Review

Sophie Marianne Korn* and Andreas Schlundt*

Structures and nucleic acid-binding preferences of the eukaryotic ARID domain

<https://doi.org/10.1515/hsz-2021-0404>

Received November 1, 2021; accepted January 17, 2022;

published online February 7, 2022

Abstract: The DNA-binding AT-rich interactive domain (ARID) exists in a wide range of proteins throughout eukaryotic kingdoms. ARID domain-containing proteins are involved in manifold biological processes, such as transcriptional regulation, cell cycle control and chromatin remodeling. Their individual domain composition allows for a sub-classification within higher mammals. ARID is categorized as binder of double-stranded AT-rich DNA, while recent work has suggested ARIDs as capable of binding other DNA motifs and also recognizing RNA. Despite a broad variability on the primary sequence level, ARIDs show a highly conserved fold, which consists of six α -helices and two loop regions. Interestingly, this minimal core domain is often found extended by helices at the N- and/or C-terminus with potential roles in target specificity and, subsequently function. While high-resolution structural information from various types of ARIDs has accumulated over two decades now, there is limited access to ARID-DNA complex structures. We thus find ourselves left at the beginning of understanding ARID domain target specificities and the role of accompanying domains. Here, we systematically summarize ARID domain conservation and compare the various types with a focus on their structural differences and DNA-binding preferences, including the context of multiple other motifs within ARID domain containing proteins.

Keywords: ARID containing proteins; AT-rich interacting domain; chromatin remodeling; DNA-binding; DNA-protein complex; structural biology.

Introduction

DNA-binding proteins (DBPs) control a plethora of biological processes, such as transcriptional regulation and DNA replication (Luscombe et al. 2000). Typically, DBPs comprise one or more DNA-binding domains (DBDs) that recognize either single-stranded (ss) or double-stranded (ds) DNA. The structured DBDs are often interspersed by intrinsically disordered regions (IDRs) that are involved in protein-protein interactions (PPIs) (Liu et al. 2006; Minezaki et al. 2006) or modulate the transition from non-specific to specific DNA-binding (Douceff and Clore 2008). Target DNA specificities of DBPs cover a broad spectrum. Non-specific DNA polymerases or histones do not discriminate DNAs on a sequence level. In turn, highly specific transcription factors (TFs) precisely recognize short, conserved motifs within promoters of their target genes, where specificity is mainly driven by a combination of DNA sequence and shape, i.e. groove width and bending (Slattery et al. 2014).

One of the most common structured DNA-binding elements is the evolutionary highly conserved helix-turn-helix (HTH) motif, present in a large number of repressor and activator proteins [e.g. summarized in (Aravind et al. 2005)]. With its minimalistic arrangement of only two α -helices connected by a turn, the HTH is a widespread component of DBPs (Gribskov and Burgess 1986; Littlefield et al. 1999; Ogata et al. 1996). While similar in the general composition, vastly differing structural contexts allow the HTH to act as a versatile element (Aravind et al. 2005).

Among the HTH-comprising DBDs, the AT-rich interactive domain (ARID) is one of the most ancient representatives, and found across eukaryotic kingdoms. ARIDs had their first mention in the mid-1990s, when – in search for AT-rich DNA-binders – two groups discovered the murine B cell-specific trans-activator of IgH transcription (*Bright*) and Dead ringer protein (*Dri*) of *Drosophila melanogaster*, respectively (Gregory et al. 1996; Herscher et al. 1995). Identification of the two responsible DBDs revealed a highly conserved domain, introduced as ‘ARID’, reflecting its suggested binding preference. Meanwhile, a large

*Corresponding authors: Sophie Marianne Korn and Andreas Schlundt, Institute for Molecular Biosciences and Center for Biomolecular Magnetic Resonance (BMRZ), Goethe-University Frankfurt, Max-von-Laue-Str. 9, D-60438 Frankfurt, Germany, E-mail: korn@bio.uni-frankfurt.de (S.M. Korn), schlundt@bio.uni-frankfurt.de (A. Schlundt). <https://orcid.org/0000-0003-3798-3277> (S.M. Korn). <https://orcid.org/0000-0003-2254-7560> (A. Schlundt)

number of ARID containing proteins – in higher mammals often termed *Arids* – has been described, involved in biological processes like cell cycle regulation or chromatin remodeling [summarized in (Patsialou et al. 2005; Wilsker et al. 2002)]. Accordingly, *Arid* proteins predominantly

locate to the nucleus, and for some, increased cytoplasmic protein levels have been correlated with cancer (Animireddy et al. 2021) (Table 1).

