks.

We have recently analyzed the AT-hook domains of the Tip5 (TTF-I interacting protein 5) chromatin regulator protein, the large subunit of the Nucleolar Remodelling Complex.[13] The DNA binding affinities of the 4 AT-hooks of Tip5 were determined *in vitro*, and the nucleolar targeting function of the first 2 AT-hooks was reported. While studying Tip5s AT-hooks, we have found an additional G-R-P peptide motif within the Tip5 sequence, which is flanked by basic amino acid patches at a larger distance than the G-R-P core of a canonical AT-hook. Thus, we named the new sequence motif extended AT-hook (eAT-hook). This novel protein domain is approximately 3 times longer than AT-hooks, and the basic amino acids are spaced apart from the G-R-P core in it. Here we show the characterization of the nucleic acid binding activities of Tip5s eAT-hook, define the eAT-hook sequence motif and identify human and mouse proteins with this sequence feature. Finally, we show that the eAT-hooks of 3 different proteins bind RNA with one order of magnitude higher affinity than DNA and therefore suggest a possible

function for the eAT-hook in diverse nucleic-acid-dependent cellular processes.

Results

The eAT-hook of Tip5 exhibits DNA binding activity

In addition to the 4 canonical AT-hooks, we have identified in the amino acid sequence of Tip5 another G-R-P tripeptide motif, which resembles the core of the AT-hook domain (**Fig. 1A**). The additional G-R-P tripeptide is flanked by basic amino acid patches located at a 6–20 amino acids distance, which is larger compared to their typical 1–5 amino acids distance detected in canonical AT-hooks. This observation led us to speculate about the existence of a new kind of DNA binding domain, which we termed extended AT-hook (eAT-hook). This description refers to the patches of basic amino acids located distant from the core motif. To experimentally test that Tip5s eAT-hook is able to bind DNA, we first cloned, expressed and purified a GST-tagged version of this protein domain (eAT), furthermore point mutant (eATmut) and deletion mutant (eATcore) versions of it. In the eATmut construct 2 arginine residues were mutated to aspartic acid. These point mutations should interfere with the DNA binding as described previously for an AT-hook domain.[14] The eATcore deletion mutant was truncated at its N- and C-termini, removing the basic amino acid patches of the peptide, but the core motif was not mutated (**Fig. 1B**). Since Coomassie-stained gels of different purifications of the eAT-hook consistently showed a dominant slower migrating and a secondary faster migrating protein, both proteins were subjected to mass spectrometry analysis prior to performing functional studies. MALDI-TOF measurements identified the larger protein as the full-length GST-tagged eAT-hook, and the smaller one as its C-terminally truncated version (**Fig. S1**).

After assessing the purity of the isolated protein, microscale thermophoresis (MST) experiments were performed with GST-tagged Tip5 eAT-hook peptides. A DNA fragment from the enhancer region of the mouse rDNA ("rDNA En"), which has been shown to be bound efficiently by the canonical AT-hooks of Tip5,[13] was used as binding partner at a constant concentration of 50 nM. Note that the same oligonucleotide concentration was used in all subsequent MST experiments to allow direct comparison of the results. The equilibrium binding constant concentration value (EC50), i.e., the eAT-hook concentration at which 50% of the DNA is bound by the protein, was calculated from MST data, resulting in 10 ± 0.8 μM. In contrast, none of the eAT-hook mutants did measurably bind to this DNA fragment (**Fig. 1C**).

To verify the DNA binding activity of Tip5s eAT-hook, gel shift assays were performed. We included here a peptide containing the first and second AT-hook of Tip5 as positive control. Two other AT-rich sites, namely an "rDNA IGS" fragment and the previously characterized HMGA1 binding site of the interferon β ("IFNβ") promoter were selected in addition to the "rDNA En" fragment. The results clearly show that the eAT-hook binds to all 3 DNA fragments with similar affinity and the observed
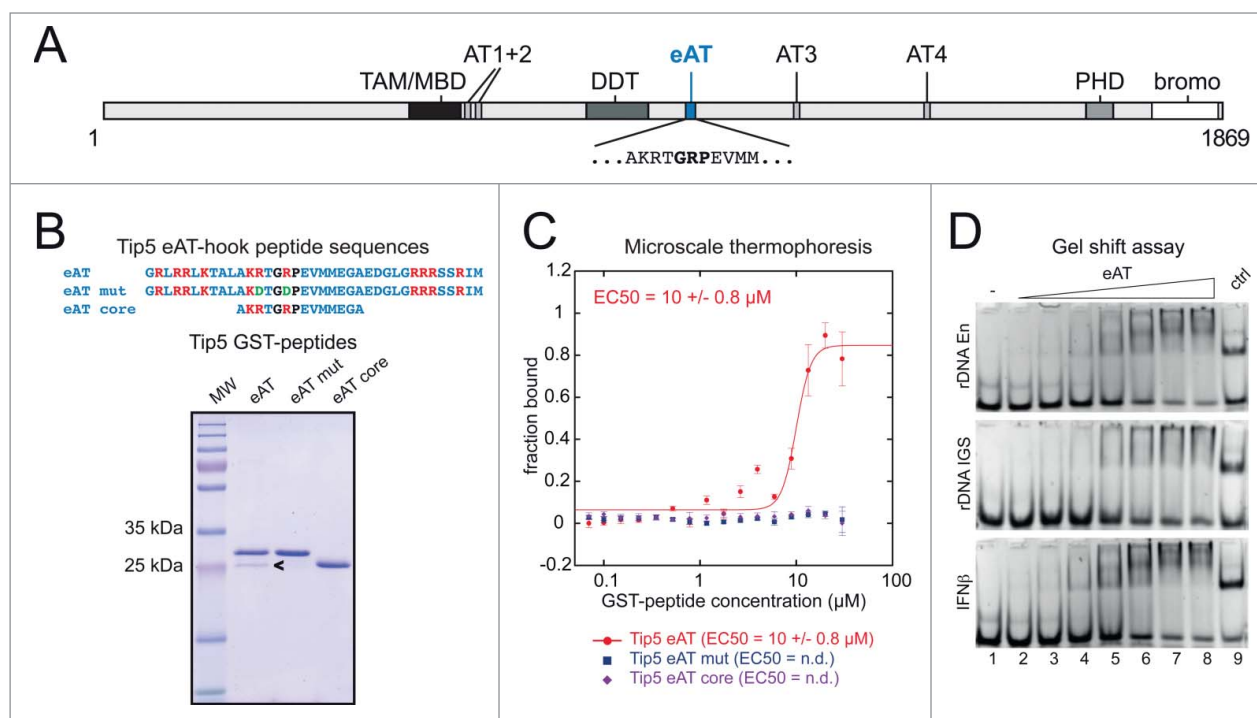
Figure 1. DNA binding features of Tip5s eAT-hook domain. (**A**) Domain structure of Tip5, the large subunit of the NoRC complex. The extended AT-hook is labeled with eAT in blue; TAM domain, Tip5-ARBP-MeCP1 domain; ATx, AT-hook; DDT, domain in different transcription and chromosome remodeling factors; PHD, plant homeodomain zinc-finger; bromo, bromodomain. The numbers below the scheme indicate amino acids and the core sequence of the eAT-hook is shown. (**B**) Amino acid sequences of Tip5 eAT-hook peptides are shown on the top. Red letters mark positively charged amino acids, green letters indicate R-to-D mutations, and core motif G and P amino acids are labeled in black. Coomassie gel image of purified GST-tagged Tip5 eAT-hook peptides is shown below the sequences. The arrowhead indicates a C-terminally truncated version of GST-Tip5-eAT. (**C**) Microscale thermophoresis (MST) experiments with Tip5 eAT-hook peptides. The indicated amounts of the different peptides were analyzed for DNA binding at constant, 50 nM concentration of a 29 bp, Cy5-labeled "rDNA En" oligonucleotide. The fraction of bound DNA was plotted against the peptide concentration, and the concentration where half of the oligonucleotide is bound, the EC50 value, was determined. The binding data of Tip5-eATmut and Tip5-eATcore peptides were obtained in parallel with Tip5-eAT and normalized to the difference between the minimum and maximum values of the Tip5-eAT binding experiment. The data points of the plot represent mean±standard deviation values of 4 measurements. (**D**) Gel shift assay analyses of Tip5-eAT-hook DNA binding. Lane 1 shows in the upper panel an rDNA enhancer ("rDNA En") probe, in the middle panel an rDNA intergenic spacer ("rDNA IGS") probe and in the lower panel an IFNβ promoter ("IFNβ") probe alone. Increasing amounts (9, 18, 36, 72, 144, 216, 252 pmol) of GST-Tip5-eAT were added to 1.25 pmol DNA and 72 pmol GST-Tip5-AT1+2 was used as positive control in lane 9.

supershifts suggest that more than one eAT-hook molecule binds to one DNA molecule at higher peptide concentrations (**Fig. 1D**).

