JBC RESEARCH ARTICLE

Arid5a uses disordered extensions of its core ARID domain for distinct DNA- and RNA-recognition and gene regulation

Received for publication, April 10, 2024, and in revised form, May 23, 2024 Published, Papers in Press, June 10, 2024,
https://doi.org/10.1016/j.jbc.2024.107457

Julian von Ehr[1,2], Lasse Oberstrass[3], Ege Yazgan[4,5], Lara Ina Schnaubelt[1], Nicole Blümel[4], Francois McNicoll[4], Julia E. Weigand[3], Kathi Zarnack[4,5], Michaela Müller-McNicoll[4,6], Sophie Marianne Korn[1,7,*], and Andreas Schlundt[1,8,*]

From the [1]Institute for Molecular Biosciences and Biomolecular Resonance Center (BMRZ), Goethe University Frankfurt, Frankfurt, Germany; [2]IMPRS on Cellular Biophysics, Frankfurt, Germany; [3]University of Marburg, Department of Pharmacy, Institute of Pharmaceutical Chemistry, Marburg, Germany; [4]Institute for Molecular Biosciences, and [5]Buchmann Institute for Molecular Life Sciences, Goethe University Frankfurt, Frankfurt, Germany; [6]Max-Planck Institute for Biophysics, Frankfurt, Germany; [7]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York, USA; [8]University of Greifswald, Institute of Biochemistry, Greifswald, Germany

Reviewed by members of the JBC Editorial Board. Edited by Ronald Wek

AT-rich interacting domain (ARID)-containing proteins, Arids, are a heterogeneous DNA-binding protein family involved in transcription regulation and chromatin processing. For the member Arid5a, no exact DNA-binding preference has been experimentally defined so far. Additionally, the protein binds to mRNA motifs for transcript stabilization, supposedly through the DNA-binding ARID domain. To date, however, no unbiased RNA motif definition and clear dissection of nucleic acid–binding through the ARID domain have been undertaken. Using NMR-centered biochemistry, we here define the Arid5a DNA preference. Further, high-throughput *in vitro* binding reveals a consensus RNA-binding motif engaged by the core ARID domain. Finally, transcriptome-wide binding (iCLIP2) reveals that Arid5a has a weak preference for (A)U-rich regions in pre-mRNA transcripts of factors related to RNA processing. We find that the intrinsically disordered regions flanking the ARID domain modulate the specificity and affinity of DNA binding, while they appear crucial for RNA interactions. Ultimately, our data suggest that Arid5a uses its extended ARID domain for bifunctional gene regulation and that the involvement of IDR extensions is a more general feature of Arids in interacting with different nucleic acids at the chromatin–mRNA interface.

Among the large number of DNA-binding proteins (DBPs), ARIDs compose a distinct family of nuclear proteins with manifold functions in cellular processes alongside transcriptional regulation (reviewed in (1, 2)). ARID proteins are classified with respect to their shared DNA-binding domain, named AT-rich interactive domain (ARID), reflecting the supposed preference for AT-rich DNA (3, 4). Beyond that, ARID-containing proteins—further referred to as 'Arids' for the sake of clear distinction from the ARID domain—are diverse in size and domain architecture, based on which the 15 known human Arids are divided into seven subfamilies (5). All ARID domains share a conserved fold, comprising a minimal core structure of six α-helices (H1 to H6, Fig. 1), with H3/4 and H5 forming a central helix-turn-helix (HTH) motif, a widespread DNA-binding unit of DNA-binding domains (5–7). Turn-containing motifs, similar to the HTH, are in principle also capable of recognizing dsRNA (8). It is thus not surprising that the general ability of nucleic acid–binding proteins to interact with both DNA and RNA (DRBPs) is conceived more widespread than previously thought (9). Still, most DRBPs are assumed to exploit distinct domains to interact with DNA and RNA, respectively, as for example, known for Sox2 (10) and SAFB proteins (11). Yet, certain domains, such as the zinc finger motifs, were early found to interact with both types of nucleic acids, for example, described for the *Xenopus laevis* protein TFIIIA (12).

Arid5a is the only Arid representative described as capable of binding both RNA and DNA (13). The Arid5 family members 5a and 5b share the least conserved domain architecture among Arids. Their ARID domains, however, are 73% identical (5). The large divergence of Arid5a and 5b reflects distinct functions: Arid5b is categorized as a transcriptional coactivator with essential roles in adipogenesis and liver development, involving chromatin interaction (14). The existing high-resolution information for an Arid5b ARID–DNA complex has been obtained with the supposedly specific dsDNA consensus motif 5′-AATA[CT]-3′ (15, 16). However, the motif has merely been questioned in single-nucleotide exchanges and, more importantly, motif expansions have not been tested. At the same time, a possible capability of the Arid5b ARID domain to interact with (ds) RNA has not been investigated, as is true for all other Arids.

Arid5a, significantly smaller in size, has been classified mainly as a transcriptional repressor (17), for example, for nuclear hormone receptors (18). On the other hand, Arid5a is thought to function actively in the transcription of specific genes and support gene de-repression by histone acetylation together with Sox9 (19). In 2013, Arid5a was termed an RNA-

* For correspondence: Sophie Marianne Korn, smk2305@cumc.columbia.edu; Andreas Schlundt, schlundt@bio.uni-frankfurt.de.

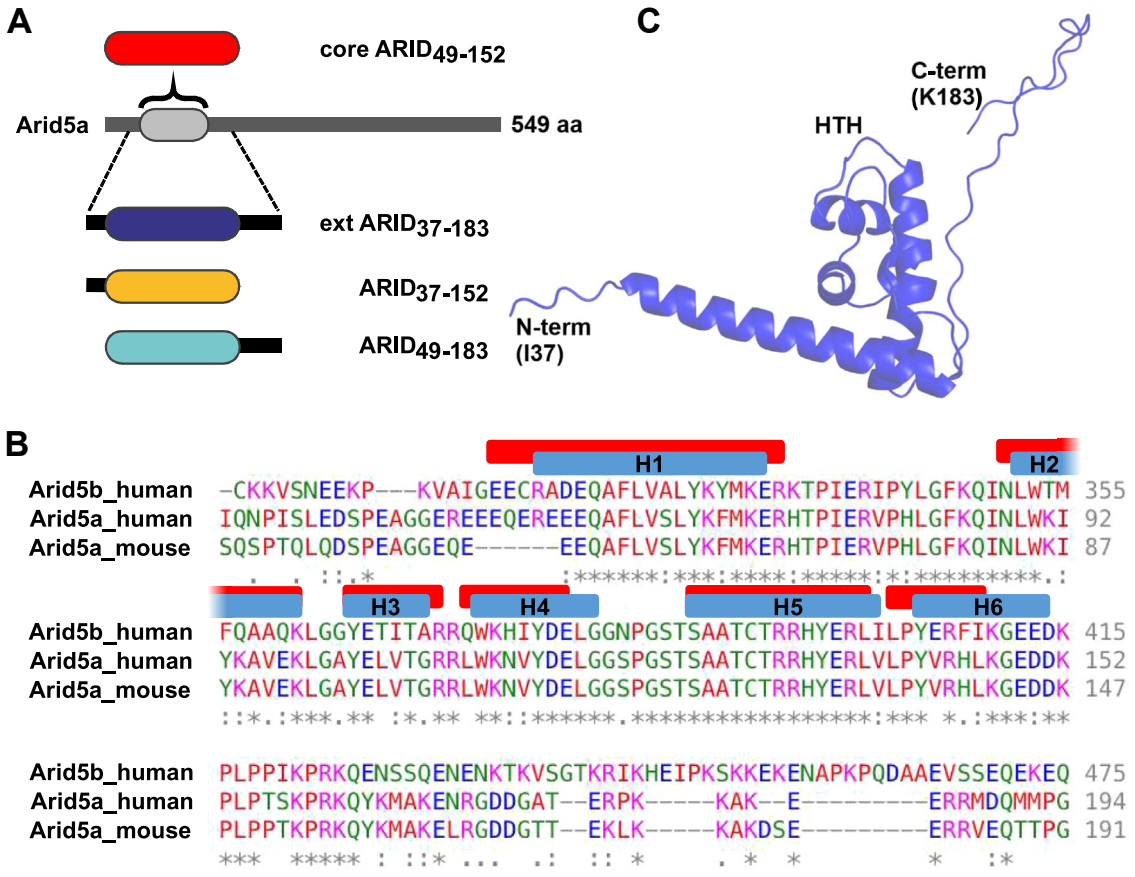


Figure 1. The Arid5a ARID domain is a minimal ARID core motif. *A*, domain architecture of full-length (fl)-Arid5a and overview of ARID domain boundaries used in this study. *B*, comparison of human and mouse Arid5a ARID sequences with the human Arid5b ARID, as obtained by Clustal Omega (88). The ARID secondary structure elements (helices H1 to H6) are indicated above the sequence for human Arid5b (*blue*, PDB 1IG6) and human Arid5a (*red* (68)). For full sequence alignment of Arid5a human with mouse, see Fig. S2. *C*, structural model of the Arid5a core ARID domain as derived from a RoseTTAFold (75) run with the sequence of $ARID_{37-183}$. The model represents member 1 of an ensemble (see Fig. S6).

binding protein (RBP), stabilizing the mRNA of *Il-6* (13), thus counteracting the degradation mediated by the regulatory RBPs Regnase-1 and Roquin (20). Follow-up work suggested additional targets of Arid5a in an immunological context, among them *Stat3* (21) and *Ox40* (22), soon categorizing the protein as pro-inflammatory factor. In these studies, the ARID domain is claimed to interact with particular RNA stem-loop structures, known to exist in *Stat3* (21), *Ox40* (23, 24), and possibly also the *Il-6* 3′-UTR, suggesting shape-specific recognition of RNA *cis*-regulatory elements similar to Regnase-1 and Roquin (23, 25–27). Though RNA recognition has indirectly been attributed to the Arid5a ARID domain in mice (21), a direct proof for its interaction with RNA is still missing. At the same time, we have no insight how Arid5a uses its ARID domain to distinguish between specific DNA- and RNA-binding and whether flanking regions are involved.

Arid5a was initially found differentially expressed in tissues unrelated to the adaptive immune system, but with a clear nuclear localization, in line with transcriptional regulation (17). Recent work has extended both findings with the protein being able to shuttle upon lipopolysaccharide stimulation in immune cells (28), a prerequisite for transcript protection against cytoplasmic nucleases. It remains unexplored how RNA motif preferences of the Arid5a ARID domain are related to this, but they should exist independent of cellular localization. There is to date no systematic study identifying the transcriptome targeted by Arid5a independently of its immunological role.

In Arids, the so-called core ARID can appear as N- and/or C-terminally extended domain (Fig. 1), that is, additional helices (H0, H7) or intrinsically disordered regions (IDRs) enlarge the interface with nucleic acids and likely modify preferences. The strength of IDRs, modifying the function and specifics of DBPs, has recently been brought up for transcription factors (TFs), many with a previously unknown affinity to RNA mediated through their IDRs (29). While well conceivable as a more general feature, *for example*, for compartmentalizing (co)-transcriptional processes, no structural proofs exist for a simultaneous or mutually exclusive interaction of protein domains with RNA and DNA. Similarly, the lack of high-resolution ARID structures with DNAs—with only few exceptions—has hindered us from identifying concepts of specific target recognition through core domains and in combination with flanking regions. As such, most motifs assigned to individual Arids are derived from genetic studies or do not unambiguously define the ARID domain as responsible for interactions. And, the currently known studies have not addressed binding of RNAs by Arids.

We here present a systematic analysis of the Arid5a ARID domain towards specific DNA- and RNA-binding. Using a combination of NMR and EMSAs, we compare nucleic acid recognition of the core with the IDR-extended ARID. Our work provides unambiguous proof for the dual nucleic acid recognition by the domain. We provide in-depth evidence for its preference towards specific AT-DNA motifs, while RNA Bind-n-Seq (RBNS) reveals a preference for an unexpected CAGG-CAG consensus motif, accompanied by a general preference for AU-rich motifs (Data Table S1). We find that the ARID-flanking IDRs strongly modulate affinity for complex RNA and nonspecific DNA sequences. We show that Arid5a exists in the nucleus under unstressed conditions and perform the first individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP2) experiment to map Arid5a-binding sites throughout the transcriptome and identify an *in vivo* (A) U-rich consensus target RNA motif. We find Arid5a to bind RNA-processing related nascent transcripts. While we suggest Arid5a to mainly exert DBP functions, we show that extended ARID domains of other Arids have a similar capacity to interact with RNA. Thus, we stress the idea of Arids as more general dual nucleic acid–binding proteins. We suggest an essential role of the ARID-extending IDRs in nucleic acid recognition, in particular for—but not restricted to—Arid5a.

Results

Highly conserved ARID domains show distinct DNA-binding preferences

Doubts have evolved over recent years to whether all name giving AT-rich interactive domains of the 15 human Arids share exclusive preference for AT-rich sequences (recently reviewed by Korn and Schlundt (5)). Indeed, controversial sequence preferences reported for a number of Arids gave reasons to unbiasedly probe for individual target sequences (1, 30–32). We thus picked representative ARID domains from three subfamilies that had been described to target DNA with different sequence preferences (Fig. 2*A*). While some literature describes Arid1a to bind DNA nonspecifically through its ARID domain, the ARID domains of Arid5b and JARID1a are suggested to be specific for AT- and GC-rich dsDNA, respectively (1, 16, 32). We used fluorescently labeled AT- or GC-rich dsDNA to monitor preferences of these ARID domains in EMSAs (Figs. 2*B* and S1). Interestingly, the ARID domains of Arid1a and JARID1a are less specific for AT-rich DNA than the ARID domain of Arid5b (see also Fig. S1). Furthermore, and in line with multiple studies (30, 31, 33), the Arid1a ARID domain displays similar affinity for a GC-rich dsDNA, supporting its nonspecificity for DNA. In summary, the data argue against ARID domains as exclusive AT-binders and raise the need to carefully *de novo* define and interpret available consensus motifs for the individual domains despite their highly conserved fold.

The Arid5a ARID domain uses an extended binding interface with AT-rich DNA

Because of the above-described variance in the DNA-binding preferences of ARID domains, we first decided to investigate the Arid5a ARID domain's sequence preference. Although 9mer dsDNA sequences were sufficient for binding, we observed a minor increase in affinity with longer dsDNAs plateauing at 13 bp length (Fig. S3) and thus used 13mers for our study. In EMSAs, we tested $ARID_{37\text{-}183}$ against fluorescently labeled dsDNAs, either GC-rich based on Jarid1a binding to a "CCGCCC" motif (32) or with variations of a central AT-stretch based on a published motif for the closely related Arid5b ARID (16) (Table S3 and Fig. S3). We find that the Arid5a ARID domain clearly favors AT-rich dsDNA, and a central "AATA" motif is important as evident from EMSA-derived affinities around 0.8 to 2.3 μM (13merAT WT, var1, var2, and var4) contrasting the DNAs without four consecutive A/Ts, for which no K_D could be derived (Fig. S4).

To investigate differential complex formation on the residue-resolved level, we next used NMR and performed ^{1}H-^{15}N-heteronuclear single quantum coherence (HSQC) titrations of either 13merAT or 13merGC dsDNA to the extended $ARID_{37\text{-}183}$ and plotted the combined chemical shift perturbations (CSPs) over the protein sequence (Fig. 3, *A* and *B*, see Experimental procedures section for details regarding CSP calculation). With this experiment, we sought to (i) identify the precise interface(s) of the ARID domain with DNA beyond its core fold and (ii) spot potential differences in CSP patterns caused by the two dsDNA ligands. The titrations clearly show that $ARID_{37\text{-}183}$ binds to both the 13merAT and the 13merGC dsDNA. However, different exchange regimes—fast exchange for 13merGC and intermediate exchange for 13merAT (insets Fig. 3*A*)—support the significantly higher affinity of Arid5a ARID to AT-rich than GC-rich DNA observed in EMSAs (Fig. S4). Of note, maximum CSPs within core ARID residues are much smaller for 13merGC than for the AT-rich DNA (Fig. 3, *A* and *B*). Interestingly, CSP differences of flanking IDR residues, especially the C-terminal extension (residues 150–160), are less pronounced between GC- and AT-rich DNA targets than within the core domain, suggesting the contribution of IDRs to DNA-binding is less or nonspecific.