Structural information of ARID domains in complex with target DNAs is rare. Hence, a detailed understanding

Table 1: Overview of the available structural information of ARID domains in the 15 human *Arid* proteins within seven sub-families and six additional eukaryotic, non-human *Arids* from *S. cerevisiae* (S.c.), *A. thaliana* (A.t.), *Mus musculus* (M.m.) and *D. melanogaster* (D.m.).

Sub-family	Protein (alternative names)	Full-length (aa)	ARID domain (aa) ^a	PDB IDs (references)	Cellular localization (references) ^f
Human <i>Arids</i>					
<i>Arid1</i>	<i>Arid1a</i> , B120, BAF250A, hOSA1 p270, hELD	2285	1017–1108	1RYU ^b (Kim et al. 2004)	N, (C localization observed) (Bateman et al. 2016; Guan et al. 2012)
	<i>Arid1b</i> , BAF250B, BRG1-binding protein, hELD/OSA1, hOsa2, p250R	2236	1053–1144	2EH9 ^c 2CXY ^c	N, (C localization associated to cancer) (Animireddy et al. 2021)
<i>Arid2</i>	<i>Arid2</i> , BAF200, Zipzap/p200	1835	13–105	–	N
<i>Arid3</i>	<i>Arid3a</i> , Bright, Dri-like protein 1, E2FBP 1	593	238–330	4LJX ^c 2KK0 ^b (Liu et al. 2010)	N/C (Schmidt et al. 2009) (Kim and Tucker 2006)
	<i>Arid3b</i> , Bright-like	561	215–307	–	Like <i>Arid3a</i>
	<i>Arid3c</i> , Bright-like	412	113–205	–	Like <i>Arid3a</i> (Tidwell et al. 2011)
<i>Arid4</i>	<i>Arid4a</i> , RBBP-1	1257	309–401	6L87 ^{b,e} (Gong et al. 2021)	N
	<i>Arid4b</i> , SAP180, BRCAA1, RBBP-1-like 1	1312	306–398	–	N
<i>Arid5</i>	<i>Arid5a</i> , MRF-1	594	55–147	–	N/C (Higa et al. 2018)
	<i>Arid5b</i> , MRF1-like protein, MRF-2	1188	318–410	1IG6 ^b 2OEH ^b (Cai et al. 2007)	N
<i>Jarid1</i>	<i>Jarid1a</i> , KDM5A, RBBP-2	1690	84–174	2JXJ ^b (Tu et al. 2008) 5CEH ^{c,e} (Vinogradova et al. 2016) 5K4L ^{c,e} (Labadie et al. 2016) 5V9P/5V9T ^{c,e} (Liang et al. 2017)	N
	<i>Jarid1b</i> , KDM5B, PLU-1, CT31, RBP2-H1	1544	97–187	–	N
	<i>Jarid1c</i> , KDM5C, SmcX, Xe169	1560	79–169	2JRZ ^b (Koehler et al. 2008)	N
	<i>Jarid1d</i> , KDM5D, H-Y, SmcY	1539	79–169	2YQE ^b	N
	<i>Jarid2</i> , JMJ	1246	621–713	–	N
Non-human <i>Arids</i>					
	<i>Swi1</i> (S. c.), ADR6, GAM3	1314	406–493	2LI6 ^b (Wang et al. 2012) 1KN5/1KKX ^b (Tu et al. 2001)	N
	<i>Arid4</i> (A. t.)	747	566–670	6LQF ^{b,e} (Tan et al. 2020)	N
	<i>Jarid1b</i> (M. m.), KDM5B, PLU-1	1544	97–187	2EQY ^b	N
	<i>Jarid2</i> (M. m.)	1234	619–711	2RQ5 ^b (Kusunoki et al. 2009)	N
	<i>Dri</i> (D. m.), Protein retained	911	293–385	1KQQ ^b (Iwahara and Clubb 1999) 1C20 ^b (Iwahara et al. 2002)	N
	<i>Kdm5</i> (D. m.), Llid, RBP2-H	1838	226–316	2LM1 ^b	N