The eAT-hook is a novel putative nucleic acid binding domain

In order to identify sequences which are similar to Tip5s eAT-hook, peptide motifs that contain more than 3 basic amino acid residues on both sides of the G-R-P core motif in a distance of 6–20 amino acids, i.e. $(K/R)_{>3}$-X_{6-20}-GRP-X_{6-20}-$(K/R)_{>3}$ sequences, were systematically searched in the human and mouse proteome datasets of the UniProt database (www.uniprot.org). Similar to the sequence homology-based search for AT-hooks,[4] the position of the core G-R-P tripeptide was fixed in our sequence analysis. After removing duplicate hits, 80 and 60 different eAT-hooks were identified in the human and mouse proteomes, respectively (**Table S1**). Next we characterized these eAT-hooks using the pLogo motif visualization tool, which scales residue heights relative to their statistical significance.[15] Human and mouse eAT-hook sequences were compared to all G-R-P containing peptides of the

respective proteomes and the results show a statistically significant overrepresentation of basic amino acids, primarily arginine residues at specific positions. A symmetric appearance of positively charged blocks at a distance of 12–15 amino acids from the G-R-P core, together with an overrepresentation of arginine and lysine amino acids at positions 2, 4, 40 and 42, i.e., 17 and 19 amino acids away from the core motif, were identified as characteristic features of human eAT-hooks (**Fig. 2A**). Though the overrepresentation of basic amino acids at the aforementioned positions is less prominent, the pLogo visualization of mouse eAT-hooks provides a similar result (**Fig. S2A**).

As the canonical AT-hook is a short protein motif and only 22 amino acids sequences were analyzed in the original sequence logos,[4] we wondered whether N- and C-terminal extensions of AT-hook peptides show similar sequence patterns to eAT-hooks or differ from them. Notably, the lists of human and mouse AT-hooks slightly vary between different sequence motif repositories, such as InterPro (AT-hook ID: IPR017956), Pfam (AT-hook ID: PF02178) and SMART (AT-hook ID: SM00384). We
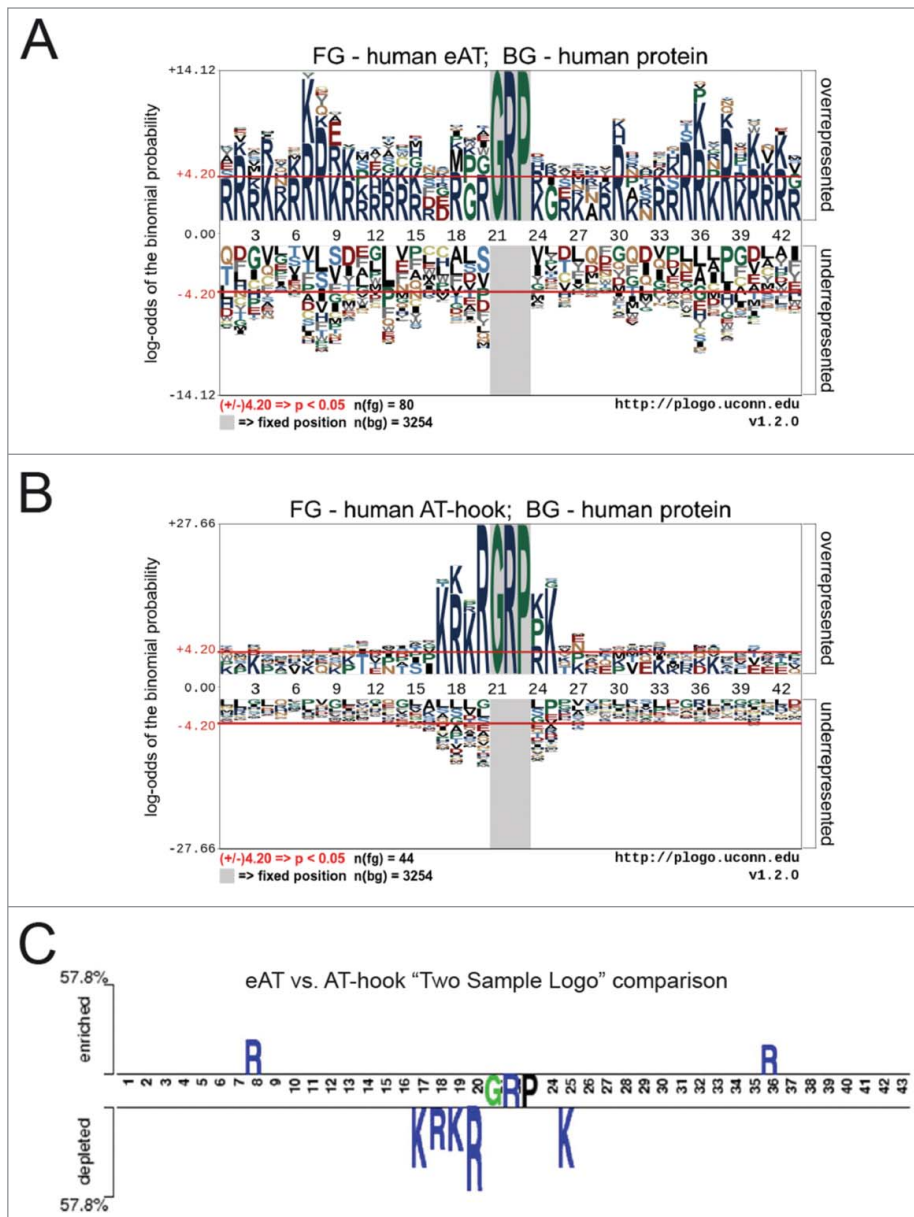
Figure 2. Sequence logo visualizations of eAT-hooks and AT-hooks. (**A**) pLogo visualization of human eAT-hook peptide sequences. Fixed G-R-P positions are labeled with gray shading. The title indicates that eAT-hooks were used as foreground (FG) and all G-R-P-containing peptides of the human prote-ome as background (BG) sequences. The n(fg) and n(bg) values at the bottom left of the pLogo indi-cate the number of aligned foreground and background sequences used to generate the image, respectively. The red horizontal bars on the pLogo correspond to P = 0.05. Note the symmetrical appearance of basic amino acids 12–15 residues away from the G-R-P core motif on both sides. (**B**) pLogo visualization of human AT-hook peptide sequences. Labels are as described in (**A**). (**C**) Two Sample Logo of the differences between eAT-hooks and AT-hooks. The combined groups of human and mouse eAT-hooks vs. AT-hooks were compared. Bonferroni correction (P < 0.05) was applied to eliminate spurious statistical significance. Note the highly significant overrepresentation of R resi-dues at positions 8 and 36, as well depletion of the KRKR motif at position 17–20 and the K residue at position 25 in eAT-hooks.

Additionally we removed duplicated entries from the AT-hook lists, and hypothetical proteins, which also appeared occasionally as redundant sequences. Finally, we generated a set of 44 human and 24 mouse lengthened AT-hook sequences for further sequence analyses (**Table S1**). Human and mouse AT-hook pLogo images were created, and the sequence visualization clearly shows the expected, characteristic pattern of this sequence motif (**Fig. 2B**; **Fig. S2B**). Importantly, the highly significant overrepresentation of basic amino acids in the immediate vicinity of the core motif does not coincide with an overrep-resentation of arginine and lysine resi-dues at positions in larger distances from the central G-R-P tripeptide.

The pLogo results and the fact that the eAT-hook lists contain only a few canoni-cal AT-hook sequences already suggested that eAT- and AT-hooks are distinct pro-tein motifs. To confirm this assumption, differences between AT-hook and eAT-hook motifs were investigated using the Two Sample Logo sequence analysis tool.[16] Combined multiple sequence alignments of human and mouse eAT-hooks or AT-hooks were generated, and they were subsequently analyzed by apply-ing 2 different settings (**Fig. 2C**; **Fig. S2C**). We observed a highly significant overrepresentation of R residues at posi-tions 8 and 36, as well depletion of the KRKR motif at position 17–20 and the K residue at position 25 in eAT-hooks, when spurious statistical significance is eliminated by Bonferroni correction (P < 0.05, **Fig. 2C**). Taken together, these results clearly demonstrate that eAT-hooks are not simply elongated canonical AT-hooks, but distinct sequence motifs.

Besides sequence motif analyses, the group of eAT-hook-containing proteins was subjected to functional profiling using the FatiGO web-tool, which is included in the Babelomics suite.[17,18] In order to obtain statistically significant information, AT-hook containing sequences were removed from the eAT-hook list prior to the analysis. Although

selected the human and mouse AT-hook lists of the SMART repository because they contained the most of the sequence entries. The sizes of the AT-hooks were extended to 43 amino acids, which is the predetermined size of eAT-hooks.

nucleic acid metabolism-related gene ontology terms (e.g. DNA binding, RNA binding etc.) associate with approximately half of the proteins, no gene ontology term was found to characterize the eAT-hook proteins with statistical significance. This result

suggests yet unidentified functions for many of these proteins and implies that their molecular function ontology terms could be further complemented.