From the CSP plots, we concluded that the Arid5a ARID domain interacts with DNA through residues in loop L1 and the HTH motif (H4-L2-H5) (Fig. 3*B*). This is in good agreement with the reported DNA-binding interface found for other ARID domains (16, 34–36) and an R-to-A mutant in murine Arid5a (corresponding to R133 in the human version, see Fig. 1*B*) incapable of DNA-binding (21). Importantly, our data reveal an additional contribution of residues K152 and L154 within the C-terminal extension. Mapping significant CSPs obtained for the AT–DNA interaction on an Arid5a ARID RoseTTAFold model clearly shows them to cluster in the canonical DNA-binding interface (Fig. 3*C*).

To investigate the potential contribution of both N- and C-terminal extensions to DNA-binding in more detail, we created constructs of the ARID domain with either the separate N- ($ARID_{37\text{-}152}$) or C-terminal ($ARID_{49\text{-}183}$) extension and compared their DNA interaction with 13merAT to the core domain ($ARID_{49\text{-}152}$) and the extended $ARID_{37\text{-}183}$ (Figs. 3*D*, S7, and S8). In contrast to the N-terminal IDR, the C-terminal

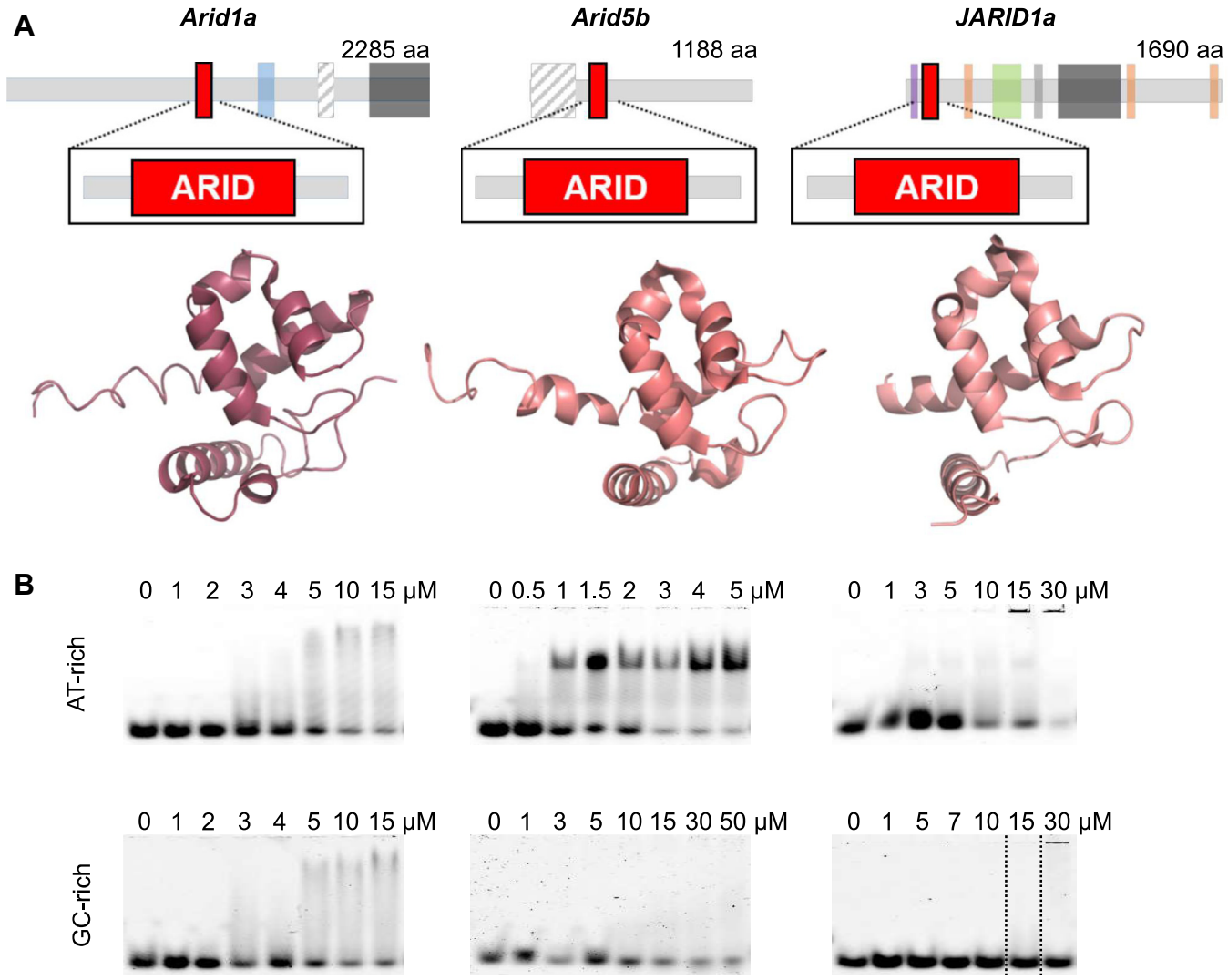


Figure 2. DNA-binding preferences of ARID domains from three different human Arid sub-families. *A*, domain architecture of Arid1a, Arid5b, and Jarid1a, with the ARID domain indicated by a *red* box and their structures depicted below (Arid1a: PDB 1RYU (35); Arid5b: PDB 1IG6, unpublished; and Jarid1a: PDB 2JXJ (32)). Other annotated domains are HIC1 (*light blue*), BAF250-C (*dark gray*), JMJN (*purple*), JMNC (*green*), PHD (*orange*), and ZnF (*light gray*). Merely predicted domains are striped *black* and *white*. *B*, the DNA-binding preference of extended ARID domains comprising the minimal core ARID plus 18 N- and 20 C-terminal residues were studied by EMSAs with either 10 nM 13merAT or 13merGC fluorescently labeled dsDNA (Table S3). Protein concentrations are shown above each lane in μM, and all experiments have been carried out in standard Arid5a buffer. Of note, the EMSA gel for Jarid1a with 13merGC has been spliced to skip additional concentrations to better align with the 13merAT EMSA above (indicated by the lines). Uncropped gels (and replicates) are given in the source data file.

extension shifted the ARID–DNA interaction towards an NMR-observed intermediate-to-slow exchange regime (Figs. 3*D* and S7), supported by observable changes in the EMSA patterns (Fig. S9). The latter does not only support the higher affinities for C-terminally extended ARID constructs ($ARID_{49\text{-}183}$ and $ARID_{37\text{-}183}$) but also reveals the formation of more prominent complex bands for these two constructs, indicating DNA–protein complexes sufficiently tight to maintain their integrity in the native gel condition and that are less pronounced in ARID domains devoid of the C-terminal extensions.

To confirm sequence-specific DNA recognition in the 13merAT DNA compared to 13merGC, we titrated increasing concentrations of $ARID_{37\text{-}183}$ to the respective DNAs and monitored effects on imino protons (Fig. 3*E*). We undertook a complete assignment of 13merAT imino resonances, which allowed a base pair–resolved analysis (Fig. S5). In line with the EMSA-observed stable complex formation, we found strong line broadening within the 13merAT DNA after addition of the protein. As expected, this effect is more pronounced for the central base pairs of the 13merAT DNA suggested to form the interface with ARID (Fig. 3*C*) and including the central AATA motif, as compared to the flanking terminal base pairs (compare residues G2/12 and T7), which, however, still show weak CSPs. In contrast, the 13merGC merely displayed minor line broadening upon $ARID_{37\text{-}183}$ addition, more evenly distributed over all imino signals. This supports a weak, but nonspecific interaction with the GC-rich DNA, driven by electrostatic interactions with the DNA backbone rather than base-specific contacts.

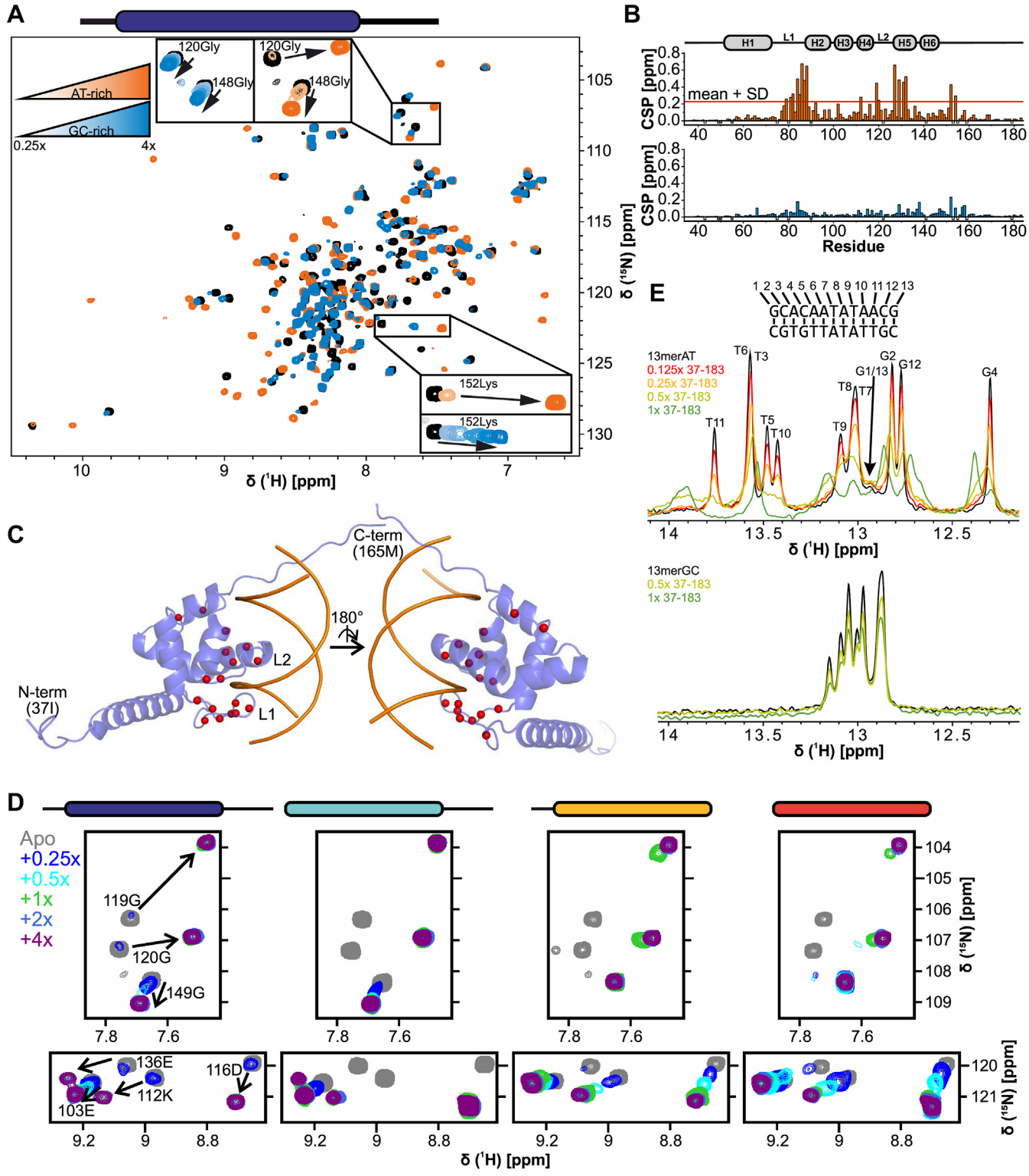


Figure 3. Arid5a interacts with AT-rich DNA through loops in its core ARID and the C-terminal IDR. *A*, ^{1}H-^{15}N-HSQC overlay of $ARID_{37-183}$ alone and after titration with 4-fold 13merAT (*orange*) or 13merGC (*blue*). Insets show all titration points and assignments. Spectra were recorded at a constant protein concentration of 70 μM with 17.5, 35, 70, 140, and 280 μM dsDNA at 600 MHz and 298 K. *B*, chemical shift perturbation (CSP) plot of $ARID_{37-183}$ upon titration with 4-fold 13merAT (*upper panel*) or 13merGC (*lower panel*). Negative bars in *light gray* and *gray* show prolines and unassigned residues, respectively. Significantly shifting peaks, that is, above mean + 1 SD, are shown in Fig. S8*B*. See methods section for details on the quantification of CSPs in this manuscript. *C*, RoseTTAFold (75) model of $ARID_{37-183}$ as picked from an ensemble (see Fig. S6) highlighting highest CSPs from (b) (above mean + 1 SD) in *red*. For simplification, only residues 37 to 165 are shown. The DNA is shown for orientation revealing the putative interface (see Experimental procedures section). *D*, ^{1}H-^{15}N-HSQC zoom-ins of four Arid5a ARID constructs with/without N- and/or C-terminal extension—as indicated above—showing spectra of proteins alone and when titrated with 13merAT dsDNA. Spectra were recorded at a constant protein concentration of 70 μM with 17.5, 35, 70, 140, and 280 μM dsDNA at 600 MHz and 298 K. *E*, 1D imino proton spectra of 13merAT (*upper panel*) and 13merGC (*lower panel*) upon titration of DNAs with $ARID_{37-183}$. See also Fig. S5. Spectra were recorded at a constant DNA concentration of 80 μM with 10, 20, 40, and 80 μM protein at 600 MHz and 298 K. All experiments have been carried out in standard Arid5a buffer.

Mutational studies of Arid5a confirm key residues for DNA-binding

To confirm the ARID DNA-binding interface, we designed protein mutants by replacing selected residues, located either in L1 or the HTH motif, by alanine. Residues were chosen based either on their high CSPs observed in the $ARID_{37-183}$ titration with 13merAT (K85 and Q86) or on literature and sequence comparison to other Arids—especially Arid5b—and their key DNA-binding residues (R78A, R109A, T125A/S126A) (1, 5, 16). Mutations were introduced both in the core $ARID_{49-152}$ and extended $ARID_{37-183}$ background, to further elucidate the role of IDRs in this context. As the spectra for the mutants only showed minor local CSPs (Fig. S10), we were able to unambiguously transfer most assignments from the WT spectra to the mutants (see also Experimental procedures section). ^{1}H-^{15}N-HSQC spectra of proteins alone and in the presence of 2-fold molar excess 13merAT dsDNA were recorded to quantify the effect of single, double, triple and quadruple mutations on DNA-binding (Figs. 4, S11, and S12).