^aAs annotated in UniProt (UniProt, 2021). ^bNMR solution structure. ^cCrystal structure. ^dCryo-EM structure. ^eStructure of macromolecular complex or tandem domains, including ARID domain. ^fDescribed cellular localization: N – nucleus; C – cytoplasm.

of how *Arid* proteins – and more specifically the ARID domains – bind DNA is still missing, also based on a very limited knowledge about the preferred target sequences for individual ARIDs. While their name implies a clear

preference for AT-rich sequences, several exceptions describe binding of either non-sequence specific dsDNA (Gong et al. 2021) or a preference for GC-rich sequences (Tu et al. 2008) (Table 2). Increasing evidence emphasizes

Table 2: Overview of suggested target DNA sequences for the 15 human and additional eukaryotic *Arids* from *S. cerevisiae* (S.c.), *A. thaliana* (A.t.), *Mus musculus* (M.m.) and *D. melanogaster* (D.m.), including affinities and the method of determination where applicable.

Protein	DNA sequences (5'–3')	K_D (μ M)	Construct ^c	References	Approach used	Structure
<i>Arid1a</i>	N.S.	–	Fragment	Dallas et al. (2000)	PCR-selection/ EMSA	–
	CCTTCCTTCC	–	ARID	Nie et al. (2000)	EMSA	–
	N.S.	–	Fragment	Patsialou et al. (2005)	GST-pulldown	–
	CGCGAATTCGCG ^a	3.4	ARID	Kim et al. (2004)	NMR	–
	CATCAATAAATTAGAAATTAA ^a	0.09	Fragment	Chandler et al. (2013)	EMSA	–
	CGCGAATTCGCG ^a	0.07	ARID	Sandhya et al. (2018)	ITC	–
	AATA or TTAT	0.35– 5.5 ^d	ARID	Maulik et al. (2019)	ITC/NMR	Model ^e
<i>Arid1b</i>	N.S.	–	Fragment	Wang et al. (2004)	GST-pulldown	–
<i>Arid2</i>	N.S.	–	Fragment	Patsialou et al. (2005)	GST-pulldown	–
<i>Arid3a</i>	ATGCAAAT PuATa/tAA ^b PuATTAA ^b	–	FL	Gregory et al. (1996), Herrscher et al. (1995), and Rhee et al. (2014)	ChIP	MD simul. using 2KK0 (<i>Arid3a</i>) versus 1C20/1KQQ (<i>Dri</i> free/bound)
<i>Arid3b</i>	TGGGATTACAG	–	FL	Bobbs et al. (2015)	ChIP-Chip	–
<i>Arid3c</i>	Like <i>Arid3a</i>	–	FL	Tidwell et al. (2011)	EMSA	–
<i>Arid4a</i>	N.S.	–	Fragment	Patsialou et al. (2005)	GST-pulldown	–
	N.S.	14– 24 ^d	PWWP-ARID	Gong et al. (2021)	EMSA/ITC	–
<i>Arid4b</i>	N.I.	–	–	–	–	–
<i>Arid5a</i>	AATAT	1	ARID	Own data, unpublished	EMSA	–
<i>Arid5b</i>	AATA(C/T)	$\geq 2.7^d$	ARID	(Whitson et al. 1999)	EMSA	–
	ACAATATAACGTCG	–	ARID	Cai et al. (2007)	NMR	20EH ^f
<i>Jarid1a</i>	N.S.	–	Fragment	Patsialou et al. (2005)	GST-pulldown	–
	CCGCC	1.6– 2.1	ARID	Tu et al. (2008)	SELEX, EMSA	–
<i>Jarid1b</i>	N.S.	–	Fragment	Patsialou et al. (2005)	GST-pulldown	–
	GCACA/C, AATTAAA	–	Fragment	Scibetta et al. (2007) and Yao et al. (2010)	SELEX, EMSA	–
<i>Jarid1c</i>	N.I.	–	–	–	–	–
<i>Jarid1d</i>	N.I.	–	–	–	–	–
<i>Jarid2</i>	N.S.	–	Fragment	Patsialou et al. (2005)	GST-pulldown	–
	CCG-repeats, GA-rich	–	FL	Peng et al. (2009)	ChIP	–
	N.S.	–	ARID	Kusunoki et al. (2009)	NMR	–
<i>Swi1</i> (S. c)	N.S.	–	ARID	Wang et al. (2012)	SPR, NMR	–
<i>Arid4</i> (A.t)	TTTAGATCTAAA	21.6	ARID-PHD	Tan et al. (2020)	EMSA, ITC	6LFQ ^g
<i>Dri</i> (D.m)	CCTGTATTGATGTGG	0.03	ARID	Iwahara et al. (2002)	EMSA, NMR	1KQQ ^h
<i>Kdm5</i> (D.m)	N.I.	–	–	–	–	–