The eAT-hook domains of PTOV1 and GPBP1 display similar DNA binding properties to Tip5s eAT-hook

After classifying the new eAT-hook domain as a functional DNA binding motif in Tip5 and identifying a large number of eAT-hooks in the human and mouse proteomes, we set out to test the DNA binding activity of eAT-hook domains found in other proteins. Two candidate proteins, which are involved in nucleic acid metabolism but do not contain known nucleic acid binding domains, were selected. The PTOV1 (Prostate Tumor Overexpressed 1) protein was initially reported as a factor overexpressed in prostate carcinoma cells. In addition, it has been suggested to have some regulatory function in cell cycle progression and was observed to shuttle between the cytoplasm and the nucleus.[19,20] The human PTOV1 protein consists of 416 amino acids, of which the first 43 residues represent the eAT-hook. In addition, the protein possesses two Med25 (Mediator complex subunit 25, PTOV activation and synapsin 2; Pfam ID: PF11232) motifs, which are predicted transactivation domains. The other selected eAT-hook protein, GPBP1 (GC-rich Promoter Binding Protein 1, also known as Vasculin, vascular wall–linked protein) was described as GC-rich promoter binding protein, which is able to induce transcription on TATA-box deficient promoters.[21,22] The human GPBP1 is a 473 amino acids protein, which contains an eAT-hook at its N-terminus and a VL1 motif (Vascular protein family Vasculin-like 1; Pfam ID: PF15337) at the C-terminus (**Fig. 3A**). The VL1 motif is linked to transcription factor function, but its exact role has not been revealed so far.

To analyze the DNA binding activity of the eAT-hooks of PTOV1 and GPBP1, GST-tagged wild-type (eAT), point mutant (eATmut) and deletion mutant (eATcore) versions of the protein domain were cloned. In the eATmut constructs the R-G-R-P and G-R-P cores of PTOV1s and GPBP1s eAT-hook were changed to D-G-D-P and G-D-P, respectively. The eATcore deletion mutants were truncated at their N- and C- termini, removing the basic amino acid patches of the peptides (**Fig. 3B**). GST-tagged eAT-hook proteins were expressed, affinity purified and analyzed on Coomassie gels prior to functional analyses (**Fig. 3C**). MST experiments were performed using the "rDNA En" DNA fragment and the same experimental setup as described for Tip5s eAT-hook. The analysis of the MST data yielded 4.0 ± 0.7 μM and 7.0 ± 1.7 μM equilibrium binding constant concentration values for the eAT-hooks of PTOV1 and GPBP1, respectively. Similar to Tip5, none of the eAT-hook mutants did measurably bind to this DNA fragment (**Fig. 3D, E**).

In order to confirm the results of MST analyses, gel shift assays were performed using PTOV1s eAT-hook peptides (**Fig. 3F**). The different eAT-hook proteins were incubated with the AT-rich "rDNA En" sequence under identical experimental conditions and the DNA binding properties of the different GST-eAT proteins were monitored on native polyacrylamide gels. The results clearly show binding of the

target DNA by PTOV1s eAT-hook. This is illustrated by the appearance of high molecular weight protein-DNA complexes in lanes 5 and 6. Apparently, with an increase of the protein amount, all DNA was bound to the protein and the DNA was supershifted in lanes 7 and 8. As expected the eATmut and eATcore peptides displayed no significant DNA binding activity in these experiments. In summary PTOV1s and GPBP1s eAT-hooks showed very similar DNA binding behavior to Tip5s eAT-hook.

eAT-hooks bind to RNA with higher affinity than to DNA

Recent publications already suggested a role in RNA interactions for canonical AT-hooks of the HMGA1 and Tip5 proteins.[23,24] Importantly, it was shown that a stem-loop structure is necessary for the proper RNA binding of the first AT-hook domain of HMGA1.[23] These observations in combination with a recent report implicating an RNA component in PTOV1 function[25] made it tempting to speculate about a possible RNA binding activity for the eAT-hook domains. To test this hypothesis, MST analyses with the described GST eAT-hook peptides of Tip5, PTOV1 and GPBP1 were performed. For these experiments two RNA oligonucleotides, which were already used in in vitro protein–RNA interaction studies in our lab, were selected[26] (**Fig. 4A**). As predicted by the RNAfold web tool,[27] oligonucleotide "RNA1" harbours a large stem-loop structure with a long double-stranded region, while "RNA2" comprises only a small hairpin structure and a large loop region. Thus, they provided an interesting structural diversity for the intended assay. For the MST measurements, both RNAs were incubated with dilution series of the GST-tagged Tip5, PTOV1 and GPBP1 eAT-hooks under identical experimental conditions. The analysis was conducted as described previously,[13] and all data points were normalized to the bound fraction of the RNA molecules and plotted against the protein concentration. The results revealed a strong RNA binding activity for all 3 eAT-hook domains tested. Interestingly, all proteins display a higher affinity for the "RNA1" molecule, with EC50 values of 187 ± 10 nM for Tip5s (**Fig. 4B**), 203 ± 6 nM for PTOV1s (**Fig. 4C**) and 378 ± 18 nM for GPBP1s eAT-hook (**Fig. 4D**). All three proteins also bind to the "RNA2" molecule, however a 2- to 3-fold decrease in binding affinity was detected. The EC50 values increased to 569 ± 40 nM for Tip5s, 600 ± 30 nM for PTOV1s and 938 ± 92 nM for GPBP1s eAT-hook. In summary the results of the MST analyses show that all tested eAT-hook domains bind to RNA with approximately one order of magnitude better than to DNA, and the secondary structure of RNA influences the binding affinity.

eAT-hooks bind preferentially to RNA in competitive nucleic acid binding assays

To further dissect the function of the eAT-hook domains in nucleic acid binding, competitive MST analyses were performed and the preference for RNA versus DNA was monitored. For this purpose, a FAM-labeled "RNA1" oligonucleotide and a Cy5-labeled "rDNA En" DNA oligonucleotide were mixed in equimolar ratio and MST measurements were carried out and

analyzed as described above. The results strongly support the observed preference for RNA binding of the 3 different eAT-hook domains. Tip5s eAT-hook binds under the competitive conditions with an EC50 of 277 ± 8 nM to the RNA (**Fig. 4E**), whereas PTOV1s eAT-hook displays an EC50 of 252 ± 10 nM (**Fig. 4F**) and GPBP1s eAT-hook reveals an EC50 of 373 ± 19 nM (**Fig. 4G**), which are comparable to the binding affinities measured in the absence of competitor DNA. In all 3 cases no interaction with the DNA was detected, the proteins bound exclusively to RNA.
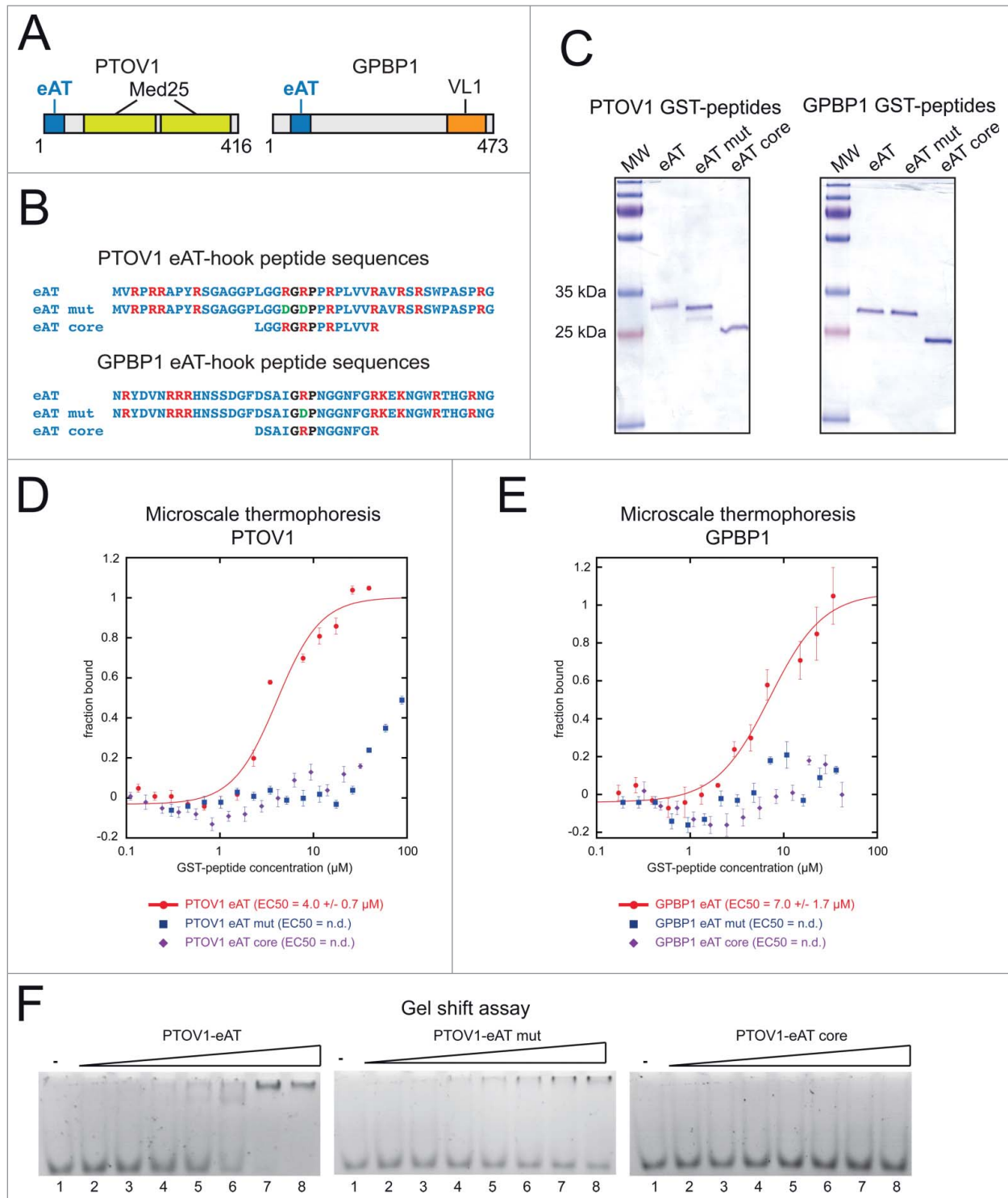
Figure 3. For figure legend, see page 870.