Mutation of T125 and S126, located in the HTH at the transition of loop 2 to helix 5, strongly impaired DNA-binding of the core ARID domain (Fig. 4*A*). This is in line with their expected role in making specific contacts with an AT base pair in the DNA major groove, as suggested by the complex structure of the closely related Arid5b ARID domain with AT-rich dsDNA (15). Loop 1 mutations (K85A/Q86A) on the other hand—despite high CSPs (Fig. 3)—did not inhibit DNA-binding and likely exhibit no crucial DNA contacts. Of note,

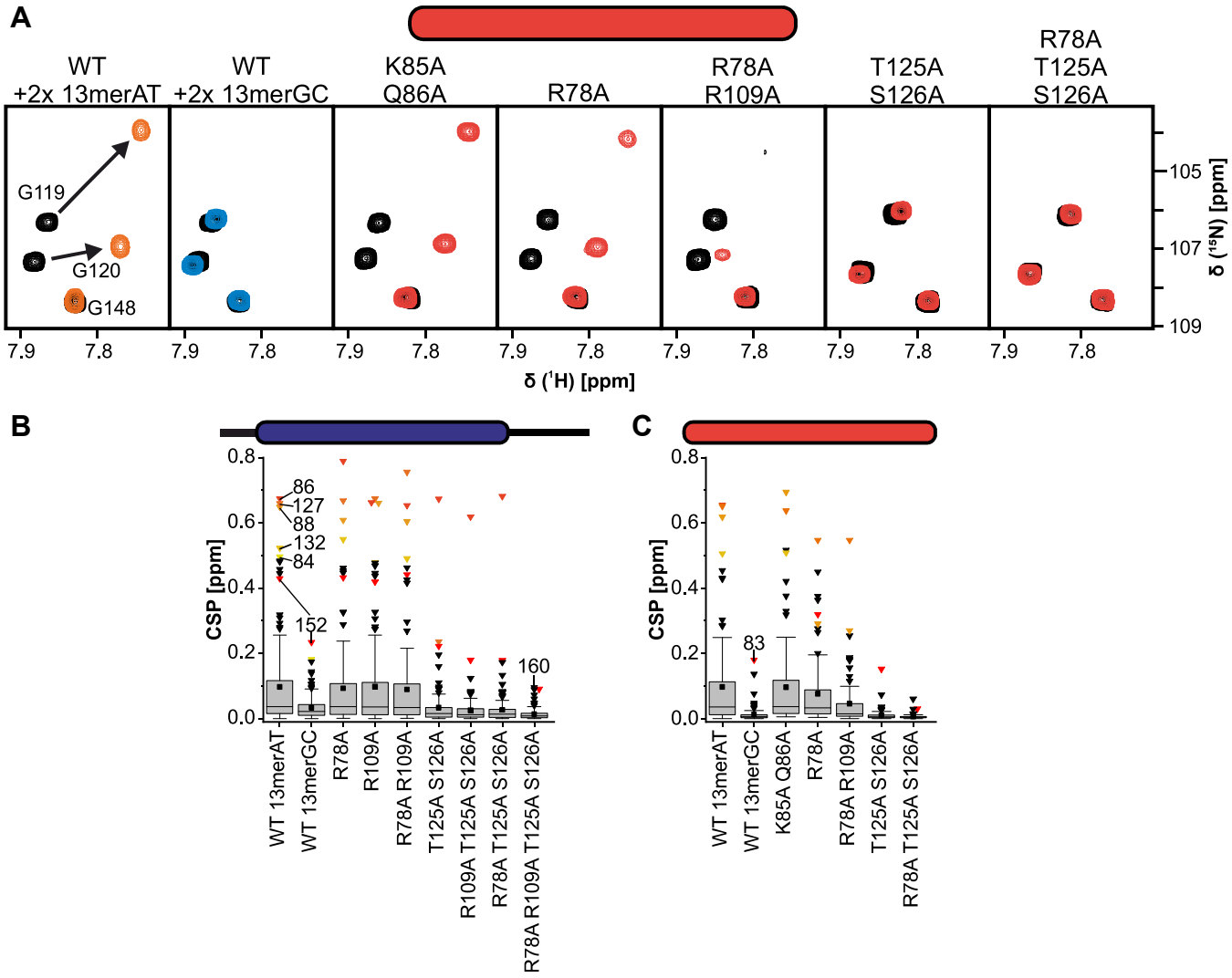


Figure 4. Mutational studies of the Arid5a ARID domain reveal the central core-binding residues. *A*, ^{1}H-^{15}N-HSQC zoom-ins of $ARID_{49-152}$ WT and mutants overlaying apo protein spectra (*black*) with those of samples containing 2-fold 13merAT dsDNA (*orange/red*) or 2-fold 13merGC dsDNA (*blue*). Spectra were recorded in standard Arid5a buffer at a protein concentration of 70 μM with 140 μM dsDNA for the complex sample at 600 MHz and 298 K. *B*, boxplot of CSP quantifications of $ARID_{37-183}$ WT and mutants upon addition of 2-fold 13merAT dsDNA. *C*, boxplot of CSP quantifications of $ARID_{49-152}$ WT and mutants upon addition of 2-fold 13merAT dsDNA. For comparison and as a reference, each boxplot also shows the CSPs of the WT with 2-fold 13merGC dsDNA. The box represents the interquartile range from the 25th to the 75th percentile with a whisker coefficient of 1.5 for outliers and further outliers shown as *black/colored triangles*. The median is shown as a *horizontal line* within boxes and mean values are indicated by *black squares*. The five highest CSPs for 13merAT and the highest CSP for 13merGC with $ARID_{37-183}$ or $ARID_{49-152}$ are color coded for direct comparison between WT and mutants. Source data with all CSPs are provided as a Source Data file.

the effect of DNA-binding mutations was less pronounced in presence of the extending IDRs, evident when comparing global CSPs between $ARID_{37\text{-}183}$ and $ARID_{49\text{-}152}$ (Fig. 4, *B* and *C*). We thus conclude that the C-terminal extension to the ARID domain can compensate for mutations within the core ARID domain by a general, but nonspecific mode of increasing affinity for dsDNA.

In vitro *RNA-binding of the Arid5a ARID domain*

Arid5a was recently identified to stabilize the *Ox40* mRNA in murine CD4+ T cells by direct interaction with a stem-looped structure in its 3′-UTR, known as an alternative decay element (ADE) (22). In doing so, Arid5a interferes with controlled degradation of the *Ox40* transcript by the nuclease Regnase through targeting the same *cis*-regulatory element. We wondered if the ARID domain in Arid5a was responsible for the underlying complex formation with the RNA stem-loop and used NMR spectroscopy to observe atom-resolved binding of the ARID domain to the ADE element (Figs. 5 and S13). Interestingly, the minimum core $ARID_{49\text{-}152}$ showed only marginal interactions, even with a high stoichiometric excess of the 19-nt ADE (Figs. 5, *D* and *E* and S13), when judged by the magnitude of CSPs compared to the AT-DNA before (see Fig. 3*B* for comparison), and in fact is rather reminiscent of binding to GC-DNA. However, similar to DNA-binding, the basic C-terminal extension in $ARID_{37\text{-}183}$ contributed to an increased ADE interaction as indicated by both visible CSPs within this region and slightly increased CSPs for the core $ARID_{49\text{-}152}$ domain (Fig. S13*B*), suggesting the extension to carry an essential role in Arid5a-based mRNA regulation *in vivo*.

While our data are the first structural proof of a direct ARID–RNA interaction, we were surprised by the observed

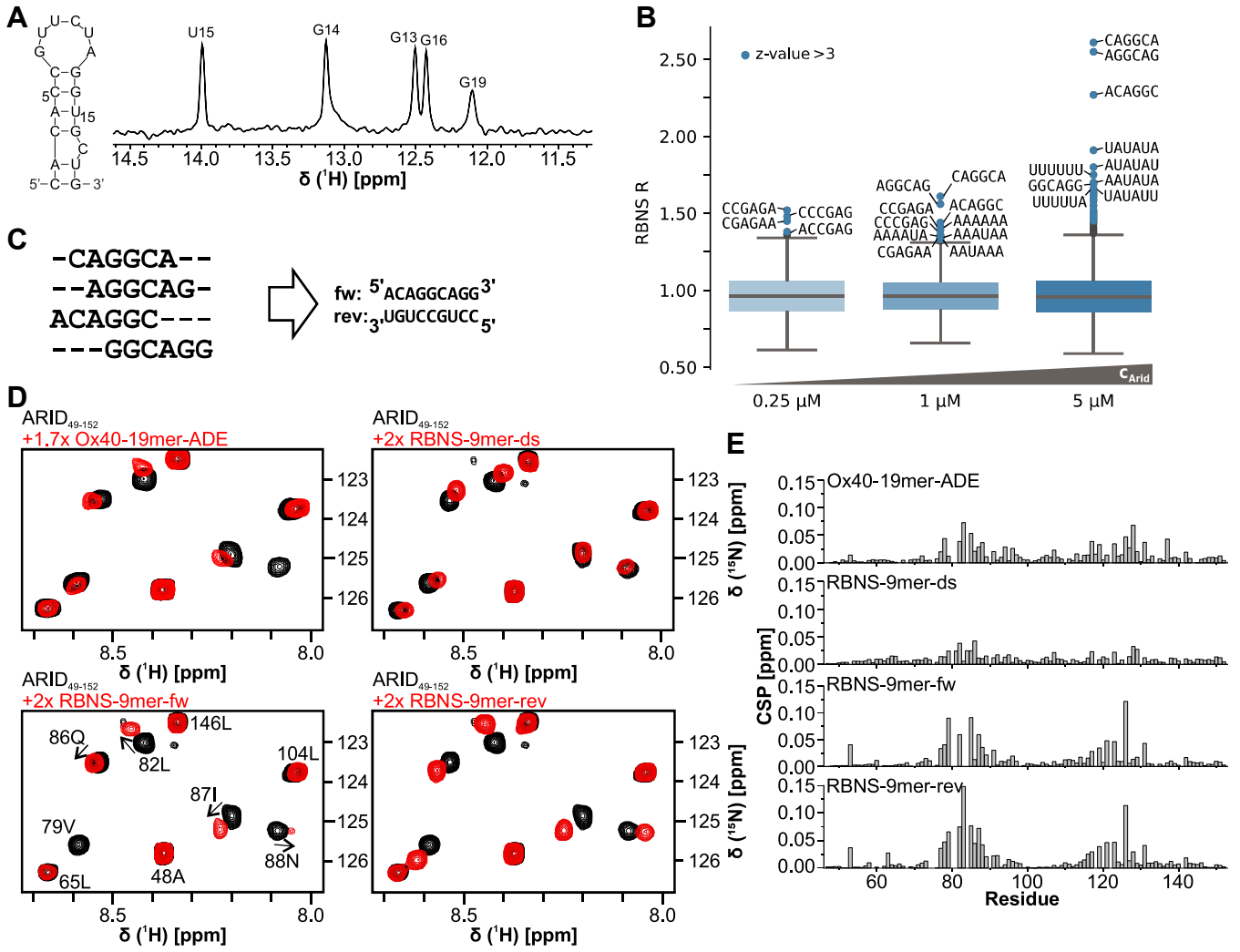


Figure 5. The Arid5a ARID domain binds RNA motifs with moderate affinity. *A*, the *Ox40*-ADE forms a stem-loop element, confirmed by the depicted imino-proton spectrum. Assignments have been transferred from Janowski *et al.*, 2016 (23). Spectrum measured with 20 µM RNA at 600 MHz and 298 K. *B*, enrichment of all 6-mers at 0.25, 1, and 5 µM $ARID_{49\text{-}152}$ concentration from RNA bind-and-seq (RBNS). Values greater than three SDs above the mean are highlighted in *blue*. For the highest significant, 10 motif sequences are given. RBNS experiments have been carried out in 25 mM Tris–HCl, pH 7.5, 150 mM KCl, 3 mM $MgCl_2$, 0.01% Tween, 500 µg/ml BSA, 1 mM DTT. *C*, RBNS-based 9mer sequences that can be clustered from the enriched 6mers containing AGGC. *D*, zoom-ins of ^{1}H-^{15}N-HSQC spectra of apo $ARID_{49\text{-}152}$ (40 µM for ADE/70 µM for RBNS-RNAs) overlaid with 1.7-fold molar excess of ADE RNA or 2-fold molar excess of RBNS-9mer RNAs. Spectra were measured at 600 MHz (for ADE RNA) or 700 MHz (for RBNS-RNAs) at 298 K in standard Arid5a buffer. *E*, CSP plots of $ARID_{49\text{-}152}$ upon addition of *Ox40*-ADE or RBNS-9mer RNAs as shown in (*D*).

moderate binding affinity. To this end, we decided to set up an unbiased search for a general consensus RNA target motif of the core ARID fold, which had not been undertaken prior to this study. We thus performed RBNS to test the ARID domain's capability to interact with specific RNA motifs. This *in vitro* high-throughput assay allows to identify the binding preferences of an RBP (37, 38). A pulldown is performed with a 20 nt random RNA pool flanked by short constant adapter sequences with different concentrations of Strep-tagged RBP ($ARID_{49\text{-}152}$: 0.25, 1, 5 μM). The constant regions are then used to add sequencing adapters for subsequent analysis by next-generation sequencing. We obtained ∼35 to 50 million unique reads for each ARID protein concentration. By comparing the frequencies of k-mers in the input library with the pulldown libraries, we were able to identify enriched 6-mers (Fig. 5*B* and Data Table S1). The motifs found here can be broadly divided into two types: (i) those that contain AGGC as a core motif and no uracil and (ii) those that are rich in AU. Analysis of enriched 5- and 7-mers yielded similar results (Fig. S14*A* and Data Table S1). Complex binding motifs were calculated to get an insight into the environment of the binding sites (Fig. S14*C*). Clustering of the AGGC core motifs results in the 9-mer (A)CAGGCA(GG) (Fig. 5*C*). Based on this, we designed two reverse-complementary 9-mer RNAs in agreement with our minimal length for affine DNA-binding (Figs. 5*C* and S3). The structural features of the identified binding motifs were estimated by calculating the average base pairing probability with RNAfold *in silico*. Here, the AGGC-containing motifs show almost no base pairing and appear unstructured, while the AU-rich ones show no particular preference for being structured or unstructured (Fig. S14*D*).

We tested $ARID_{49\text{-}152}$ binding to the (A)CAGGCA(GG) motif both as ssRNA with the forward (fw) and reverse (rev) strand individually as well as their annealed dsRNA (Fig. 5*C*). Comparison of CSPs similarly to RBNS data reveals a clear preference of ss versus dsRNA; yet within the ssRNA context, specificity for a defined motif is not particularly pronounced (Figs. 5, *D* and *E* and S14*D*). Unexpectedly, the protein regions interacting with ssRNA are the same as are interacting with dsDNA, with residues 80 to 90 and 120 to 130 showing the highest CSPs, which raises the question of a so-far unknown single-stranded nucleic acid–binding mode by ARID. Furthermore, titrations with the RBNS-based ssRNAs indeed show stronger CSPs compared with the ADE despite the larger size and the previously suggested specificity of Arid5a for the ADE interaction mediated by the ARID domain (22). Those findings are in line with the lack of ADE-related motifs observed in RBNS. Altogether, our data suggest a previously unknown RNA sequence preference specific to the core ARID domain, which may indicate possibly unidentified (m)RNAs bound by Arid5a *in vivo*.

IDRs increase the ARID RNA-binding affinity in a length-dependent manner

To this stage, our data suggest the ARID core domain to prefer ssRNA over dsRNA and folded RNA and an obvious contribution of the C-terminal IDR to the binding affinity for the ADE. Consequently, we wondered how the ARID-extending IDRs would influence the RNA-binding capacity of other RNA sequences that are longer and more complexly folded. We started with a prolonged dsRNA (19mer_ds), with a central AU-rich core and GC-stabilized flanking regions (Figs. 6 and S15). We recorded ^{1}H-^{15}N-HSQCs of the minimal core and the extended ARID domain in the presence and absence of RNA. The core ARID domain interacted only weakly with the 19mer_dsRNA, indicated by minor CSPs in the spectral overlay (Fig. 6*A*). In contrast, severe line broadening of the IDR-extended $ARID_{37\text{-}183}$ suggested a strongly increased interaction with the 19mer_dsRNA (Fig. 6*B*). This suggests an increasing contribution of IDRs in the case of extended RNA stretches, likely reasoned by the steric possibilities and high density of charges.

To test this hypothesis, we used a previously described physiologically relevant target sequence (13) located in the 3′-UTR of the *Il-6* mRNA. The 129-nt sequence likely represents the naturally occurring RNA-folding complexity, providing stretches of ssRNA (loops) and base-paired regions (Fig. S15*G*). Strikingly, while the ARID core domain still binds with only moderate affinity to the RNA in 1.2-fold molar excess, the extended $ARID_{37\text{-}183}$ strongly interacts already in substoichiometric concentrations (<0.1×) (Fig. 6, *C* and *D*). Our results suggest that the ARID IDRs drive RNA-binding affinity in dependence of the provided density of negative charge. To confirm this hypothesis, we further compared EMSA-derived apparent affinities to RNAs of increasing length and see a clear correlation between affinity and RNA length (Fig. 6*E*). Likely, this effect is also supported by more than one protein binding to the larger RNAs (see *right* panel).