N.S. – non-sequence specific; N.I. – no information available. ^aJust one sequence tested – no conclusion about different target DNA preferences possible. ^bSuggested by homology to *Dri* and *BRIGHT*, respectively. ^cProtein constructs used for DNA-binding studies: FL – full-length protein; Fragment – fragment of full-length protein, including ARID; ARID – isolated ARID domain; Tandem domains are indicated by their corresponding abbreviations (see also Figure 1). ^dVariable K_D values found depending on different DNA length and sequence. ^emodelled using Haddock (van Zundert et al. 2016). ^fPRE-based Haddock-refined NMR structure in complex with DNA. ^gCrystal structure of tandem domain in complex with DNA and H3K4me3 peptide. ^hNMR structure in complex with DNA.

Figure 1: Domain architectures of the 15 human *Arid* proteins (in the seven sub-families *Arid1-5* and *Jarid1* and 2) aligned as centered to the ARID domain.

Four selected non-human *Arids* (caskets) are included, placed above their closest human paralogs – if applicable. Other protein domains and motifs are given (see bottom of the figure). The length of each protein (amino acids, aa) is shown on the right. *Arid* protein sequences and annotated domains were taken from UniProt (UniProt 2021) and related literature. Uncharacterized, folded regions were predicted from AlphaFold (Jumper et al. 2021).

the pivotal role of *Arids* in a large number of human diseases, especially cancer. While some members of the *Arid* family actively function as tumor suppressors (Bluemn et al. 2019; Chang et al. 2019; Luchini et al. 2015), over-expression of others provokes opposing effects in tumor growth [reviewed in (Lin et al. 2014)]. This underlines the high clinical relevance of *Arid* proteins and creates a need for the comprehensive and precise understanding of the versatile *Arid* protein functions, in particular DNA-binding.

In this review we summarize available data on the ARID domains found within individual protein sub-families. After comparing sequence and fold conservation, we provide a detailed overview of the current data on DNA sequence specificities for ARID containing proteins, including the few complex structures with DNA. This overview mainly concentrates on the human *Arid* proteins, yet we include the available structural information of *Arids* from *Saccharomyces cerevisiae* (*Swi1*), *D. melanogaster* (*Dri* and *Kdm5*) and *Arabidopsis thaliana* (*Arid4*) for comparison (Table 1, Figure 1).

Classes of *Arid* proteins

In human cells, to date a total of 15 ARID containing proteins have been described, divided in seven sub-families (Table 1). Most human *Arid* proteins were discovered as key players in epigenetic regulation, either by direct interaction with chromatin structures, e.g. represented by the *Jarid1* sub-family (Secombe and Eisenman 2007), or as part of chromatin remodeling complexes as shown for *Arid1a/b* or *Arid2* (Euskirchen et al. 2012). The underlying mechanisms are often incompletely understood and typically involve multiple co-factors.