Deletion of the eAT-hook domain impairs nucleocytoplasmic shuttling of PTOV1

Previous studies revealed that PTOV1 shuttles between the cytoplasm and the nucleus, further it associates with ribonucleoprotein particles which also shuttle.[20,25] Because eAT-hooks displayed a strong RNA binding activity in the MST experiments, we decided to test the influence of the eAT-hook on PTOV1 shuttling *in vivo*. For this 2 mammalian expression vector constructs were cloned, encoding a GFP-tagged, full-length PTOV1 fusion protein (PTOV1 WT), as well an N-terminal deletion mutant lacking the eAT-hook comprising first 48 amino acids (PTOV1 ΔeAT). Since the eAT-hook presents the only characterized nucleic acid interaction domain of PTOV1, this deletion was likely to interfere with its functions in RNA- or DNA-related processes. The two constructs were transfected in HeLa or HCT116 cells and their expression was monitored on western blots, furthermore sub-cellular localization of the expressed GFP fusion proteins was examined by fluorescence microscopy (**Fig. 5**; **Fig. S3**). Strikingly, PTOV1 WT and PTOV1 ΔeAT revealed a clear difference in their localization patterns. PTOV1 WT expressing cells show 3 distinct patterns of GFP signal distribution with cells displaying mainly nuclear localization (1), cells revealing evenly dispersed signals in the nucleus and the cytoplasm (2), and cells showing a higher amount of GFP signal in the cytoplasm (3) (**Fig. 5A**; **Fig. S3A**). These results were in good agreement with previously published reports, describing similar patterns, caused by the shuttling mechanism of PTOV1.[20,25,28] In contrast, PTOV1 ΔeAT shows one particular localization pattern with the GFP signals predominantly located in the nucleus and displaying only very faint cytoplasmic signals (**Fig. 5A**; **Fig. S3A**). To get a more detailed view on the differential distribution of the GFP fusion proteins throughout the cells, the transfected cells were examined with a higher magnification, too (**Fig. 5B,C**). In addition to the 3 different localization patterns of PTOV1 WT, an accumulation of the GFP signal can be observed around the nucleus in cells that contain higher amounts of the protein in the cytoplasm (**Fig. 5B**). The PTOV1 ΔeAT expressing cells exhibit the primary nuclear localization as seen before (**Fig. 5C**). The possibility that this effect is simply due to the loss of a nuclear export signal present in the deleted region is unlikely, because a search for this motif using a web based *in silico* prediction tool[29] delivered a negative result. These observations clearly suggest a role for the eAT-hook in nucleocytoplasmic shuttling and indicate its importance in PTOV1 function.

Discussion

Characterization of the eAT-hook domain

When investigating the function of Tip5s AT-hooks, we have identified an AT-hook-like domain, which was termed extended AT-hook due to the relatively large distances of basic amino acid patches from the core G-R-P motif. Microscale thermophoresis and gel shift assays revealed a similar DNA binding activity of the eAT-hook to the canonical AT-hooks of Tip5. The characterization of the novel Tip5 eAT-hook domain as a functional non-canonical AT-hook set the stage for the identification of a large group of proteins holding this novel putative DNA-binding domain. The widespread distribution of the $(K/R)_{>3}$-X_{6-20}-GRP-X_{6-20}-$(K/R)_{>3}$ sequence throughout the human and mouse proteomes with 80 and 60 different candidate proteins, resembling about twice the number of already mapped AT-hooks, further underlines the relevance of this new motif. Bioinformatics analyses of sequence alignments using the pLogo and Two Sample Logo web tools demonstrate that eAT-hook and AT-hook sequence motifs are clearly distinct from each other. Notably, the central sequence region of some eAT-hooks can also be identified as canonical AT-hook (see **Table S1**), suggesting that these domains may exhibit the functions of both eAT- and AT-hooks.

The results gained from gel shift and MST analysis of the eAT-hooks of the PTOV1 and GPBP1 proteins emphasize the observations made for the Tip5 eAT-hook motif. Remarkably, all 3 domains display DNA binding affinities in a range comparable to canonical AT-hooks. Moreover, deletion and point mutants of these eAT-hook peptides lose DNA binding activity, further validating the nucleic acid binding capacity of this protein domain. Surprisingly, additional investigations hinted at a strong RNA-binding activity for eAT-hooks. Therefore the data presented here indicates that the eAT-hook is a new functional nucleic acid binding domain, which likely mediates RNA-dependent functions in many proteins. We would like to note that several bioinformatically-identified eAT-hooks may still represent
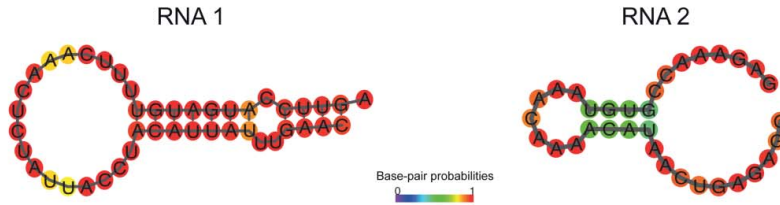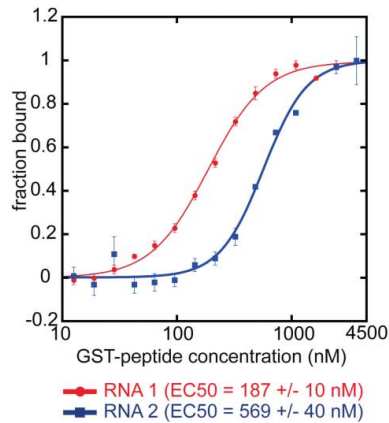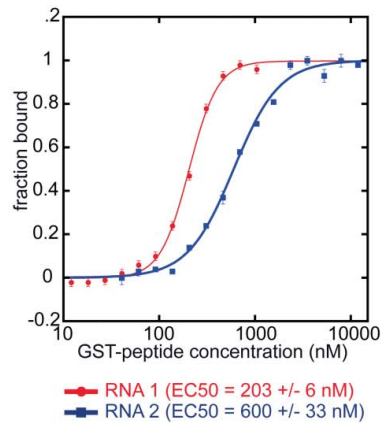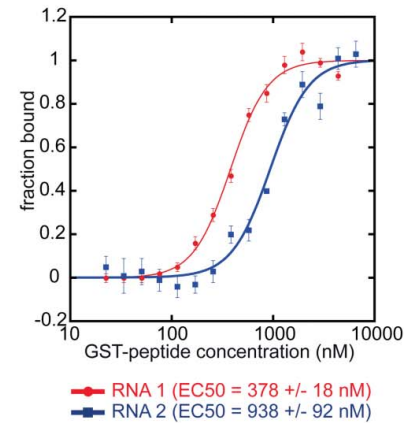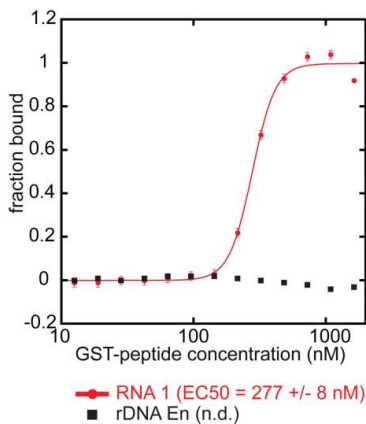
Figure 3 (**See previous page**). DNA binding features of PTOV1s and GPBP1s eAT-hooks. (**A**) Domain structures of PTOV1 and GPBP1. The extended AT-hooks are labeled with eAT in blue; Med25 (Mediator complex subunit 25, PTOV activation and synapsin 2; Pfam ID: PF11232) domains of PTOV1 are labeled in yellow; VL1 (Vascular protein family Vasculin-like 1; Pfam ID: PF15337) motif of GPBP1 is shown in orange. The numbers indicate amino acids. (**B**) Amino acid sequences of PTOV1 and GPBP1 eAT-hook peptides are shown below the schemes. Red letters mark positively charged amino acids, green letters indicate R-to-D mutations, and core motif G and P amino acids are labeled in black. (**C**) Coomassie gel picture of GST-tagged, purified PTOV1 and GPBP1 eAT-hook peptides. (**D**) MST experiments with PTOV1 eAT-hook peptides. The indicated amounts of the different PTOV1 eAT-hook peptides were analyzed for DNA binding at constant, 50 nM concentration of a 29 bp, Cy5-labeled "rDNA En" oligonucleotide. The fraction of bound DNA was plotted against the peptide concentration, and the concentration where half of the oligonucleotide is bound, the EC50 value, was determined. The binding data of PTOV1-eATmut and PTOV1-eATcore peptides were obtained in parallel with PTOV1-eAT and normalized to the difference between the minimum and maximum values of the PTOV1-eAT binding experiment. The data points of the plot represent mean±standard deviation values of 4 measurements. (**E**) MST experiments with GPBP1 eAT-hook peptides. The experiments with the GPBP1 eAT-hook peptides were carried out as described in (**D**). (**F**) Gel shift assay analyses of PTOV1-eAT-hook DNA binding. PTOV1-eAT, PTOV1-eATmut and PTOV1-eATcore peptides were tested for DNA binding as indicated. Increasing amounts of the peptides (6, 9, 13, 20, 30, 45, 68 pmol) were added to 1.25 pmol of the "rDNA En" probe. Lane 1 shows the probe alone.