In conclusion, our results show that Arid5a is principally capable of interacting with RNA. The intrinsically low affinity of the core ARID domain is compensated by its IDR extensions in a nonspecific manner. Consequently, those nonspecific interactions favor RNAs of increasing size, while the core ARID domain remains restrictive to very specific sequences.

iCLIP2 reveals that Arid5a binds to ssRNA in cells with a preference for U-rich stretches

Our data so-far suggest that Arid5a is capable of tight interactions with available RNAs, albeit primarily driven by charge interactions. On the other hand, the core ARID domain shows a particular preference for short motifs, which still may steer specific interactions with RNAs more selectively. We speculated that those motifs will be embedded in a larger RNA context *in vivo*, where RNP complex formation will be supported and modulated by the presence of the ARID-flanking regions and potentially further regions of Arid5a beyond that. Motivated by those assumptions, we performed individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP2) (39) with full-length Arid5a in murine P19 cells. To our knowledge, this has been the first CLIP experiment carried out with an Arid protein to date.

Murine P19 cells express *Arid5a* mRNA (Fig. S16*A*), but a specific antibody suitable for iCLIP is lacking. Hence, we

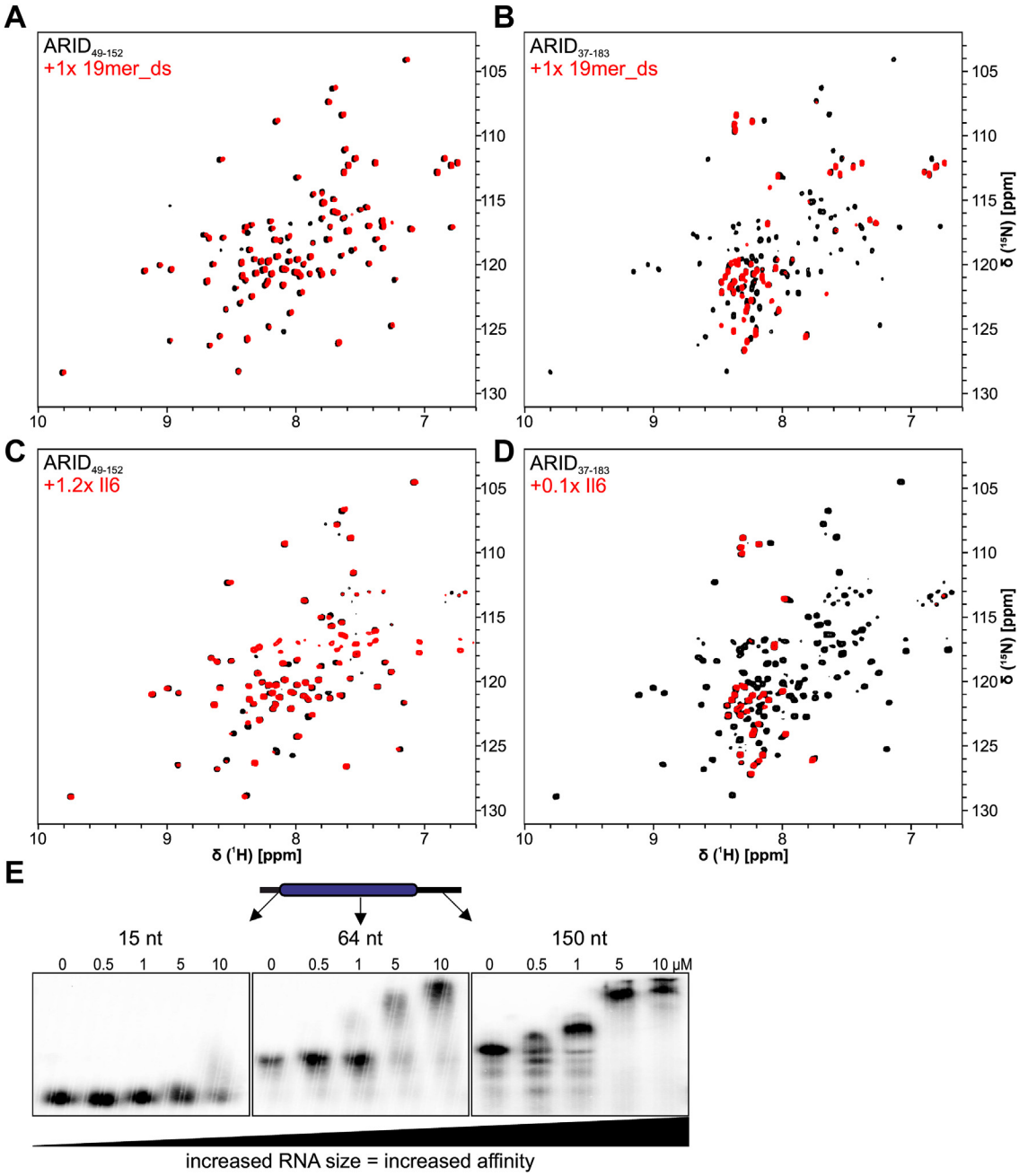
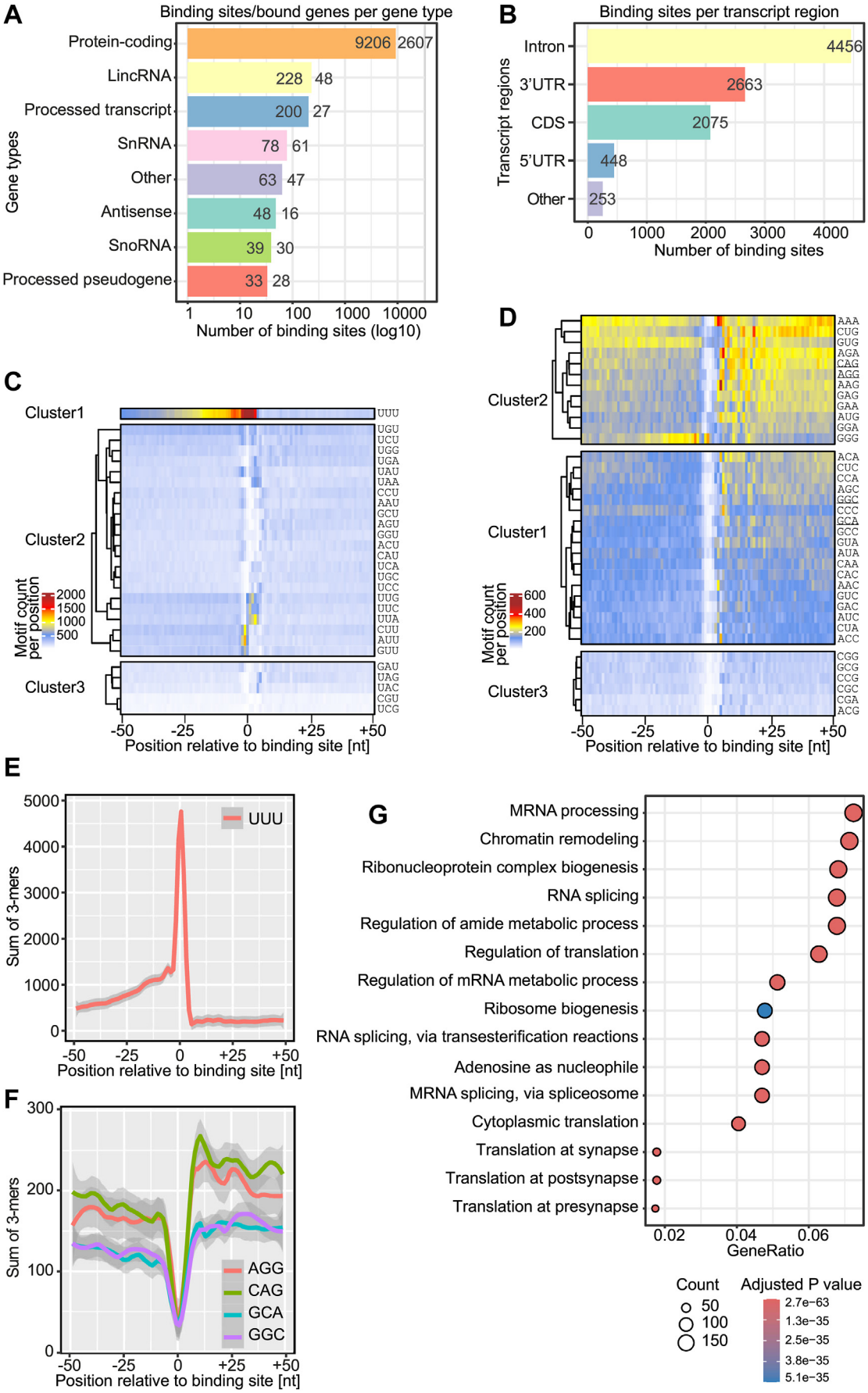


Figure 6. Arid5a ARID domain binding to RNA. *A* and *B*, ^{1}H-^{15}N-HSQCs of ARID$_{49\text{-}152}$ (*A*) or $_{37\text{-}183}$ (*B*) without RNA (*black*) or with 1-fold 19mer_ds RNA (*red*). *C* and *D*, ^{1}H-^{15}N-TROSY-HSQC of ARID$_{49\text{-}152}$ (*C*) or $_{37\text{-}183}$ (*D*) without RNA (*black*) or with 1.2-fold and 0.1-fold human Interleukin-6 mRNA (*red*), respectively. Protein concentration for all NMR measurements was 50 µM. Spectra were measured at 600 (*A* and *B*) or 900 MHz (*C* and *D*) and 298 K. *E*, EMSAs showing that increased RNA size leads to more affine binding by ARID$_{37\text{-}183}$. All experiments have been carried out in standard Arid5a buffer.

generated an expression plasmid with the murine full-length Arid5a (see sequence alignment and conservation with human Arid5a in Fig. S2) fused to a C-terminal GFP-tag (mArid5a-GFP, Table S2). P19 WT cells were transfected in three replicates and subjected to the iCLIP2 procedure using an anti-GFP antibody (Fig. S16*B*, see Experimental procedures). All three replicate experiments were highly reproducible and gave rise to more than 7 million crosslinks (Fig. S16, *B* and *C*) and 9895 binding sites with an optimal width of 7 nucleotides (nt) that were used for downstream analysis. Arid5a-binding sites are found predominantly in 2607 protein-coding genes but also in 48 lncRNAs and other noncoding RNAs (Fig. 7*A*). A 3-mer enrichment analysis reveals that Arid5a crosslinks preferentially at U-rich stretches (Fig. 7*C*). The ramp-like enrichment pattern indicates that Arid5a sits at the very 3′-end of polyU stretches (Fig. 7, *C* and *E*). We hypothesized that this positioning of Arid5a may result from a specific interaction downstream of the polyU stretches *via* the ARID domain, while its adjacent IDRs crosslink to Us in a fixed Arid5a-RNA orientation. U-rich stretches were also found enriched by RBNS (Fig. S14, *B* and *C*), suggesting that this preference is not merely due to a UV crosslinking bias for U. To test our


A
Binding sites/bound genes per gene type
Protein-coding
LincRNA
Processed transcript
SnRNA
Other
Antisense
SnoRNA
Processed pseudogene
Gene types
9206 2607
228 48
200 27
78 61
63 47
48 16
39 30
33 28
1 10 100 1000 10000
Number of binding sites (log10)
B
Binding sites per transcript region
Intron
3'UTR
CDS
5'UTR
Other
Transcript regions
4456
2663
2075
448
253
0 1000 2000 3000 4000
Number of binding sites
C
Cluster1
Cluster2
Cluster3
Motif count per position
2000
1500
1000
500
UUU
UGU UCU UGG UGA UAU UAA CCU AAU GCU AGU GGU ACU CAU UCA UGC UCC UUG UUC UUA CUU AUU GUU
GAU UAG UAC CGU UCG
-50 -25 0 +25 +50
Position relative to binding site [nt]
D
Cluster2
Cluster1
Cluster3
Motif count per position
600
400
200
AAA CUG GUG AGA CAG AGG AAG GAG GAA AUG GGA GGG
ACA CUC CCA AGC GGC CCC GCA GCC GUA AUA CAA CAC AAC GUC GAC AUC CUA ACC
CGG GCG CCG CGC CGA ACG
-50 -25 0 +25 +50
Position relative to binding site [nt]
E
Sum of 3-mers
5000
4000
3000
2000
1000
0
UUU
-50 -25 0 +25 +50
Position relative to binding site [nt]
F
Sum of 3-mers
300
200
100
0
AGG
CAG
GCA
GGC
-50 -25 0 +25 +50
Position relative to binding site [nt]
G
MRNA processing
Chromatin remodeling
Ribonucleoprotein complex biogenesis
RNA splicing
Regulation of amide metabolic process
Regulation of translation
Regulation of mRNA metabolic process
Ribosome biogenesis
RNA splicing, via transesterification reactions
Adenosine as nucleophile
MRNA splicing, via spliceosome
Cytoplasmic translation
Translation at synapse
Translation at postsynapse
Translation at presynapse
0.02 0.04 0.06
GeneRatio
Count
50
100
150
Adjusted P value
2.7e-63
1.3e-35
2.5e-35
3.8e-35
5.1e-35

hypothesis, we searched for enriched 3-mers downstream of the Arid5a-binding sites. Of note, we observe a general enrichment of AG-rich 3-mers (Fig. 7*D*). Moreover, the four 3-mers CAG, AGG, GGC, and GCA contained within the RBNS-enriched consensus motif (A)CAGGCA(G) (Fig. 5, *B* and *C*) are consistently enriched downstream of the binding sites (Fig. 7, *D* and *F*), suggesting that Arid5a shows a preference for this RNA motif also *in vivo*. The positioning of these 3-mers downstream to the polyU stretches might help to position the binding of the ARID domain.

Looking at the bound transcripts, we observe Arid5a across all transcript regions, including introns, 3′-UTRs, and exons (Fig. 7*B*), indicating that Arid5a binds to pre-mRNAs in the nucleus. Arid5a targets are enriched for transcripts involved in mRNA processing, chromatin remodeling, and translation regulation (Fig. 7*G*). Altogether, our data suggest that Arid5a is a sequence-specific dual DNA- and RNA-binding protein that binds to a subset of transcripts *in vivo* and may have an accessory function in chromatin-related transcript processing or in transcription regulation, possibly in an RNA-supported context (see Discussion).

Arid5a is a strictly nuclear protein and colocalizes with heterochromatin

The proposed dual function of Arid5a in gene regulation both *via* interaction with DNA/chromatin and protection against mRNA degradation in the cytoplasm (13, 21, 22) requires the protein to be present in both the nucleus and the cytoplasm. However, our iCLIP2 data show that Arid5a binds preferentially to unspliced pre-mRNAs, suggesting an exclusively nuclear function of Arid5a associated with chromatin. To test the subcellular localization of Arid5a under normal conditions, we performed confocal fluorescence microscopy of P19 WT cells transfected with mArid5a-GFP. As controls, we transfected SRSF3-GFP as marker for the nucleoplasm (40) and performed immunofluorescence for G3BP1 as a cytoplasmic marker. A plasmid-expressing GFP alone was used as an ubiquitously present protein (41).

Arid5a clearly localizes to the nucleus, and no signal is detectable in the cytoplasm (Fig. 8). However, Arid5a shows a markedly distinct localization pattern compared to the splicing regulator SRSF3, which is found in nuclear speckles and the nucleoplasm. Arid5a perfectly colocalizes with some of the bright heterochromatin dots, indicating a close proximity to silenced chromatin. Together with our iCLIP2 data, this suggests that Arid5a might use its dual nucleic acid–binding capability to interact with DNA and pre-mRNA simultaneously, for example, to detect transcribed loci and then modulate transcriptional repression as it was described earlier (17, 18) and very recently for TF with an RBP activity (42).