A

RNA oligos used for MST

RNA 1 5'FAM – AGUUCCAUGAUGUUUUCAAACUCUAUUACCUACAUUAUUUGAAC

RNA 2 5'CY5 – GAGAAACCGUGUAAACAAAACAUAACUGAGAGG

RNAfold structure prediction

RNA 1   RNA 2

Base-pair probabilities
0   1

B

Microscale thermophoresis
Tip5 - RNA

— RNA 1 (EC50 = 187 +/- 10 nM)
— RNA 2 (EC50 = 569 +/- 40 nM)

C

Microscale thermophoresis
PTOV1 - RNA

— RNA 1 (EC50 = 203 +/- 6 nM)
— RNA 2 (EC50 = 600 +/- 33 nM)

D

Microscale thermophoresis
GPBP1 - RNA

— RNA 1 (EC50 = 378 +/- 18 nM)
— RNA 2 (EC50 = 938 +/- 92 nM)

E

Microscale thermophoresis
Tip5 - RNA vs DNA

— RNA 1 (EC50 = 277 +/- 8 nM)
■ rDNA En (n.d.)

F

Microscale thermophoresis
PTOV1 - RNA vs DNA

— RNA 1 (EC50 = 252 +/- 10 nM)
■ rDNA En (n.d.)
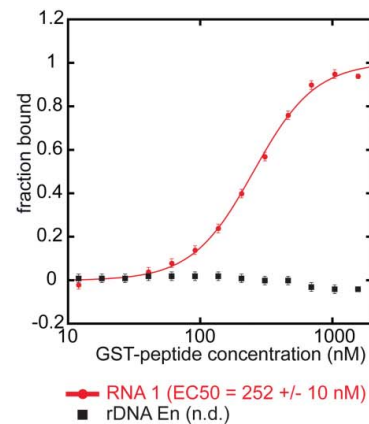
G

Microscale thermophoresis
GPBP1 - RNA vs DNA

— RNA 1 (EC50 = 373 +/- 19 nM)
■ rDNA En (n.d.)
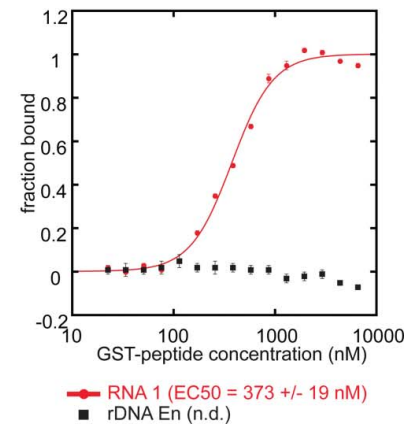
Figure 4. For figure legend, see page 872.

peptides that do not possess nucleic acid binding activity. In contrast, eAT-hooks with variations in the G-R-P core motif may also exist, and such sequences would represent false negatives of our eAT-hook search.

The importance of the eAT-hook for proper protein function *in vivo* is shown for a selected eAT-hook protein called PTOV1, which contains an N-terminal eAT-hook and no other known DNA- or RNA-binding domain. PTOV1 was initially described as a factor overexpressed in prostate carcinoma cells.[19] The exact function of the protein is so far not known, however a role for PTOV1 in regulation of cell cycle progression was implicated and the shuttling of the protein between the cytoplasm and nucleus has been revealed.[20] Interestingly, a very recent report described the association of PTOV1 with 40 S ribosomes through interaction with RACK1 (receptor of protein C kinase 1), which in turn could control translation initiation of c-Jun.[25] The authors furthermore suggested a function for PTOV1 in ribonucleoprotein shuttling. We showed that PTOV1 accumulates in the nucleus upon deletion of the eAT-hook, which demonstrates the importance of this domain in nucleocytoplasmic shuttling of the protein, and implies that it may have a regulatory role in the aforementioned biological process.

The relationship of eAT-hooks to other short K/R-rich protein sequence motifs

The discovery of various non-canonical AT-hooks already emphasized that it is not possible to present a single sequence motif, which allows the identification of all short peptides with AT-hook-like function. The analysis of the β′ subunits of DNA-dependent RNA polymerases suggested that such sequence motifs appeared very early in evolution possibly playing a role in nucleic acid – protein interactions.[5] Since AT-hook-like motifs show high sequence variability, we have not intended to search for all possible eAT-hook-like sequences, but rather focused on the characterization of several eAT-hooks and generated a sequence logo for human and mouse $(K/R)_{>3}$-X_{6-20}-GRP-X_{6-20}-$(K/R)_{>3}$-type eAT-hooks. This sequence logo was significantly different from the sequence logo of 43-amino-acids-sized AT-hooks.

The K/R richness of the eAT-hook motif also raised the question about its possible similarity to short K/R-rich protein

sequence motifs, i.e. nuclear localization signals (NLSs) and RGG/RG motifs. Although NLSs frequently overlap nucleic acid binding domains,[30] and eAT-hooks are rich in basic amino acids, they do not represent NLSs. A characteristic feature of NLSs is lysine clustering, but the eAT-hook sequence logo illustrates an arginine-rich sequence motif. The arginine-richness, on the other hand, clearly supports the occasional occurrence of RGG/RG motifs within eAT-hooks. In the 80 human eAT-hooks altogether 1 di-RGG, 2 tri-RG, and 8 di-RG motifs can be identified. Remarkably, the role of RGG/RG motifs in mediating nucleic acid - protein interactions was highlighted recently.[31]

Possible roles for the RNA-binding activity of eAT-hook proteins

Surprisingly, the tested eAT-hook domains of Tip5, PTOV1 and GPBP1 displayed a strong RNA binding activity in MST experiments. The affinities of the 3 different domains measured for RNA are in the hundred nanomolar range and approximately 10- to 20-fold higher than their affinity for DNA molecules. The results of nucleic acid binding analyses also suggest a structural preference of RNA binding. All three tested eAT-hooks bind to the RNA oligonucleotide comprising a helix-loop structure with an almost 3-fold higher affinity than to the RNA featuring a small stem – large loop structure. Noteworthy, this behavior parallels the efficient binding of the first AT-hook of HMGA1 to the double-stranded helix-loop structure of 7SK snRNA.[23] Moreover, in competitive MST assays the presence of DNA does not interfere with the RNA interactions, further validating the preferred RNA binding of the eAT-hook proteins.