RNA-binding is a more widespread capacity of ARID domains

Prior to this study, RNA-binding had only been described for Arid5a but for none of the other 14 human Arids. Driven by the observations for Arid5a above, we wondered if RNA-binding was a more general feature within this protein family. To test this, we used the closely related Arid5b as well as Arid1a and Jarid1a to perform analogous ^{1}H-^{15}N-HSQC measurements with and without the 19mer_dsRNA (Fig. S15, *A*–*E*). Surprisingly, all three Arids showed obvious line broadening with Jarid1a being most affected and very similar to $ARID_{37\text{-}183}$. Arid5b and Arid1a showed less but still mentionable line broadening. This indicates that also other Arids are able to bind RNA *in vitro* and hints at so-far unexplored functions of these proteins. To corroborate protein-observed HSQC spectra, we recorded imino proton spectra of the 19mer_dsRNA with and without the Arids in order to examine the influence of protein binding on the RNA chemical shifts (Fig. S15*F*). We did not observe significant CSPs for the imino peaks, but minor line broadening indicates that all Arid proteins interact with the RNA backbone, suggesting little specificity for the herein provided RNA motif. Nonetheless, these results reveal that some—if not all—Arid proteins are generally able to interact with RNA through their respective ARID domains, and RNA-binding competence is thus not a unique observation for Arid5a. This opens up interesting questions for further detailed studies in the future into whether and how RNA-binding is of functional relevance for them (as suggested for Arid5a).

Discussion

A recent study suggests more than 100 TFs are actively involved in splicing through their DRBP function (43). The ability to interact with both DNA and RNA is either conferred by a combination of specialized domains, for example, in Sox2 (10) or SAFB2 (44) or by the dual exploitation of one domain (12, 45). Recently, the role of IDRs for DNA- and RNA-recognition, often *via* the same sequences (46), has come into focus, but specificity parameters like in Arid5a remain elusive based on the lack of structure-derivable knowledge.

Arid proteins are categorized as exclusive DNA-binders with only one exception: Arid5a is capable of binding RNA, with specific target mRNAs and a responsible folded motif presented in earlier work (2, 13). However, not only the structural basis of this unique observation has remained unresolved, but also a clear understanding of the precise target nucleic acid preferences of Arid5a, all of which are expected to involve regions beyond the core ARID domain. In support, prior data on Arid5a RNA-binding had been achieved with the full-length protein (21, 22, 28), while RNA-binding is abolished in the absence of the ARID domain (13). The latter, as well as a

Figure 7. Binding preferences of Arid5a in endogenous RNAs determined by iCLIP2. *A*, binding site distribution of Arid5a in different gene biotypes. *B*, Arid5a-binding sites in transcript regions of protein-coding genes. *C*, heatmap showing clusters of 3-mers starting/ending with U around Arid5a-binding sites in a window of ±50 nt. *D*, heatmap showing clusters of all other 3-mers around Arid5a-binding sites in a window of ±50 nt. RBNS-derived 3-mers are underlined. *E*, frequency of UUU per position in a window of ±50 nt around the binding sites. *F*, frequency of the 3-mers AGG, CAG, GCC, and AGC from the RBNS consensus motif per position in a window of ±50 nt around the binding sites. *G*, functional enrichment analysis (Gene Ontology Biological Process) using for transcripts with Arid5a-binding sites. CDS, coding sequence; UTR, untranslated region; lincRNA, long intergenic non-coding RNA; snRNA, small nuclear RNA; snoRNA, small nucleolar RNA.

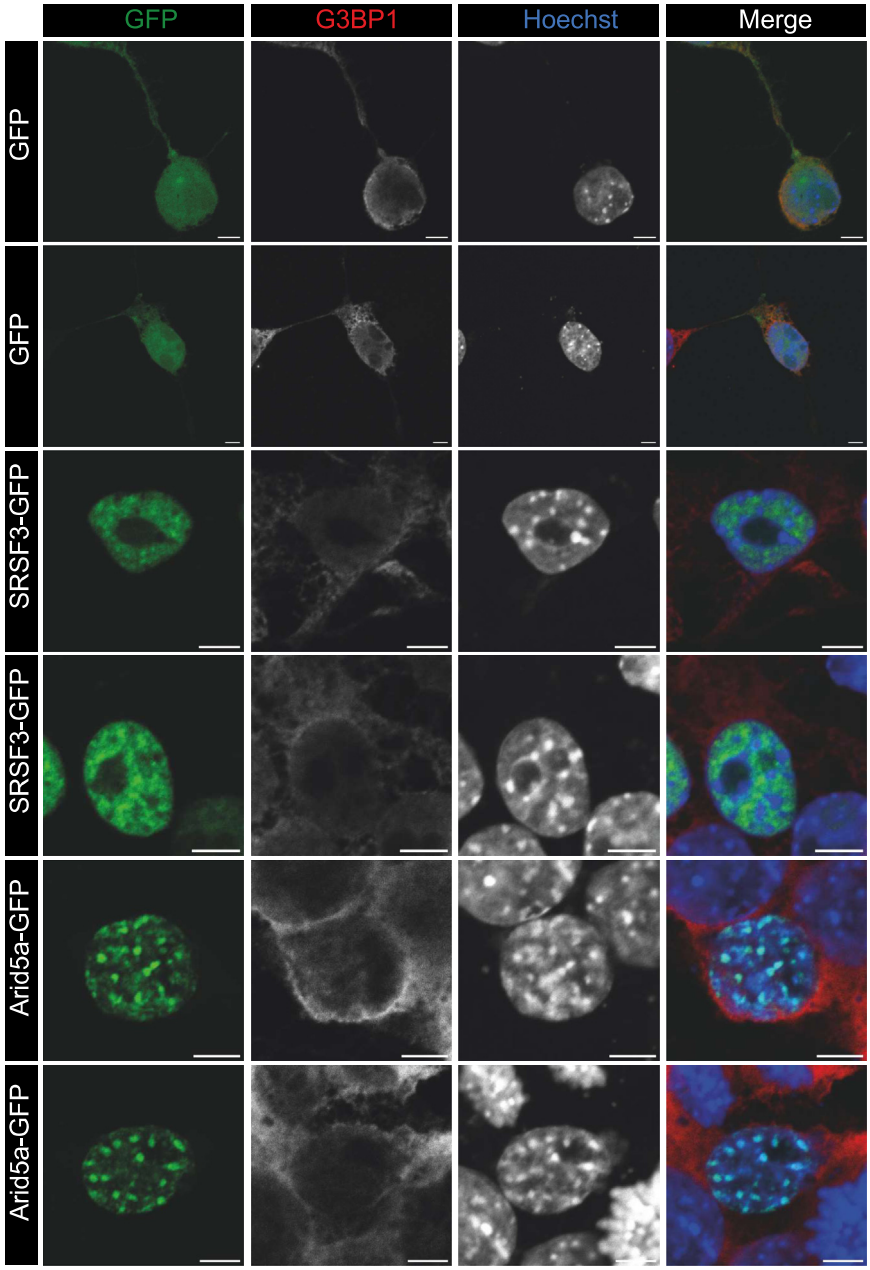


Figure 8. Subcellular localization and RNA-binding of Arid5a. Representative confocal micrograph showing that Arid5a-GFP localizes to the nucleus of murine P19 cells and colocalizes with bright heterochromatin dots. GFP was used to stain the entire cell and SRSF3-GFP as marker for nuclear speckles. Staining for G3BP1 in all cells labels the cytoplasm, and Hoechst stains chromatin. All zoom-ins are twofold. Scale bars represent 5 μm.

study involving a mutant within the core ARID (21), claim that the domain is sufficient for RNA-binding but ignore contributions from sequence elements directly adjacent. Altogether, an atom-resolved proof of the Arid5a ARID domain interacting with DNA and RNA in an isolated, *in vitro* setup had been missing.

We here provide a detailed interrogation of the Arid5a ARID-preferred DNA target DNA motif, focusing on the core domain, but taking into account contributions of N- and C-terminally extending IDRs. With a core "AATA/TATT" sequence, we find that the core ARID prefers a similar DNA target motif as its related family partner Arid5b (15, 16). This was unexpected considering the core ARIDs of both proteins share a sequence identity of only 70.2%, and the extended domains an even lower 58.3%, respectively. Interestingly, the regions involved in DNA-binding (L1 and H4-L2-H5) share a sequence similarity of 97.6% and identity of 85.4%, explaining their preference for identical DNA motifs (Fig. 1*B*). This is further supported by the finding that amino acids analogous to L2-residue T125 in Arid5a are determinants of DNA preference (32). T125 is both conserved in Arid5a between species and between Arid5a and 5b. Finally, early work had already suggested Arid5a to interact with multiple AT-rich sites, but not with a precise motif (17). In contrast, other members of the Arid family do not necessarily prefer AT-rich sequences, as, for example, reported for Arid1a (31, 35) and JARID1a (32)

with a serine and lysine, respectively, at this position. Our data (Fig. 2) confirm that Arid1a and JARID1a can bind AT- and GC-rich DNA equally strong. Similarly, we do not confirm Jarid1a to exclusively bind GC-rich DNA, thus contradicting the previous suggestions (32).

The co-existence of Arid5a and 5b in higher eukaryotes remains enigmatic, seeing their shared DNA target motif preference of the core ARID domain. Notably, literature does not list an overlap of genes regulated in transcription. Our findings suggest that a fine-tuning of DNA targets may take place through modulation by the non-identical IDRs. This is supported by earlier findings, in which Arid5b was shown to interact with DNA *via* its C-terminal extension (36).

Our data show that the positively charged C-terminal IDR also supports the affinity of Arid5a to DNA. Notably, the NMR data reveal a larger relative contribution to binding of GC DNA. This suggests this region provides a general support in DNA engagement, ultimately allowing the core ARID domain to selectively encounter AT motifs (Fig. 9*A*). While here, we suggest opposing charges to drive encounter complex formation, a recent study found negatively charged IDRs to accelerate specific motif search (47), likely preventing too tight interactions. We, however, did not find a similar contribution from the negatively charged N-terminal extension. Certainly, nature has established multiple modes of fine-tuning DNA-recognition through IDRs (48–50), including roles for hydrophobic sequences as recently shown by Jonas *et al.* (51).

The strong similarity in DNA-binding between Arid5a and Arid5b raises the question why to date only Arid5a was found to bind RNA. A sequence conservation within the extended ARID domains of the two proteins below 60% supports the hypothesis that the IDRs play a central role in (distinctive) RNA-binding competence. For Arid5a, we here unambiguously provide an atom-resolved proof for its proposed dual nucleic acid–binding competence (Fig. 9*A*), while no work had shown RNA-binding by robust *in vitro* experiments before. Unexpectedly, we found only weak binding of the utilized Arid5a constructs to the previously described *Ox40* ADE motif (22–24), while visibly enhanced by the IDRs. Our RBNS approach (unprecedented for an ARID domain) suggested short ssRNAs, superior to the ADE. In support, these motifs were partially found *in vivo*, demonstrated by the first iCLIP2 experiment with an Arid protein.

In general, binding to sequence- and size-equivalents of DNA (Fig. S17) revealed the subordinated affinity of the ARID domain to RNA. In fact, we find that RNA binding shows a CSP pattern reminiscent of nonspecific DNA-binding by NMR (Fig. S18). We do not rule out that we missed a complex-folded

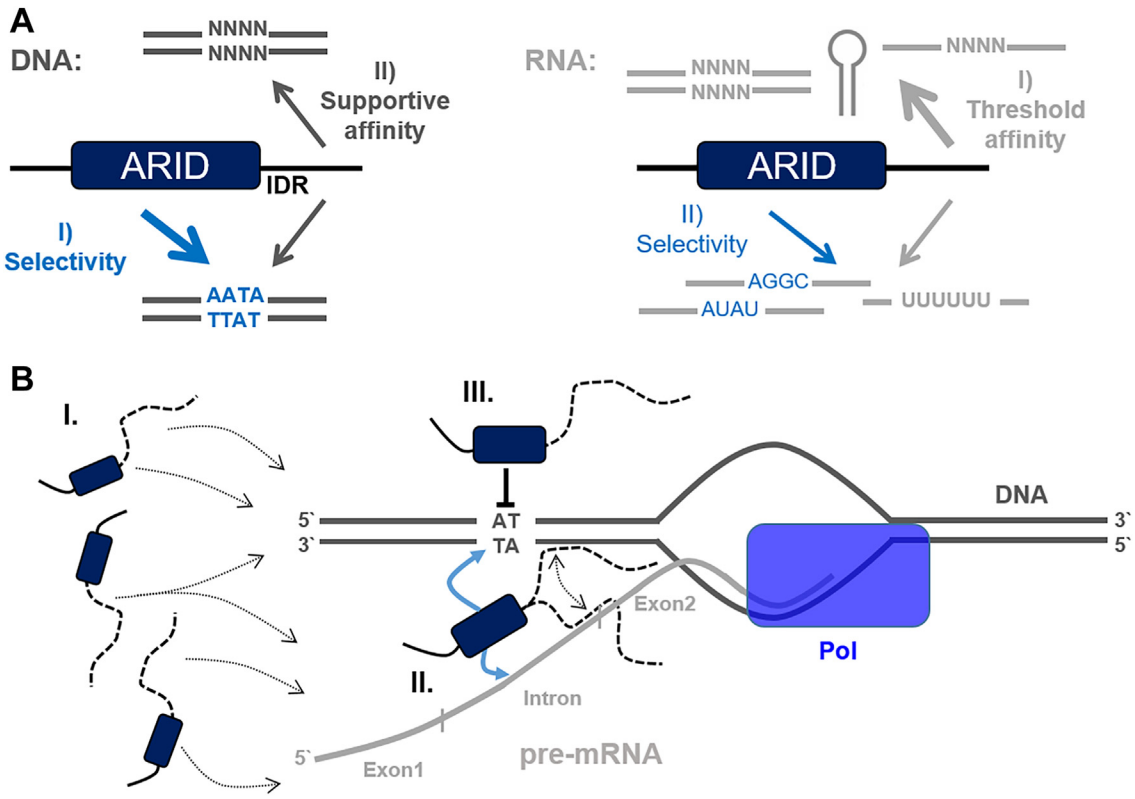


Figure 9. Summary of Arid5a interacting with nucleic acids. *A*, overview of possible and preferred interactions of the extended ARID domain, $ARID_{37\text{-}183}$, with DNAs (*left*) and RNAs (*right*) with an apparent hierarchy of affinity and selectivity. *B*, hypothetical model of transcription modulation by Arid5a integrating the *in vitro* and *in vivo* findings of Arid5a's specific and nonspecific nucleic acid interactions mediated by the core domain (*dark blue rectangle*) and the IDR extensions (*broken lines*): Recruitment of Arid5a to DNA/RNA ('scanning') will primarily locate the protein to AT-rich DNA promoter/enhancer regions (I). Increase of local Arid5a concentration through recognition of intron-exon boundaries in nascent transcripts closely located to transcribing DNA (II). Arid5a binding could allow productive transcription simply by relieving the block of DNA promoter/enhancer regions (III). Alternatively, Arid5a recruitment to pre-mRNA can lead to recognition of the gene's promoter region and its silencing.

RNA motif preferentially bound by the ARID domain, similar to the unique binding of ADE and CDE elements by the Roquin ROQ domain (23, 52).

The nuclear localization of Arid5a is supported by early characterization of the protein in different tissue types (17). More recent work identified Arid5a as specific RBP in stimulated immune cells, including export to the cytosol (13, 21, 22, 28, 53). Our data reveal a full nuclear localization, while we did not perform iCLIP2 under differential conditions and do not question a possible engagement with specific transcripts outside the nucleus. Still, we doubt Arid5a is a broadly acting RBP, neither in the nucleus nor cytoplasm, as it crosslinked much less to RNA than, for example, the splicing factor SRSF5 with ∼40 times more binding sites in a similar approach (54).

Apart from the above, the general capability of interacting with RNA had not yet been tested for other Arid proteins to our best of knowledge. When we compare findings for Arid5a to Arid1a, 5b, and Jarid1a, our data hint at a more common RNA-binding capability of ARID domains than expected. This suggests unknown functions for Arids with respect to gene regulation, for example, at the interface of transcriptional and posttranscriptional levels as very recently shown for Arid1a (55). Considering a difference in affinity between DNA and RNA, as found here for the Arid5a ARID domain, we can speculate whether RNA-binding functions require high protein or target RNA concentrations. In such a scenario, for example, Arid5a will automatically expand from DNA-binding (transcription regulation) to RNA (transcript)-binding as a consequence of its own abundance. For example, Arid5a may first act as transcriptional repressor on the chromatin level, while it then stabilizes or blocks transcripts from translation at a later stage, including its abundance-based co-export from the nucleus, thus fulfilling a regulatory role on multiple levels.

In fact, TFs possibly involve simultaneous RNA-binding as a feedback mechanism in transcription or for recruitment to transcriptional start sites, for example, *via* (l)ncRNAs. Similar to the emerging role of circRNAs for RBPs (56), RNAs may also act as sponges for excessive DBPs (57) *via* IDR interactions. DNA- and RNA-binding is a strong indicator for subcompartmental clustering of transcriptional processes, for example, for co-transcriptional splicing (43) or miRNA processing. The latter was suggested for SAFB2 (58, 59) as a *bona fide* example of a DRBP (44). Our iCLIP2 and microscopy data now suggest a similar role for Arid5a, which could function in a mechanism of RNA-induced transcriptional silencing or activation in line with differential regulation of transcription in Arid5a k.o. conditions (60). Both scenarios will involve the core ARID domain binding to dsDNA and to particular ssRNAs. The observed ramping effect in our iCLIP2 data suggests that Arid5a recognizes steep changes in nucleotide composition, that is, from longer U-rich stretches to purine-rich sequences (in accordance with RBNS), using its core ARID domain and the flanking IDRs. Such changes in nucleotide composition occur at intron-exon junctions in pre-mRNAs (Fig. 9*B*). We speculate that Arid5a normally binds to DNA but hops on nascent RNAs in the vicinity when such boundaries emerge and could thereby discriminate normal pre-mRNAs from spurious transcripts. The fact that we detect bound transcripts by iCLIP2 suggests that Arid5a binding rather prevents silencing of the locus, perhaps through loss of DNA-binding, but this requires further investigation.

The support by its adjacent IDRs additionally allows for protein-regulatory features steerable *via* posttranslational modifications (PTM). Likewise, IDRs are by default susceptible to proteolysis, an excellent tool to disrupt functional protein moieties (61, 62). Similar to the described PTMs in more distal parts of Arid5a (63), PTMs in the extended ARID domain may be relevant with respect to DNA *versus* RNA preference and general affinity.

We here focused on solution NMR spectroscopy as a valuable tool to *en-detail* correlate chemical shift information with binding modes. CSP patterns, that is, trajectories, magnitudes, and exchange regimes, are unambiguous indicators of a protein domain's preference for nucleic acid as recently shown in similar studies by us (44, 64) and others (65, 66). As such, CSPs can be used to compare DNA and RNA-binding by Arid5a, and consequently, the approach is transferable to other nucleic acid–binding domains of interest. The straightforward NMR-centered biochemical setting will on the longer run also allow to unambiguously read-out selective inhibition of one or both DNA and RNA functions as intended for Arid5a earlier (67).

Experimental procedures

Arid *protein construct design and mutagenesis*

Human Arid5a ARID constructs used in this study were designed and cloned as described previously (68). In brief, we used different domain boundaries, comprising the minimal ARID core ($ARID_{49\text{-}152}$) alone or extended either N- ($ARID_{37\text{-}152}$) or C- ($ARID_{49\text{-}183}$) terminally or both ($ARID_{37\text{-}183}$), with the numbers representing the natural sequence in the full-length context (Fig. 1*A*). ARID-coding DNA sequences for human Arid1a (residues 999–1132), Arid5b (residues 300–434), and Jarid1a (residues 66–198) were designed to comprise the minimal core ARID plus 18 N- and 20 C-terminal amino acids. They were obtained from Eurofins Genomics, optimized for *Escherichia coli* codon usage, and sub-cloned into the pET24d-derived vector pET-Trx1a (Gunter Stier, EMBL/BZH Heidelberg) (69, 70) by *NcoI/XhoI* restriction and subsequent ligation. Minimal core ARID domains were generated using the respective oligonucleotides listed in Table S1.

Arid5a ARID point mutations, in either the minimal $ARID_{49\text{-}152}$ or extended $ARID_{37\text{-}183}$ context, were introduced by site-directed mutagenesis (Tables S1 and S2). Constructs with multiple nonadjacent mutations were cloned in subsequent steps. A gene encoding for murine full-length Arid5a (fl-Arid5a) (Eurofins Genomics) was cloned into the vector pEGFP-N1 (CLONTECH) (Tables S1 and S2) to obtain Arid5a-GFP, for transfection and imaging in human and murine cell lines. Cloning was performed *via* Gibson assembly (71). Briefly, the PCR-linearized pEGFP-N1 and fl-Arid5a with homologous 5′- and 3′-ends were mixed in the reaction, incubated for 60 min at 50 °C, and transformed into *E.coli* Dh5α.

To create an Arid5a production vector for a recombinant ARID domain with Strep-tag for RBNS experiments, we amplified the *Arid5a* gene from pET-Trx1a_ARID$_{49\text{-}152}$ and cloned it into pET_TRX_Bsa_StrepTag-N *via* Golden Gate Assembly (72) using *Bsa*I restriction sites (Table S1). The resulting fusion protein then contains a His$_6$-Tag followed by a thioredoxin-tag (TRX), a TEV cleavage site, a Twin-Strep-tag (IBA Lifesciences, Göttingen), and ARID$_{49\text{-}152}$. The amino acid sequences are listed in Table S6.

Protein expression and purification

Proteins were expressed and purified as described recently (68), with an additional cation-exchange chromatography step. An ENrich S 5 × 50 column (Bio-Rad) was equilibrated with low-salt buffer (20 mM Bis-Tris, 40 mM NaCl, 2 mM DTT, 0.02% NaN_3, pH 7.2) and subsequently loaded with size-exclusion chromatography–purified protein, buffer-exchanged to low-salt buffer in Amicon ultra centrifugal filters (MWCO: 3 kDa), and concentrated to 5 ml. Pure protein was eluted with a gradient of 0 to 50% high-salt buffer (20 mM Bis–Tris, 2 M NaCl, 2 mM DTT, 0.02% NaN_3, pH 7.2) and a flow rate of 1.25 ml/min. Pure protein (determined by SDS-PAGE) was pooled and re-buffered to the final standard Arid5a buffer (20 mM Bis–Tris, 150 mM NaCl, 2 mM TCEP, 0.02% NaN_3, pH 6.5) in Amicon ultra centrifugal filters (MWCO: 3 kDa) and subsequently used for NMR, EMSA, and RBNS experiments.

DNA ligand constructs

All DNA oligonucleotides used in this study were obtained from Sigma-Aldrich. dsDNA was obtained through annealing of complementary strands (5 min at 98 °C followed by cooling down to room temperature). An overview of herein used DNAs is given in Tables S3 and S5.

RNA in vitro *transcription*

Unlabeled RNAs from 15 nt in length and longer were produced by in-house optimized *in vitro* transcription (IVT) and purified either from a linearized plasmid or from annealed oligonucleotides (Table S4) as described in (64). Briefly, plasmid DNA was linearized with *HindIII* prior to IVT by in-house–expressed T7 RNA polymerase. Alternatively, complementary oligonucleotides (Sigma-Aldrich) were annealed and used as templates for IVT. RNAs from preparative-scale (10–20 ml) transcription reactions (4 h at 37 °C) were precipitated with 1.5 volumes 2-propanol overnight at −20 °C. RNAs were separated on 12 to 18% denaturing polyacrylamide gels and visualized by UV shadowing. The excised RNA-fragments of expected length were eluted into 0.3 M NaOAc overnight and subsequently washed, concentrated, and buffer-exchanged to the experimental buffer.

RNAs below 15 nt in length (Table S5) were obtained from Dharmacon Horizon in 1-μmol-scale quantities, deprotected, and desalted. Each RNA was dissolved in the respective volume of ddH_2O to a final concentration of 3 mM.

In vitro *transcription of the RBNS input pool*

As template, a T7 promoter-containing oligonucleotide was annealed to an equimolar quantity of RBNS T7 template oligonucleotide (a random 20-mer flanked by partial Illumina primers). 500 fmol template were transcribed overnight at 37 °C with 200 mM Tris–HCl pH 8.0, 20 mM magnesium acetate, 8% (v/v) DMSO, 20 mM DTT, 20 mM spermidine, 4 mM nucleoside triphosphates (each), and self-made T7 RNA polymerase. The RBNS pool was purified by PAGE (polyacrylamide gel electrophoresis). Oligonucleotide sequences are given in Table S7.

NMR spectroscopy

NMR experiments were performed at the Frankfurt BMRZ using Bruker Avance III/Avance Neo spectrometers of 600, 700, and 900 MHz proton Larmor frequency, equipped with cryogenic probes, and using Z-axis pulsed field gradients. All measurements containing protein were performed at 298 K in standard Arid5a buffer containing 20 mM Bis–Tris pH 6.5, 150 mM NaCl, 2 mM TCEP, 0,02% NaN_3 supplemented with 5% (v/v) D_2O. Topspin versions 3 and 4 were used for data acquisition and processing. Graphical plots of spectra were created using the program NMRFAM-Sparky (73) version 1.470.

NMR backbone resonance assignments of WT ARID constructs were taken from BMRB entries 51,811 and 51,812 (68). Amide assignments of mutant ARID versions were accomplished by directly transferring the majority of assignments for peaks matching both spectra. Assignments for shifted peaks were transferred to the closest neighbor and/or with most obvious fit, which led to an unambiguous assignment completeness of 92 to 98% in the mutant ARID versions. All assignment transfers for individual apo and DNA-bound mutants are summarized in the Source Data file compared to 135 total amide assignments for the ARID$_{37\text{-}183}$ apo spectra and 126 assignments for ARID$_{37\text{-}183}$ with 2x 13merAT, as well as a total of 101 resonances for the apo and DNA-bound ARID$_{49\text{-}152}$ WT. Note that all significantly perturbed residues in DNA-binding were successfully re-assigned for comparison and later use in the box plot analysis.

NMR titrations were performed by preparing two initial samples: (i) a protein apo sample and (ii) a sample comprising protein in the presence of the maximum DNA/RNA concentration. All intermediate titration points were mixed from those samples subsequently (from high to low) to avoid side effects of protein dilution. For each sample, we monitored protein peaks by recording ^{15}N-(TROSY)-HSQCs and DNA/RNA imino peaks by acquisition of 1D imino proton spectra. For HSQC-spectra, we typically recorded 128 and 2048 points in the indirect ^{15}N and ^{1}H direct dimensions, respectively, with spectral widths of 32 ppm (offset at 116.5 ppm) and 16 ppm. For 70 μM samples used in titrations, we recorded 32 scans per increment, while 40 scans per increment were recorded for 50 μM samples. For the DNA 1D imino proton spectra, we recorded a second set of experiments, where the DNA concentrations were kept constant at 80 μM and the protein

concentration varied (10 μM, 20 μM, 40 μM, 80 μM). Spectra were recorded with 8192 points and 512 scans for 13merAT and 2560 points and 256 scans for 13merGC. The spectral width was set to 23.5 ppm and 21 ppm for 13merAT and 13merGC, respectively. Analysis of spectra and quantification/plotting of CSPs from titrations were performed in the CCPNMR Analysis 2.5 software (74). Significance of CSPs was defined as above average plus one SD, if not indicated differently. ^{1}H-^{15}N-CSPs were calculated in ppm according to Equation 1:

$$CSP = \sqrt{(0.15 \times \delta N)^2 + (\delta H)^2} \quad (1)$$

For the full integration of CSPs into statistics and graphical depiction, we used box plots according to the OneSampletTest (descriptive statistics) in OriginPro 2021b. Each box represents the interquartile range from the 25th to the 75th percentile. Whiskers show deviating values with a coefficient of 1.5. Values further beyond this threshold are shown in black or colored triangles, with colored triangles representing the five highest CSPs from $ARID_{37\text{-}183}$ with 13merAT or the single highest CSP of $ARID_{37\text{-}183}$/$ARID_{49\text{-}152}$ with 13merGC. Those colors are used throughout the panel for comparison.

For the assignment of imino protons in the 13merAT, we recorded a ^{1}H-^{1}H-NOESY at 278 K with a spectral width of 22 and 15 ppm and 4096 and 266 points for the direct and indirect proton dimensions, respectively. The mixing time was set to 300 ms. Based on this, we transferred the assignment to 298 K in a peak-traceable temperature series (Fig. S5).

Structures and structure models

Five structural models of Arid5a were generated *ab initio* with RoseTTAfold (75) using residues 37 to 183 from the sequence deposited in Uniprot (76) under ID Q03989 (see Fig. S6). We confirmed the secondary structural elements within the ARID domain by a comparison of the generated models with secondary chemical shift data, obtained in earlier work (68). The 13merAT dsDNA was modeled using the program Avogadro (77) from its primary sequence as B-DNA. PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.) was used to align both the Arid5a models and the 13merAT dsDNA model individually to the structure of Arid5b ARID in complex with DNA (PDB 2OEH, (16)). To visualize the extended Arid5a ARID domain binding to DNA, the aligned models were than manually arranged to each other by means of a slight positional adjustment to exclude steric clashes caused by the nonidentical sequences of DNAs and between Arid5a and 5b.

Electrophoretic mobility shift assay

To decipher the interaction of RNA/DNA and protein, we used EMSAs with radioactively labeled RNA (rEMSA) and fluorescently labeled DNA. The RNA was *in vitro* transcribed with T7-RNA polymerase and labeled with γ-^{32}P according to a protocol by Nahvi and Green (78). We used 30 pmol of RNA, which was dephosphorylated at the 5′-end with 3 μl of Quick-CIP (5000 U/μl, NEB) in a total volume of 20 μl according to manufacturer's instructions. Next, we performed a phenol/chloroform extraction and precipitated the RNA with ethanol and sodium acetate in the presence of 20 μg glycogen for 30 min at −20 °C. The precipitated RNA was pelleted for 15 min at 16,000*g* at 4 °C. The pellet was resuspendend in 10 μl ddH_2O from which 5 μl were used for the ^{32}P-labeling. Therefore, 1.5 μl γ-^{32}P-ATP (10 pmol, Hartmann Analytic), 2 μl T4-PNK buffer (NEB), 2 μl T4-PNK (10 U/μl, NEB), and 9.5 μl H_2O_{MQ} were added. The reaction was incubated for 60 min at 37 °C to allow phosphorylation followed by 10 min at 80 °C to inactivate the kinase. To finally purify the radioactively-labeled RNA, we used *NucAway Spin Columns* (Thermo Fisher Scientific) according to manufacturer's instructions. Finally, the RNA was refolded (4 min 95 °C, cooled down on ice water) and diluted to a final volume of 400 μl and stored at −20 °C.

For the fluorescently labeled DNA, complementary DNA oligonucleotides were used (Table S3), with one oligonucleotide 5′-labeled with fluorescein- (FAM) and the other one unlabeled. Complementary oligonucleotides (100 μM) in Arid5a buffer (20 mM Bis–Tris, 150 mM NaCl, 2 mM TCEP, 0.02% NaN_3, pH 6.5) were mixed in a 1:1 ratio and heated to 95 °C for 5 min before cooling down to allow for the annealing of dsDNA.

EMSA reactions were prepared in a final volume of 20 μl. Therefore, we mixed 0.6 μg of yeast tRNA (Roche), 10 mM $MgCl_2$, Arid5a buffer (20 mM Bis–Tris, 150 mM NaCl, 2 mM TCEP, 0.02% NaN_3, pH 6.5), and respective amounts of protein. Finally, 2 μl labeled RNA or DNA were added and the reaction (final concentration of fluorescent ligand was 10 nM and of ^{32}P-ligand ≤1 nM) incubated for 10 min at room temperature (22–24 °C). Immediately before loading 10 μl onto a 6-% polyacrylamide gel, 3 μl of loading buffer were added. Gel-electrophoresis was run for either 40 min at 80 V for DNA or 80 min at 80 V for RNA. The gels were imaged with a Typhoon Imager (GE Healthcare) either in the glass plates (for DNA) with a laser at 488 nm excitation and an emission filter at 520 nm or dried and indirectly imaged by phosphor imaging (for RNA).