In recent years the discovery of an ever-growing number of functional ncRNAs, working as guides, regulators, and scaffolds for the assembly and targeting of chromatin modifying complexes, revealed a completely new layer of gene regulatory processes.[32] Now that numerous novel non-coding roles have been discovered for various RNA classes, it is becoming increasingly important to map RNA-dependent regulatory networks and to further identify and characterize the underlying individual RNA-protein interactions in detail. Strikingly, 2 studies involving HMGA1 and Tip5 already illustrated the capability of AT-hook domains to mediate RNA interactions in regulatory complex formation.[23,24] For HMGA1 this

Figure 4 (**See previous page**). Nucleic acid binding features of Tip5s, PTOV1s and GPBP1s eAT-hooks. (**A**) RNA oligonucleotides used in MST analysis of eAT-hook RNA binding. The sequences of the RNAs used in the MST assay are depicted. "RNA1" was labeled with a FAM moiety on the 5′ end (red sequence). The "RNA2" oligonucelotide contained a Cy5 label on the 5′ end (blue sequence). The RNAfold web tool was used for structure prediction of the different RNAs. The colors of the different nucleotides represent the calculated basepair probabilities. (**B**) MST analysis of Tip5 eAT-hook binding to RNA. 50 nM of Cy5 labeled "RNA1" (red data points) or FAM labeled "RNA2" (blue data points) oligonucleotides were incubated with serial dilutions of the Tip5 eAT-hook peptide and MST measurements were performed. The thermophoresis signals were normalized to the fraction bound and plotted to the protein concentration, depicted on a logarithmic scale in nM concentration. The data points of the plot represent mean±standard deviation values of 4 measurements. (**C**) PTOV1 eAT-hook RNA binding. The MST experiment was performed as described in (**B**). (**D**) GPBP1 eAT-hook RNA binding. The MST experiment was performed as described in (**B**). (**E**) Competitive MST analysis of Tip5s eAT-hook nucleic acid binding. 50 nM of FAM-labeled "RNA1" (red data points) together with 50 nM Cy5-labeled "rDNA En" (black data points) oligonucleotides were incubated with serial dilutions of the Tip5 eAT-hook peptide and MST measurements were performed. The thermophoresis signals were normalized to the fraction bound and plotted to the protein concentration, depicted on a logarithmic scale in nM concentration. The data points of the plot represent mean±standard deviation values of 4 measurements. (**F**) Competitive MST analysis of PTOV1s eAT-hook nucleic acid binding. The MST experiment was performed as described in (**E**). (**G**) Competitive MST analysis of GPBP1s eAT-hook nucleic acid binding. The MST experiment was performed as described in (**E**).
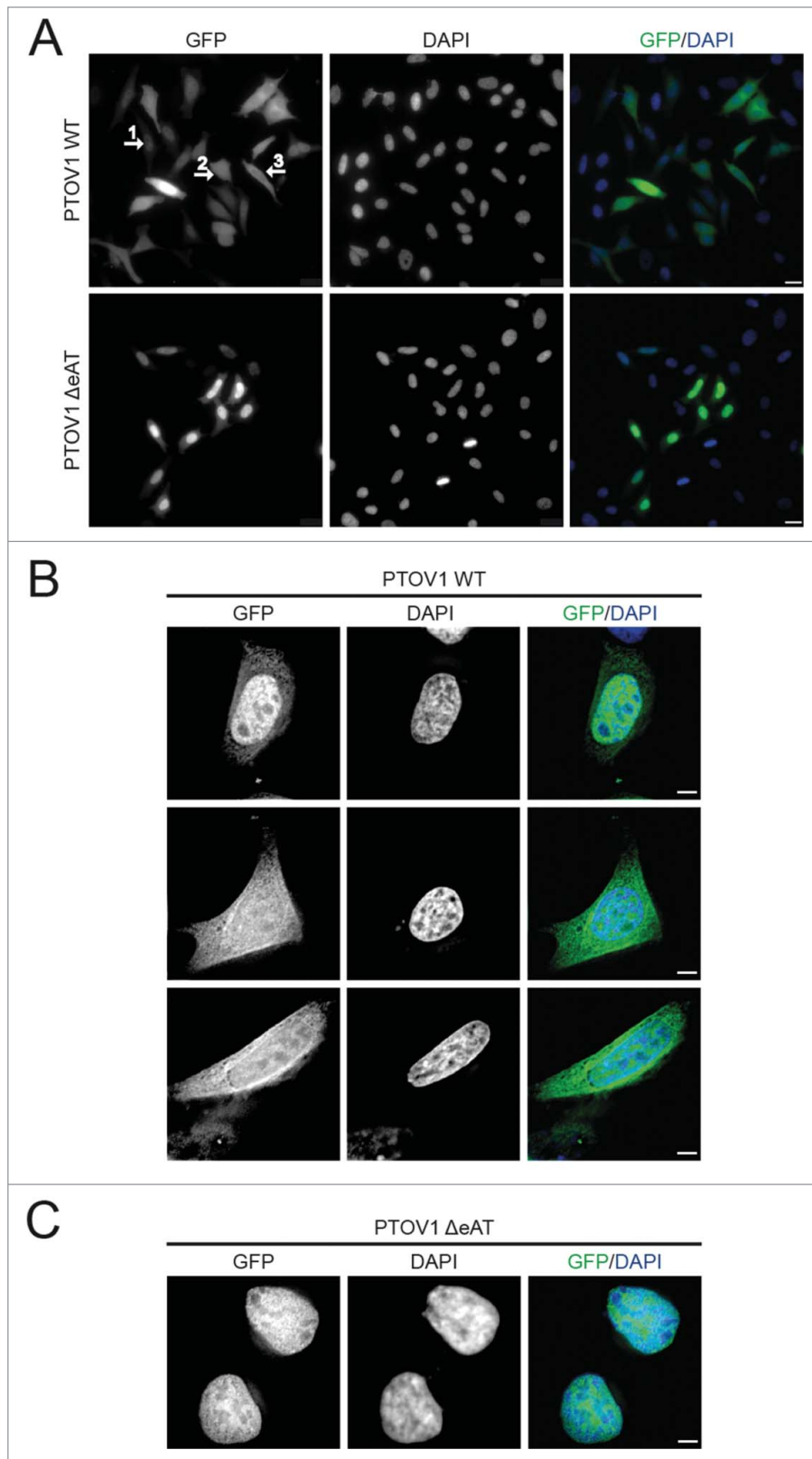
Figure 5. The eAT-hook domain influences the subcellular localization of PTOV1. (**A**) Deletion of the eAT-hook domain leads to a distinct subcellular localization of PTOV1. HeLa cells were transfected with a GFP-tagged wild type (WT) and an eAT deletion mutant of PTOV1 (ΔeAT) as indicated on the left. Cells were fixed 24 h post-transfection and immunostaining of GFP was performed. DNA was counterstained with DAPI. The numbered arrows in the PTOV1 WT picture mark cells with different nuclear-cytoplasmic distributions of the protein. Bars indicate 15 μm. (**B**) Three different localization patterns of the GFP-tagged PTOV1 protein. Individual cells were selected from the experiment described in (**A**). The protein shows predominantly nuclear localization in the upper panel, a nearly equal distribution between the nucleus and the cytoplasm in the middle panel, and cytoplasmic enrichment with perinuclear accumulation in the lower panel. Bars indicate 5 μm. (**C**) Nuclear localization of GFP-tagged PTOV1-ΔeAT. A representative image shows 2 cells from the experiment described in (**A**) with predominant nuclear localization of the protein. The bar indicates 5 μm.

deacetylation of a lysine residue within the second AT-hook leads to increased binding of the evolutionarily conserved non-coding pRNA, and to enhanced heterochromatin formation at the rDNA locus.[24,34] Additionally, ncRNA binding of the insulator protein CTCF was also described recently and the RNA-binding domain was mapped to the C-terminus of the protein.[35] This chromatin regulator protein does not possess known RNA-binding domains but it contains an AT-hook at its C-terminus.[36] In the light of these facts, the RNA binding of the eAT-hook could suggest a role for eAT-hook proteins in RNA-mediated processes within the cell. It is tempting to speculate that eAT-hooks and AT-hooks of some proteins may not only tether and anchor them to particular DNA structures and sequences as originally predicted for AT-hook proteins,[4] but may have a function as RNA-binding factors. Concerning eAT-hooks, the chromatin-dependent regulatory functions of

results in a ribonucleoprotein complex consisting of the ncRNA 7SK, HMGA1 and P-TEFb playing a role in the transcriptional regulation of target genes.[23,33] In Tip5 the

Brg1 (Brahma related gene 1), Dot1L (Dot1 [Disrupter of telomere silencing protein 1]-like), and Tip5 could involve as well eAT-hook-RNA interactions, since these proteins have been shown to associate with specific lncRNAs.[37,38] The identification of eAT-hooks in the Rps8, Mrp-S12, Rrp12 and Srsf10 proteins suggests a possible involvement of this domain also in the function of specific ribonucleoprotein complexes, for instance in ribosome and spliceosome function. In addition, 19 eAT-hook proteins can be found in current catalogs of human RNA-binding proteins,[39-41] which have been identified mostly by their ability to bind mRNA (**Table S1**). Although the aforementioned eAT-hook proteins can already be classified as factors that bind different RNA molecules, many other eAT-hook proteins are yet uncharacterized. The discovery of the presented novel sequence motif and the characterization of selected eAT-hooks set the ground to explore the function of eAT-hook proteins, and to better understand their role in RNA- and chromatin-dependent biological processes.

Materials and Methods

Plasmid DNA constructs

GST-eAT-hook constructs were prepared by inserting the coding sequences for the peptides listed in **Figures 1B and 3B** into BamHI/SalI digested pGEX4T3 plasmid DNA. The pGEX4T3-Tip5-AT1+2 plasmid encoding the GST-tagged GKRGRPRNNEKAKNKEVPKVKRGRGRPPKIKMPE peptide of Tip5 has been previously described.[13] Synthetic DNA oligonucleotides with flanking BamHI/SalI restriction sites were used for the cloning of eATcore constructs of Tip5, PTOV1 and GPBP1. The Tip5 eAT coding DNA fragment was PCR amplified using primers with flanking BamHI/SalI restriction sites. The DNA inserts for the cloning of eATmut and eAT constructs of PTOV1 and GPBP1 were obtained from synthetic plasmid DNA (GeneArt Gene Synthesis, Life Technologies). The pEGFP-PTOV1 plasmid DNA was cloned by inserting the full-length cDNA of PTOV1 into EcoRI/XbaI digested pEGFP-C2. The insert was PCR amplified from the clone 5928807 (IMAGE ID) of the NIH_MGC_47 library (Source Bioscience). To obtain pEGFP-PTOV1-ΔeAT, the pEGFP-PTOV1 construct was EcoRI/BamHI digested, the ends refilled with the Klenow enzyme and re-ligated. Novel plasmid DNA sequences were analyzed by sequencing.

Expression and purification of recombinant proteins

The GST fusion proteins were expressed in the *E. coli* Rosetta 2 strain. Transformed bacteria were induced at $OD_{600} = 0.6$ by adding 0.5 mM IPTG and incubated for 3 hours at 37°C. Cells were then harvested and affinity purified on Glutathione Sepharose 4B beads (Pharmacia) in batch according to the manufacturer's protocol. Protein concentrations of the elution fractions were measured by using a Qubit fluorometer (Invitrogen).

Mass spectrometry analysis

GST-eAT proteins were separated on SDS–polyacrylamide gels. After electrophoresis, gels were stained with Coomassie brilliant blue G-250 and the 2 bands appearing in the GST-Tip5-eAT elution fraction were excised. Gel bands were destained, digested with trypsin, and the digestion reactions were analyzed by MALDI-TOF mass spectrometry as previously described.[42] Spectra were manually interpreted.

Protein – nucleic acid binding assays

Gel shift assays and microscale thermophoresis (MST) measurements were performed as described in our previous studies.[13,26] To analyze the nucleic acid binding activity of recombinant eAT-hooks, the following oligonucleotides were used: double-stranded synthetic DNA oligonucleotides (sense strand sequences): "rDNA En" (gel shift): 5′-TGG ATC TTT TTT TTT TTT TTT CTT TTT TCC TCC A, "rDNA IGS:" 5′-AAG TCG TGC CTA AAA TAA ATA TTT TTC TGG CCA A, "IFNβ:" 5′-GGA GAA GTG AAA GTG GGA AAT TCC TCT GAA TAG A, "rDNA En" (MST): 5′-Cy5 TCT TTT TTT TTT TTT TTC TTT TTT CCT CC. RNA oligonucleotides: "RNA1:" 5′-FAM AGU UCC AUG AUG UUU UCA AAC UCU AUU ACC UAC AUU AUU UGA AC, "RNA2:" 5′-Cy5 GAG AAA CCG UGU AAA CAA AAC AUA ACU GAG AGG.

Bioinformatics analyses of the extended AT-hook (eAT-hook)

To identify eAT-hook proteins in the human and mouse proteomes (http://www.uniprot.org/taxonomy/complete-proteomes, April 2013 version), the data sets were downloaded as FASTA files and a text search was performed for $(K/R)_{>3}$-X_{6-20}-GRP-X_{6-20}-$(K/R)_{>3}$ sequence motifs. Redundant sequences were eliminated manually. Statistical analysis of gene ontology (GO) terms was performed using the FatiGO web-tool, which is included in the Babelomics suite (http://babelomics.bioinfo.cipf.es/functional.html).[17,18] Sequence logos of the eAT-hook and AT-hook specifying statistical significance of residue heights were visualized using the pLogo web tool (https://plogo.uconn.edu/).[15] The position of the G-R-P core tripeptide was fixed in the visualization processes. Position-specific differences in the residue compositions between AT-hook and eAT-hook motifs were visualized using the Two Sample Logo sequence analysis tool.[16] Statistically significant differences were calculated either with or without Bonferroni correction (P < 0.05), which removes some of the spurious significance as described in the Two Sample Logo User Manual (http://www.twosamplelogo.org/).

Cell culture

HeLa human cervix carcinoma cells and HCT116 colorectal carcinoma cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 4.5 g/L glucose, 4 mM GlutaMAX, 10% fetal calf serum, 100 U/ml penicillin and 100 μg/ml streptomycin and incubated at 37°C in 5% CO_2. The cells were transfected with the FuGENE HD (Roche) transfection reagent according to the manufacturer's instructions.

Antibodies

To detect proteins by protein gel blot or immunofluorescence, rabbit polyclonal α-actin (Sigma A-2066), rat monoclonal α-GFP (3H9, Chromotek) and the GFP-Booster-Atto488 (Chromotek) antibodies were used at appropriate dilutions.

Immunofluorescence

Transfected HeLa and HCT116 cells were grown on coverslips for 24 and 48 hours, respectively, and fixed in 4% paraformaldehyde in PBS for 10 min. During the last minute a few drops of 0.5% Triton X-100 in PBS were added to the cells. Afterwards cells were permeabilized in PBS with 0.5% Triton X-100 for 10 min. Cells were then washed with PBS-T (PBS with 0.1% Tween20) 3 times for 5 min. Subsequently they were incubated with the GFP-Booster-Atto488 (Chromotek) in 4% BSA in PBS-T for 1 h in a humidified chamber followed by 3 times 5 min washes in PBS-T. The second washing solution was supplemented with 50 ng/ml DAPI to counterstain the DNA. Alternatively, only DAPI staining was performed on fixed cells, and GFP fluorescence was directly visualized. Slides were mounted in Vectashield (Vector). Images were taken on a Zeiss Axiovert 200 inverted fluorescence microscope using the AxioVision software and processed with ImageJ and Adobe Photoshop.

Supplemental Material

Supplemental data for this article can be accessed on the publisher's website.

References

1. Reeves R, Nissen MS. The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. J Biol Chem 1990; 265:8573-82; PMID:1692833

2. Huth JR, Bewley CA, Nissen MS, Evans JN, Reeves R, Gronenborn AM, Clore GM. The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. Nat Struct Biol 1997; 4:657-65; PMID:9253416; http://dx.doi.org/10.1038/nsb0897-657

3. Reeves R. Nuclear functions of the HMG proteins. Biochim Biophys Acta 2010; 1799:3-14; PMID:19748605

4. Aravind L, Landsman D. AT-hook motifs identified in a wide variety of DNA-binding proteins. Nucleic Acids Res 1998; 26:4413-21; PMID:9742243; http://dx.doi.org/10.1093/nar/26.19.4413

5. Iyer LM, Koonin EV, Aravind L. Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. BMC Struct Biol 2003; 3:1; PMID:12553882; http://dx.doi.org/10.1186/1472-6807-3-1

6. Metcalf CE, Wassarman DA. DNA binding properties of TAF1 isoforms with two AT-hooks. J Biol Chem 2006; 281:30015-23; PMID:16893881; http://dx.doi.org/10.1074/jbc.M606289200

7. Siddiqa A, Sims-Mourtada JC, Guzman-Rojas L, Rangel R, Guret C, Madrid-Marina V, Sun Y, Martinez-Valdez H. Regulation of CD40 and CD40 ligand by the AT-hook transcription factor AKNA. Nature 2001; 410:383-7; PMID:11268217; http://dx.doi.org/10.1038/35066602

8. Cikala M, Alexandrova O, David CN, Pröschel M, Stiening B, Cramer P, Böttger A. The phosphatidylserine receptor from Hydra is a nuclear protein with potential Fe(II) dependent oxygenase activity. BMC Cell Biol 2004; 5:26; PMID:15193161; http://dx.doi.org/10.1186/1471-2121-5-26

9. Kim S-Y, Kim Y-C, Seong ES, Lee Y-H, Park JM, Choi D. The chili pepper CaATL1: an AT-hook motif-containing transcription factor implicated in defence responses against pathogens. Mol Plant Pathol 2007; 8:761-71; PMID:20507536; http://dx.doi.org/10.1111/j.1364-3703.2007.00427.x

10. Kubota T, Maezawa S, Koiwai K, Hayano T, Koiwai O. Identification of functional domains in TdIF1 and its inhibitory mechanism for TdT activity. Genes Cells 2007; 12:941-59; PMID:17663723; http://dx.doi.org/10.1111/j.1365-2443.2007.01105.x