Quantification of EMSAs was carried out as follows: The free band intensity was quantified in ImageQuantTL 8.1 by measuring the pixel intensity in a fixed window for each given protein concentration (see Fig. S4*F* as an example). Afterward, intensities were normalized to the lane with 0 μM protein, which was automatically set to 1. These values were then subtracted from 1 (1-free DNA) and again normalized, so that the highest values would reach 1 (only done for EMSAs with visible complex formation and 13merGC as a control). The twofold normalized data was plotted as a function of the respective protein concentrations used. The single data points were fitted by a nonlinear fit (Hill-fit) in OriginPro 2021b according to Equation 2:

$$y = V_{max} \frac{x^n}{k^n + x^n} \quad (2)$$

In the equation, "V_{max}" stands for the maximum possible bound fraction represented by the upper asymptote, "k" is the protein concentration at the transition point, and "n" is the Hill coefficient.

RNA bind-n-seq

An RBNS assay was performed with the Twin-Strep-tagged $ARID_{49-152}$ domain and a randomized input RNA pool based on reference (38). The protein was equilibrated in binding buffer (25 mM Tris–HCl, pH 7.5, 150 mM KCl, 3 mM $MgCl_2$, 0.01% Tween, 500 μg/ml BSA, 1 mM DTT) at three different concentrations (0.25, 1, 5 μM) for 30 min at 4 °C. Next, the RNA was folded by snap-cooling and added to a final concentration of 1 μM with 40 U Ribonuclease Inhibitor (moloX, Berlin) and incubated for 1 h at room temperature. A pull-down was performed by incubating the RNA/ARID mixture with 1 μl of washed MagStrep "type3" XT beads (IBA Lifesciences) for 1 h at 4 °C. Subsequently, unbound RNA was removed by washing three times with wash buffer (25 mM Tris–HCl pH 8.0, 150 mM KCl, 60 μg/ml BSA, 0.5 mM EDTA, 0.01% Tween). Afterward, the RNA–ARID complexes were eluted twice with 25 μl of elution buffer (wash buffer containing 50 mM biotin). RNA was extracted with the Zymo RNA Clean & Concentrator-5 kit (Zymo Research) according to the manufacturer's instructions. The extracted RNA was reverse transcribed into cDNA, amplified by PCR to add Illumina adapters (Table S7) and an index for each concentration (Table S8), and subjected to deep sequencing (GENEWIZ).

Next-generation sequencing data were analyzed using the RBNS pipeline as described in (79), available at https://bitbucket.org/pfreese/rbns_pipeline/overview. The sequence context was analyzed using a self-written Python script. This searches for a given motif (in this case the enriched kmers) in each read of the sequence and generates the upstream and downstream sequence logos of the given sequence. Logos are then calculated from this, which are corrected by the composition of the bases (background) in the input pool.

Culturing and transfection of P19 cells

Murine P19 WT cells were cultivated on 10-cm culture dishes pre-coated with 0.1% gelatin (in PBS) under humidified condition at 5% CO_2 and 37 °C in DMEM GlutaMAX Medium, supplemented with 10% (v/v) heat-inactivated fetal bovine serum and 100 μg/ml penicillin-streptomycin (all Gibco, Thermo Fisher Scientific). P19 WT cells were transfected with 4 μg plasmid DNA in 10-cm plates using the jetOPTIMUS Transfection reagent (Polyplus) according to the manufacturer's instructions. The cells were harvested after 24 h of incubation.

Confocal microscopy

For GFP and immunofluorescence microscopy, P19 cells were grown on precoated 10 mm glass coverslips in 10-cm plates. The coverslips were transferred into a 24-well plate and washed with 1× PBS. After removing the PBS, cells were fixed with 4% paraformaldehyde (in PBS; Thermo Fisher Scientific) for 20 min at room temperature. Fixed cells were washed twice with 1× PBS and then permeabilized in permeabilization buffer (5% BSA, 0.1% Triton in 1× PBS) for 30 min. Mouse anti-G3BP1 antibody (Abcam, ab56574) was diluted in blocking buffer (5% BSA in 1× PBS) at 2 μg/ml final concentration and incubated for 16 h overnight at 4 °C in the dark as a cytoplasmic marker. The coverslips were washed twice with 1× PBS and incubated with the secondary antibody (donkey anti-mouse coupled to Alexa Fluor 594, Abcam; 1:500 in blocking buffer) for 60 min at room temperature in the dark. After washing the coverslips twice with 1× PBS, the DNA was stained with Hoechst 34580 (Thermo Fisher Scientific) at a final concentration of 5 μg/ml in Tris-buffered saline with 0.1% Tween-20 for 30 min at room temperature in the dark. After a final wash, the coverslips were dried and mounted on ProLong Diamond Antifade Mountant (Thermo Fisher Scientific P36961).

Images were acquired with confocal laser-scanning microscope (LSM780; Zeiss) with a Plan-Apochromat 63× 1.4 NA oil differential interference contrast objective M27 using the Zen 2012 (black edition; 8.0.5.273; ZEISS). Fluorescence signal was detected with an Argon laser (GFP – 488 nm, G3BP1/Qasar – 594 nm and Hoechst – 405 nm). Fiji was used to crop the pictures with the Image crop function and to add the scale bars (80).

iCLIP2 of full-length Arid5a-GFP

iCLIP experiments were performed using the iCLIP2 protocol (39) with minor modifications. For each replicate, two 15-cm dishes of P19 cells grown to 60% confluence were transfected with 15 μg of Arid5a-GFP plasmid DNA. After 24 h, the cells were washed with ice-cold PBS, irradiated with 250 mJ/cm^2 UV light at 254 nm (CL-1000, UVP), and harvested by scraping and centrifugation. Following lysis and partial digestion with RNase I (Thermo Fisher Scientific, AM2294), immunoprecipitation of Arid5a-GFP was performed using a goat anti-GFP antibody (MPI-CBG) coupled to Dynabeads Protein G (Thermo Fisher Scientific, 10002D). Copurified, crosslinked RNA fragments were dephosphorylated at their 3′-ends using T4-PNK (NEB, M0201S) and ligated to a pre-adenylated 3′-adapter (L3-App, Table S9). To visualize protein–RNA complexes, RNA fragments crosslinked to Arid5a-GFP were labeled at their 5′ ends using T4-PNK and γ-^{32}P-ATP (Hartmann Analytic). Samples were run on a NuPAGE 4-12% Bis-Tris Protein Gel (Thermo Fisher Scientific, NP0335BOX), transferred to a 0.45 μm nitrocellulose membrane (GE Healthcare Life Science, 10600002), and visualized using a Phosphorimager. Regions of interest were cut from the nitrocellulose membrane (95 kDa to 180 kDa), and RNA was released from the membrane using Proteinase K (Roche, 03115828001). RNA was purified using neutral phenol/chloroform/isoamylalcohol (Ambion, AM9722) followed by chloroform (Serva, 39554.02) extraction and reverse transcribed using SuperScript III (Life Technologies, 18080-044) and a short RT primer (Table S9). cDNA was cleaned up using

MyONE Silane beads (Life Technologies, 37002D) followed by ligation of a second adapter containing a bipartite (5-nt + 4-nt) unique molecular identifier (UMI) as well as a 6-nt experimental barcode (39) (Lclip2.0 adapter, Table S9). iCLIP2 libraries were pre-amplified with 6 PCR cycles using short Solexa primers P5 and P3 (Table S9) and then size-selected using the ProNex Size-Selective Purification System (Promega, NG2001) in a 1:2.95 (v/v) sample:bead ratio to eliminate products originating from short cDNAs or primer dimers. The size-selected library was amplified for 6 cycles using long Solexa primers P5 and P3 (Table S9), and primers were removed using the ProNex Size-Selective Purification System (Promega, NG2001) in a 1:2.4 (v/v) sample:bead ratio. Purified iCLIP2 libraries were sequenced on a NextSeq 500 System (Illumina) using a NextSeq 500/550 High Output Kit v2 as 92-nt single-end reads, yielding between 18 and 21 million reads.

iCLIP2 analysis

iCLIP2 data were processed as described in (81). In brief, quality control was done using FastQC (version 0.11.9) (https://www.bioinformatics. Babraham.ac.uk/projects/fastqc/). Read were de-multiplexed according to the sample barcode on positions 6-11 of the reads using Flexbar (version 3.5.0, (82)) using nondefault parameters: flexbar –adapter-seq AGATCGGAAGAGCGGTTCAG –adapter-min-overlap 1 –min-read-length 15 –length-dist –umi-tags. Flexbar was also used to trim UMI and barcode regions as well as adapter sequences from read ends requiring a minimal overlap of 1 nt of read and adapter. UMIs were added to the read names and reads shorter than 15 nt were removed from further analysis. The downstream analysis was done as described in (39). Reads were mapped with STAR (v2.7.3a) (83) with nondefault parameters: STAR –alignEndsType Extend5pOfRead1 –outFilterMismatchNoverReadLmax 0.04 –outFilterMultimapNmax 1. Genome assembly (GRCm38.p6) and annotation of GENCODE (release M25) (84) were used.

Reads directly mapped to the chromosome ends were removed, as they do not have an upstream position, and no crosslink position can be extracted using Samtools (v1.10) (85), bedtools (v2.29.2) (86). PCR duplicates were removed using UMI-tools (v1.1.2) with nondefault parameters: umi_tools dedup –method unique –random-seed=100. To extract the crosslink sites, the bam files were first converted into bed files shifting one base upstream using bedtools (v2.29.2) bamtobed and shift. Then, only the first nucleotide was kept and the positions separated by the strand information using bedtools (v2.29.2) genomecov. Processed reads from three replicates were merged prior to peak calling with PureCLIP (version 1.3.1) (87) using a minimum transition probability of 1%. Significant crosslink sites (1 nt) were filtered by their PureCLIP score, removing the lowest 1% of crosslink sites. The remaining sites were merged into 7-nt wide-binding sites using the R/Bioconductor package BindingSiteFinder (version 2.0.0), filtering for sites with at least 2 positions covered by crosslink events. Only reproducible binding sites were considered for further analyses, which had to be supported by two out of three replicates. Binding sites were overlapped with gene and transcript annotations obtained from GENCODE (release 29). Binding sites in intergenic regions were removed from further analysis. Binding sites within protein-coding genes were assigned to the transcript regions, that is, intron, coding sequence, 3′-UTR, or 5′-UTR.

Motif analysis was performed by counting all possible 3-mers in a window of ±50 nt around the center points of all Arid5a-binding sites using the R/Bioconductor package Biostrings (version 2.70.1). Heatmap visualization was done separately for 3-mers starting and/or ending on U (Fig. 7*C*) and all other 3-mers (Fig. 7*D*), including *k*-means clustering with $k = 3$.

Data availability

All sequencing data are available in the Gene Expression Omnibus (GEO) under the accession numbers GSE256029 (RNBS) and GSE254818 (iCLIP2).

Supporting information—This article contains supporting information (75, 88–90).

Acknowledgments—We acknowledge excellent technical support by Katharina Targaczewski. We thank Mirko Brüggemann for advice and supervision and Anke Busch and IMB Bioinformatics Core Facility for processing the iCLIP2 data. Support by the IMB Genomics Core Facility and the use of its NextSeq 500 (funded by the Deutsche Forschungsgemeinschaft [DFG, German Research Foundation] – 329045328) is gratefully acknowledged.

Author contributions—J. v. E., S. M. K., and A. S. writing–review and editing; J. v. E., L. O., J. E. W., K. Z., M. M.-M., S. M. K., and A. S. writing–original draft; J. v. E., L. O., E. Y., N. B., S. M. K., and A. S. visualization; J. v. E., L. O., E. Y., L. I. S., N. B., F. M., and S. M. K. investigation; J. v. E., L. O., E. Y., and K. Z. formal analysis; J. v. E., S. M. K., and A. S. conceptualization; J. E. W., K. Z., M. M.-M., and A. S. funding acquisition; A. S. project administration.

Funding and additional information—The Frankfurt BMRZ (Center for Biomolecular Resonance) is supported by the Federal State of Hesse. This work was funded by the Deutsche Forschungsgemeinschaft through grant numbers SFB902/B16 and SCHL2062/2-1 and 2-2 (to A. S.), SFB902/B13 (to M. M.-M. and K. Z.), as well as SFB902/B14 and WE 5819/3-1 (to. J. E. W.), the Clusterproject EnABLE funded by the Hessian Ministry for Science and Arts (to M. M.-M.), and by the Johanna Quandt Young Academy at Goethe (grant number 2019/AS01 to A. S.).

Conflicts of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: ADE, alternative decay element; ARID, AT-rich interacting domain; CSP, chemical shift perturbation; DBP, DNA-binding protein; DRBP, DNA- and RNA-binding protein; HSQC, heteronuclear single quantum coherence; HTH, helix-turn-helix; IDR, intrinsically disordered region; IVT, *in vitro* transcription; PTM, posttranslational modifications; RBNS,

RNA Bind-n-Seq; RBP, RNA-binding protein; TF, transcription factor; UMI, unique molecular identifier.