11. Turlure F, Maertens G, Rahman S, Cherepanov P, Engelman A. A tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of LEDGF/p75 with chromatin in vivo. Nucleic Acids Res 2006; 34:1653-65; PMID:16549878; http://dx.doi.org/10.1093/nar/gkl052

12. Baker SA, Chen L, Wilkins AD, Yu P, Lichtarge O, Zoghbi HY. An AT-Hook Domain in MeCP2 Determines the Clinical Course of Rett Syndrome and Related Disorders. Cell 2013; 152:984-96; PMID:23452848; http://dx.doi.org/10.1016/j.cell.2013.01.038

13. Zillner K, Filarsky M, Rachow K, Weinberger M, Längst G, Németh A. Large-scale organization of ribosomal DNA chromatin is regulated by Tip5. Nucleic Acids Res 2013; 41:5251-62; PMID:23580549; http://dx.doi.org/10.1093/nar/gkt218

14. Bourachot B, Yaniv M, Muchardt C. The activity of mammalian brm/SNF2alpha is dependent on a high-mobility-group protein I/Y-like DNA binding domain. Mol Cell Biol 1999; 19:3931-9; PMID:10330133

15. O'Shea JP, Chou MF, Quader SA, Ryan JK, Church GM, Schwartz D. pLogo: a probabilistic approach to visualizing sequence motifs. Nat Methods 2013; 10:1211-2; PMID:24097270; http://dx.doi.org/10.1038/nmeth.2646

16. Vacic V, Iakoucheva LM, Radivojac P. Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments. Bioinformatics 2006; 22:1536-7; PMID:16632492; http://dx.doi.org/10.1093/bioinformatics/btl151

17. Al-Shahrour F, Minguez P, Tarraga J, Medina I, Alloza E, Montaner D, Dopazo J. FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. Nucleic Acids Res 2007; 35:W91-6; PMID:17478504; http://dx.doi.org/10.1093/nar/gkm260

18. Medina I, Carbonell J, Pulido L, Madeira SC, Goetz S, Conesa A, Tarraga J, Pascual-Montano A, Nogales-Cadenas R, Santoyo J, et al. Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. Nucleic Acids Res 2010; 38:W210-3; PMID:20478823; http://dx.doi.org/10.1093/nar/gkq388

19. Benedit P, Paciucci R, Thomson TM, Valeri M, Nadal M, Caceres C, de Torres I, Estivill X, Lozano JJ, Morote J. PTOV1, a novel protein overexpressed in prostate cancer containing a new class of protein homology blocks. Oncogene 2001; 20:1455; PMID:11313889; http://dx.doi.org/10.1038/sj.onc.1204233

20. Santamaría A, Fernández PL, Farré X, Benedit P, Reventós J, Morote J, Paciucci R, Thomson TM. PTOV-1, a novel protein overexpressed in prostate cancer, shuttles between the cytoplasm and the nucleus and promotes entry into the S phase of the cell division cycle. Am J Pathol 2003; 162:897-905; http://dx.doi.org/10.1016/S0002-9440(10)63885-0

21. Hsu LC, Liu S, Abedinpour F, Beech RD, Lahti JM, Kidd VJ, Greenspan JA, Yeung CY. The Murine G+C-Rich Promoter Binding Protein mGPBP Is Required for Promoter-Specific Transcription. Mol Cell Biol 2003; 23:8773-85; PMID:14612417; http://dx.doi.org/10.1128/MCB.23.23.8773-8785.2003

22. Bijnens APJJ. Vasculin, a novel vascular protein differentially expressed in human atherogenesis. Blood 2003; 102:2803-10; PMID:12842993; http://dx.doi.org/10.1182/blood-2003-01-0306

23. Eilebrecht S, Brysbaert G, Wegert T, Urlaub H, Benecke BJ, Benecke A. 7SK small nuclear RNA directly affects HMGA1 function in transcription regulation. Nucleic Acids Res 2011; 39:2057-72; PMID:21087998; http://dx.doi.org/10.1093/nar/gkq1153

24. Zhou Y, Schmitz K-M, Mayer C, Yuan X, Akhtar A, Grummt I. Reversible acetylation of the chromatin remodelling complex NoRC is required for non-coding

RNA-dependent silencing. Nat Cell Biol 2009; 11:1010-6; PMID:19578370; http://dx.doi.org/10.1038/ncb1914

25. Marqués N, Sesé M, Cánovas V, Valente F, Bermudo R, de Torres I, Fernández Y, Abasolo I, Fernández PL, Contreras H, et al. Regulation of protein translation and c-Jun expression by prostate tumor overexpressed 1. Oncogene 2014; 33:1124-34; http://dx.doi.org/10.1038/onc.2013.51

26. Schubert T, Pusch MC, Diermeier S, Benes V, Kremmer E, Imhof A, Längst G. Df31 protein and snoRNAs maintain accessible higher-order structures of chromatin. Mol Cell 2012; 48:434-44; PMID:23022379; http://dx.doi.org/10.1016/j.molcel.2012.08.021

27. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. The Vienna RNA Websuite. Nucleic Acids Res 2008; 36:W70-4; PMID:18424795; http://dx.doi.org/10.1093/nar/gkn188

28. Santamaria A, Castellanos E, Gomez V, Benedit P, Renau-Piqueras J, Morote J, Reventos J, Thomson TM, Paciucci R. PTOV1 Enables the Nuclear Translocation and Mitogenic Activity of Flotillin-1, a Major Protein of Lipid Rafts. Mol Cell Biol 2005; 25:1900-11; PMID:15713644; http://dx.doi.org/10.1128/MCB.25.5.1900-1911.2005

29. la Cour T, Gupta R, Rapacki K, Skriver K, Poulsen FM, Brunak S. NESbase version 1.0: a database of nuclear export signals. Nucleic Acids Res 2003; 31:393-6; PMID:12520031; http://dx.doi.org/10.1093/nar/gkg101

30. LaCasse EC, Lefebvre YA. Nuclear localization signals overlap DNA- or RNA-binding domains in nucleic acid-binding proteins. Nucleic Acids Res 1995; 23:1647-56; PMID:7540284; http://dx.doi.org/10.1093/nar/23.10.1647

31. Thandapani P, O'Connor TR, Bailey TL, Richard S. Defining the RGG/RG Motif. Mol Cell 2013; 50:613-23; PMID:23746349; http://dx.doi.org/10.1016/j.molcel.2013.05.021

32. Morris KV, Mattick JS. The rise of regulatory RNA. Nat Rev Genet 2014; 15:423-37; PMID:24776770; http://dx.doi.org/10.1038/nrg3722

33. Eilebrecht S, Benecke B-J, Benecke A. 7SK snRNA-mediated, gene-specific cooperativity of HMGA1 and P-TEFb. RNA Biol 2011; 8:1084-93; ; PMID:21957495; http://dx.doi.org/10.4161/rna.8.6.17015

34. Wehner S, Dörrich AK, Ciba P, Wilde A, Marz M. pRNA: NoRC-associated RNA of rRNA operons. RNA Biol 2014; 11:3-9; PMID:24440945; http://dx.doi.org/10.4161/rna.27448

35. Kung JT, Kesner B, An JY, Ahn JY, Cifuentes-Rojas C, Colognori D, Jeon Y, Szanto A, Del Rosario BC, Pinter SF, et al. Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. Mol Cell 2015; 57:361-75; PMID:25578877; http://dx.doi.org/10.1016/j.molcel.2014.12.006

36. Ohlsson R, Renkawitz R, Lobanenkov V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet 2001; 17:520-7; PMID:11525835; http://dx.doi.org/10.1016/S0168-9525(01)02366-6

37. Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W, Merkurjev D, Ohgi KA, Da Meng, Zhang J, et al. lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. Nature 2014; 500:598-602; http://dx.doi.org/10.1038/nature12451

38. Han P, Li W, Lin C-H, Yang J, Shang C, Nurnberg ST, Jin KK, Xu W, Lin C-Y, Lin C-J, et al. A long non-coding RNA protects the heart from pathological hypertrophy. Nature 2014; 514:102-6; PMID:25119045; http://dx.doi.org/10.1038/nature13596

39. Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, et al. The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. Mol Cell 2012; 46:674-90; PMID:22681889; http://dx.doi.org/10.1016/j.molcel.2012.05.021

40. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, et al. Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. Cell 2012; 149:1393-406; PMID:22658674; http://dx.doi.org/10.1016/j.cell.2012.04.031

41. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. Nat Rev Genet 2014; 15:829-45; PMID:25365966; http://dx.doi.org/10.1038/nrg3813

42. Villar-Garea A, Imhof A. Fine Mapping of Posttranslational Modifications of the Linker Histone H1 from Drosophila melanogaster. PLoS ONE 2008; 3:e1553; PMID:18253500; http://dx.doi.org/10.1371/journal.pone.0001553