References

1. Patsialou, A., Wilsker, D., and Moran, E. (2005) DNA-binding properties of ARID family proteins. *Nucleic Acids Res.* **33**, 66–80
2. Wilsker, D., Patsialou, A., Dallas, P. B., and Moran, E. (2002) ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development. *Cell Growth Differ.* **13**, 95–106
3. Gregory, S. L., Kortschak, R. D., Kalionis, B., and Saint, R. (1996) Characterization of the dead ringer gene identifies a novel, highly conserved family of sequence-specific DNA-binding proteins. *Mol. Cell Biol.* **16**, 792–799
4. Herrscher, R. F., Kaplan, M. H., Lelsz, D. L., Das, C., Scheuermann, R., and Tucker, P. W. (1995) The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family. *Genes Dev.* **9**, 3067–3082
5. Korn, S. M., and Schlundt, A. (2022) Structures and nucleic acid-binding preferences of the eukaryotic ARID domain. *Biol. Chem.* **403**, 731–747
6. Littlefield, O., Korkhin, Y., and Sigler, P. B. (1999) The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13668–13673
7. Ogata, K., Kanei-Ishii, C., Sasaki, M., Hatanaka, H., Nagadoi, A., Enari, M., *et al.* (1996) The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nat. Struct. Biol.* **3**, 178–187
8. Masliah, G., Barraud, P., and Allain, F. H. (2013) RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell Mol. Life Sci.* **70**, 1875–1895
9. Hudson, W. H., and Ortlund, E. A. (2014) The structure, function and evolution of proteins that bind DNA and RNA. *Nat. Rev. Mol. Cell Biol.* **15**, 749–760
10. Hou, L., Wei, Y., Lin, Y., Wang, X., Lai, Y., Yin, M., *et al.* (2020) Concurrent binding to DNA and RNA facilitates the pluripotency reprogramming activity of Sox2. *Nucleic Acids Res.* **48**, 3869–3887
11. Norman, M., Rivers, C., Lee, Y. B., Idris, J., and Uney, J. (2016) The increasing diversity of functions attributed to the SAFB family of RNA-/DNA-binding proteins. *Biochem. J.* **473**, 4271–4288
12. Theunissen, O., Rudt, F., Guddat, U., Mentzel, H., and Pieler, T. (1992) RNA and DNA binding zinc fingers in Xenopus TFIIIA. *Cell* **71**, 679–690
13. Masuda, K., Ripley, B., Nishimura, R., Mino, T., Takeuchi, O., Shioi, G., *et al.* (2013) Arid5a controls IL-6 mRNA stability, which contributes to elevation of IL-6 level *in vivo*. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 9409–9414
14. Yang, N., and Xu, R. M. (2013) Structure and function of the BAH domain in chromatin biology. *Crit. Rev. Biochem. Mol. Biol.* **48**, 211–221
15. Whitson, R. H., Huang, T., and Itakura, K. (1999) The novel Mrf-2 DNA-binding domain recognizes a five-base core sequence through major and minor-groove contacts. *Biochem. Biophys. Res. Commun.* **258**, 326–331
16. Cai, S., Zhu, L., Zhang, Z., and Chen, Y. (2007) Determination of the three-dimensional structure of the Mrf2-DNA complex using paramagnetic spin labeling. *Biochemistry* **46**, 4943–4950
17. Huang, T. H., Oka, T., Asai, T., Okada, T., Merrills, B. W., Gertson, P. N., *et al.* (1996) Repression by a differentiation-specific factor of the human cytomegalovirus enhancer. *Nucleic Acids Res.* **24**, 1695–1701
18. Georgescu, S. P., Li, J. H., Lu, Q., Karas, R. H., Brown, M., and Mendelsohn, M. E. (2005) Modulator recognition factor 1, an AT-rich interaction domain family member, is a novel corepressor for estrogen receptor alpha. *Mol. Endocrinol.* **19**, 2491–2501
19. Amano, K., Hata, K., Muramatsu, S., Wakabayashi, M., Takigawa, Y., Ono, K., *et al.* (2011) Arid5a cooperates with Sox9 to stimulate chondrocyte-specific transcription. *Mol. Biol. Cell* **22**, 1300–1311
20. Yoshinaga, M., and Takeuchi, O. (2019) RNA binding proteins in the control of autoimmune diseases. *Immunol. Med.* **42**, 53–64
21. Masuda, K., Ripley, B., Nyati, K. K., Dubey, P. K., Zaman, M. M. U., Hanieh, H., *et al.* (2016) Arid5a regulates naive CD4+ T cell fate through selective stabilization of Stat3 mRNA. *J. Exp. Med.* **213**, 605–619
22. Hanieh, H., Masuda, K., Metwally, H., Chalise, J. P., Mohamed, M., Nyati, K. K., *et al.* (2018) Arid5a stabilizes OX40 mRNA in murine CD4(+) T cells by recognizing a stem-loop structure in its 3'UTR. *Eur. J. Immunol.* **48**, 593–604
23. Janowski, R., Heinz, G. A., Schlundt, A., Wommelsdorf, N., Brenner, S., Gruber, A. R., *et al.* (2016) Roquin recognizes a non-canonical hexaloop structure in the 3′-UTR of Ox40. *Nat. Commun.* **7**, 11032
24. Tants, J.-N., Becker, L. M., McNicoll, F., Müller-McNicoll, M., and Schlundt, A. (2022) NMR-derived secondary structure of the full-length Ox40 mRNA 3′UTR and its multivalent binding to the immunoregulatory RBP Roquin. *Nucleic Acids Res.* **50**, 4083–4099
25. Schlundt, A., Niessing, D., Heissmeyer, V., and Sattler, M. (2016) RNA recognition by Roquin in posttranscriptional gene regulation. *Wiley Inter. Rev. RNA* **7**, 455–469
26. Mino, T., Murakawa, Y., Fukao, A., Vandenbon, A., Wessels, H. H., Ori, D., *et al.* (2015) Regnase-1 and roquin regulate a common element in inflammatory mRNAs by spatiotemporally distinct mechanisms. *Cell* **161**, 1058–1073
27. Jeltsch, K. M., and Heissmeyer, V. (2016) Regulation of T cell signaling and autoimmunity by RNA-binding proteins. *Curr. Opin. Immunol.* **39**, 127–135
28. Higa, M., Oka, M., Fujihara, Y., Masuda, K., Yoneda, Y., and Kishimoto, T. (2018) Regulation of inflammatory responses by dynamic subcellular localization of RNA-binding protein Arid5a. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E1214–E1220
29. Oksuz, O., Henninger, J. E., Warneford-Thomson, R., Zheng, M. M., Erb, H., Vancura, A., *et al.* (2023) Transcription factors interact with RNA to regulate genes. *Mol. Cell* **83**, 2449–2463.e2413
30. Maulik, A., Giri, M., and Singh, M. (2019) Molecular determinants of complex formation between DNA and the AT-rich interaction domain of BAF250a. *FEBS Lett.* **593**, 2716–2729
31. Nie, Z., Xue, Y., Yang, D., Zhou, S., Deroo, B. J., Archer, T. K., and Wang, W. (2000) A specificity and targeting subunit of a human SWI/SNF family-related chromatin-remodeling complex. *Mol. Cell Biol.* **20**, 8879–8888
32. Tu, S., Teng, Y. C., Yuan, C., Wu, Y. T., Chan, M. Y., Cheng, A. N., *et al.* (2008) The ARID domain of the H3K4 demethylase RBP2 binds to a DNA CCGCCC motif. *Nat. Struct. Mol. Biol.* **15**, 419–421
33. Sandhya, S., Maulik, A., Giri, M., and Singh, M. (2018) Domain architecture of BAF250a reveals the ARID and ARM-repeat domains with implication in function and assembly of the BAF remodeling complex. *PLoS One* **13**, e0205267
34. Iwahara, J., Iwahara, M., Daughdrill, G. W., Ford, J., and Clubb, R. T. (2002) The structure of the Dead ringer-DNA complex reveals how AT-rich interaction domains (ARIDs) recognize DNA. *EMBO J.* **21**, 1197–1209
35. Kim, S., Zhang, Z., Upchurch, S., Isern, N., and Chen, Y. (2004) Structure and DNA-binding sites of the SWI1 AT-rich interaction domain (ARID) suggest determinants for sequence-specific DNA recognition. *J. Biol. Chem.* **279**, 16670–16676
36. Zhu, L., Hu, J., Lin, D., Whitson, R., Itakura, K., and Chen, Y. (2001) Dynamics of the Mrf-2 DNA-binding domain free and in complex with DNA. *Biochemistry* **40**, 9142–9150
37. Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P. A., and Burge, C. B. (2014) RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol. Cell* **54**, 887–900
38. Lambert, N. J., Robertson, A. D., and Burge, C. B. (2015) RNA bind-n-seq: measuring the binding affinity landscape of RNA-binding proteins. *Methods Enzymol.* **558**, 465–493
39. Buchbender, A., Mutter, H., Sutandy, F. X. R., Körtel, N., Hänel, H., Busch, A., *et al.* (2020) Improved library preparation with the new iCLIP2 protocol. *Methods* **178**, 33–48

40. Schwich, O. D., Blümel, N., Keller, M., Wegener, M., Setty, S. T., Brunstein, M. E., *et al.* (2021) SRSF3 and SRSF7 modulate 3′UTR length through suppression or activation of proximal polyadenylation sites and regulation of CFIm levels. *Genome Biol.* **22**, 82
41. Seibel, N. M., Eljouni, J., Nalaskowski, M. M., and Hampe, W. (2007) Nuclear localization of enhanced green fluorescent protein homomultimers. *Anal. Biochem.* **368**, 95–99
42. [preprint] Ray, M., Zaborowsky, J., Mahableshwarkar, P., Vaidyanathan, S., Shum, J., Viswanathan, R., *et al.* (2024) Dual DNA/RNA-binding factor regulates dynamics of hnRNP splicing condensates. *bioRxiv*. https://doi.org/10.1101/2024.01.11.575216
43. Wang, C., Zong, X., Wu, F., Leung, R. W. T., Hu, Y., and Qin, J. (2022) DNA- and RNA-binding proteins linked transcriptional control and alternative splicing together in a two-layer regulatory network system of chronic myeloid leukemia. *Front. Mol. Biosci.* **9**, 920492
44. Korn, S. M., Von Ehr, J., Dhamotharan, K., Tants, J. N., Abele, R., and Schlundt, A. (2023) Insight into the structural basis for dual nucleic acid-recognition by the scaffold attachment factor B2 protein. *Int. J. Mol. Sci.* **24**, 3286
45. Talwar, T., Vidhyasagar, V., Qing, J., Guo, M., Kariem, A., Lu, Y., *et al.* (2017) The DEAD-box protein DDX43 (HAGE) is a dual RNA-DNA helicase and has a K-homology domain required for full nucleic acid unwinding activity. *J. Biol. Chem.* **292**, 10429–10443
46. Miyaji, M., Kawano, S., Furuta, R., Murakami, E., Ikeda, S., Tsutsui, K. M., and Tsutsui, K. (2023) Selective DNA-binding of SP120 (rat ortholog of human hnRNP U) is mediated by arginine-glycine rich domain and modulated by RNA. *PLoS One* **18**, e0289599
47. Wang, X., Bigman, L. S., Greenblatt, H. M., Yu, B., Levy, Y., and Iwahara, J. (2023) Negatively charged, intrinsically disordered regions can accelerate target search by DNA-binding proteins. *Nucleic Acids Res.* **51**, 4701–4712
48. Mar, M., Nitsenko, K., and Heidarsson, P. O. (2023) Multifunctional intrinsically disordered regions in transcription factors. *Chemistry* **29**, e202203369
49. Brodsky, S., Jana, T., Mittelman, K., Chapal, M., Kumar, D. K., Carmi, M., and Barkai, N. (2020) Intrinsically disordered regions direct transcription factor *in vivo* binding specificity. *Mol. Cell* **79**, 459–471.e454
50. Brodsky, S., Jana, T., and Barkai, N. (2021) Order through disorder: the role of intrinsically disordered regions in transcription factor binding specificity. *Curr. Opin. Struct. Biol.* **71**, 110–115
51. Jonas, F., Carmi, M., Krupkin, B., Steinberger, J., Brodsky, S., Jana, T., and Barkai, N. (2023) The molecular grammar of protein disorder guiding genome-binding locations. *Nucleic Acids Res.* **51**, 4831–4844
52. Schlundt, A., Heinz, G. A., Janowski, R., Geerlof, A., Stehle, R., Heissmeyer, V., *et al.* (2014) Structural basis for RNA recognition in roquin-mediated post-transcriptional gene regulation. *Nat. Struct. Mol. Biol.* **21**, 671–678
53. Zaman, M. M., Masuda, K., Nyati, K. K., Dubey, P. K., Ripley, B., Wang, K., *et al.* (2016) Arid5a exacerbates IFN-gamma-mediated septic shock by stabilizing T-bet mRNA. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11543–11548
54. Arnold, B., Riegger, R. J., Okuda, E. K., Slišković, I., Keller, M., Bakisoglu, C., *et al.* (2024) hGRAD: a versatile "one-fits-all" system to acutely deplete RNA binding proteins from condensates. *J. Cell Biol.* **223**, e202304030
55. Zhang, J., Chen, F., Tang, M., Xu, W., Tian, Y., Liu, Z., *et al.* (2024) The ARID1A-METTL3-m6A axis ensures effective RNase H1-mediated resolution of R-loops and genome stability. *Cell Rep.* **43**, 113779
56. [preprint] Okholm, T. L. H., Kamstrup, A. B., Nielsen, M. M., Hollensen, A. K., Graversgaard, M. L., Sørensen, M. H., *et al.* (2023) circHIPK3 nucleates IGF2BP2 and functions as a competing endogenous RNA. *bioRxiv*. https://doi.org/10.1101/2023.09.14.557527
57. Schreiner, S., Didio, A., Hung, L. H., and Bindereif, A. (2020) Design and application of circular RNAs with protein-sponge function. *Nucleic Acids Res.* **48**, 12326–12335
58. Huo, X., Ji, L., Zhang, Y., Lv, P., Cao, X., Wang, Q., *et al.* (2020) The nuclear matrix protein SAFB cooperates with major satellite RNAs to stabilize heterochromatin architecture partially through phase separation. *Mol. Cell* **77**, 368–383.e367
59. Hutter, K., Lohmüller, M., Jukic, A., Eichin, F., Avci, S., Labi, V., *et al.* (2020) SAFB2 enables the processing of suboptimal stem-loop structures in clustered primary miRNA transcripts. *Mol. Cell* **78**, 876–889.e876
60. Nyati, K. K., Hashimoto, S., Singh, S. K., Tekguc, M., Metwally, H., Liu, Y. C., *et al.* (2021) The novel long noncoding RNA AU021063, induced by IL-6/Arid5a signaling, exacerbates breast cancer invasion and metastasis by stabilizing Trib3 and activating the Mek/Erk pathway. *Cancer Lett.* **520**, 295–306
61. Jeltsch, K. M., Hu, D., Brenner, S., Zöller, J., Heinz, G. A., Nagel, D., *et al.* (2014) Cleavage of roquin and regnase-1 by the paracaspase MALT1 releases their cooperatively repressed targets to promote T(H)17 differentiation. *Nat. Immunol.* **15**, 1079–1089
62. Schmidt, H., Raj, T., O'Neill, T. J., Muschaweckh, A., Giesert, F., Negraschus, A., *et al.* (2023) Unrestrained cleavage of Roquin-1 by MALT1 induces spontaneous T cell activation and the development of autoimmunity. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2309205120
63. Nyati, K. K., Agarwal, R. G., Sharma, P., and Kishimoto, T. (2019) Arid5a regulation and the roles of Arid5a in the inflammatory response and disease. *Front. Immunol.* **10**, 2790
64. Korn, S. M., Dhamotharan, K., Jeffries, C. M., and Schlundt, A. (2023) The preference signature of the SARS-CoV-2 Nucleocapsid NTD for its 5′-genomic RNA elements. *Nat. Commun.* **14**, 3331
65. Sundell, G. N., Vogeli, B., Ivarsson, Y., and Chi, C. N. (2018) The sign of nuclear magnetic resonance chemical shift difference as a determinant of the origin of binding selectivity: elucidation of the position dependence of phosphorylation in ligands binding to scribble PDZ1. *Biochemistry* **57**, 66–71
66. Chaves-Arquero, B., Collins, K. M., Abis, G., Kelly, G., Christodoulou, E., Taylor, I. A., and Ramos, A. (2023) Affinity-enhanced RNA-binding domains as tools to understand RNA recognition. *Cell Rep. Methods* **3**, 100508
67. Masuda, K., and Kishimoto, T. (2018) A potential therapeutic target RNA-binding protein, Arid5a for the treatment of inflammatory disease associated with aberrant cytokine expression. *Curr. Pharm. Des.* **24**, 1766–1771
68. von Ehr, J., Korn, S. M., Weiss, L., and Schlundt, A. (2023) ^{1}H, ^{13}C, ^{15}N backbone chemical shift assignments of the extended ARID domain in human AT-rich interactive domain protein 5a (Arid5a). *Biomol. NMR Assign.* **17**, 121–127
69. Bogomolovas, J., Simon, B., Sattler, M., and Stier, G. (2009) Screening of fusion partners for high yield expression and purification of bioactive viscotoxins. *Protein Expr. Purif.* **64**, 16–23
70. Peti, W., and Page, R. (2007) Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost. *Protein Expr. Purif.* **51**, 1–10
71. Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345
72. Engler, C., Kandzia, R., and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, e3647
73. Lee, W., Tonelli, M., and Markley, J. L. (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327
74. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., *et al.* (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696
75. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., *et al.* (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876
76. UniProt, C. (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489
77. Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4**, 17
78. Nahvi, A., and Green, R. (2013) Structural analysis of RNA backbone using in-line probing. *Methods Enzymol.* **530**, 381–397

79. Dominguez, D., Freese, P., Alexis, M. S., Su, A., Hochman, M., Palden, T., *et al.* (2018) Sequence, structure, and context preferences of human RNA binding proteins. *Mol. Cell* **70**, 854–867.e859
80. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682
81. Busch, A., Bruggemann, M., Ebersberger, S., and Zarnack, K. (2020) iCLIP data analysis: a complete pipeline from sequencing reads to RBP binding sites. *Methods* **178**, 49–62
82. Roehr, J. T., Dieterich, C., and Reinert, K. (2017) Flexbar 3.0 - SIMD and multicore parallelization. *Bioinformatics* **33**, 2941–2942
83. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., *et al.* (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
84. Frankish, A., Diekhans, M., Ferreira, A. M., Johnson, R., Jungreis, I., Loveland, J., *et al.* (2019) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773
85. Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., *et al.* (2021) Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008
86. Quinlan, A. R., and Hall, I. M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842
87. Krakau, S., Richard, H., and Marsico, A. (2017) PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. *Genome Biol.* **18**, 240
88. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539
89. Lorenz, R., Bernhart, S. H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, PF, *et al.* (2011) ViennaRNA package 2.0. *Algorithms Mol. Biol.* **6**, 26
90. Hofacker, I. L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431