The classification of *Arid* proteins is based on their domain context (with exception of *Arid5*, see below) and co-occurrence of specific domains is a significant feature of individual sub-families (Figure 1). A clear correlation of ARID sequence conservation and the corresponding domain architecture suggests co-evolution of domains (Sandhya et al. 2018) (Figure 2). In many cases, ARIDs appear together with further DBDs, chromatin reader

Figure 2: Evolutionary relationship within the family of eukaryotic full-length *Arid* proteins and their ARID domains.

(A) Heatmap comparing sequence identities of all human *Arid* full-length proteins and four selected non-human representatives. (B) Heatmap of ARID core domains correlated to panel A. The inserted phylogenetic trees depict the degree of evolutionary relationships for panels A and B, respectively. Sequences were taken from annotated UniProt entries, by default isoform 1 (UniProt 2021), and aligned with Clustal Omega (Sievers et al. 2011).

domains or motifs/domains for PPIs. This indicates combinatorial interactions with DNA or other proteins, e.g. subunits of chromatin remodeling complexes, as recently shown for ARID-PHD and for ARID-HMG cassettes, both found in plants (Hansen et al. 2008; Tan et al. 2020).

Arid1a and *Arid1b* are distinct subunits of the SWI/SNF chromatin remodeling complex (Euskirchen et al. 2012). Both members of the *Arid1* sub-family are large multidomain proteins harboring at least two folded domains. Besides a centrally located ARID they comprise the C-terminal BAF250_C, including an Armadillo (ARM) repeat (Sandhya et al. 2018) suggested to be involved in PPIs within SWI/SNF (Zinzalla 2016). Interestingly, the ortholog protein *SwiI* in *S. cerevisiae* suggests a similar domain setup with a still unexplored domain predicted at its C-terminus.

Arid2 is a specific subunit of the chromatin remodeling complex PBAF that otherwise shares many elements with the SWI/SNF complex (Wang et al. 2004; Yan et al. 2005). *Arid2* harbors an N-terminally located ARID, an RFX-like DBD defined by a winged-helix type motif (Gajiwala et al. 2000) and a twin zinc finger (tZnF) located C-terminally (Hatayama and Aruga 2010). As part of the PBAF complex, *Arid2* plays an important role in the regulation of cell differentiation (Xu et al. 2012).

The sub-family of *Arid3* counts three paralogs – *Arid3a*, *b* and *c* – that feature a characteristic combination of ARID and REKLES. The REKLES domain is required for self-association and also harbors a nuclear export signal (NES) allowing *Arid3* members to shuttle between nucleus and cytoplasm (Kim et al. 2007; Kim and Tucker 2006) (Figure 1, Table 1). Murine *Arid3a* – also known as *Bright* – together with the *D. melanogaster* ortholog *Dri* are founding members of the large family of ARID containing proteins (Gregory et al. 1996; Herrscher et al. 1995).

Arid4a and *Arid4b* both are part of the mSin3/HDAC1 complex involved in epigenetic regulation and gene repression. In addition, they function as co-regulators of transcriptional activation mediated by the androgen receptor (Lai et al. 2001; Wu et al. 2013). Three of the five domains in *Arid4* belong to the ‘Royal family domains’, which are potentially able to bind to methylated histone tails: Tudor, PWWP, and chromo barrel. Yet in case of *Arid4a*, only the latter is able to recognize these epigenetically relevant marks (Gong et al. 2012). While Tudor has been demonstrated to possess dsDNA-binding activity (Gong et al. 2014), PWWP domains in general are suggested to play a role in PPIs (Stec et al. 2000).

The *Arid5* sub-family comprises two members – *Arid5a* and *Arid5b* – that share high sequence similarity exclusively within the ARID domain (Figures 2 and 3A) in an otherwise largely unstructured context. It is noteworthy that both proteins differ significantly in size, which is unique for *Arids* of one sub-family. Structure prediction within the N-terminus of the significantly longer *Arid5b* by AlphaFold (<https://alphafold.ebi.ac.uk>) (Jumper et al. 2021; Tunyasuvunakool et al. 2021) indicates the presence of an additional BAH domain; a motif that often occurs in chromatin-interacting proteins (Yang and Xu 2013). *Arid5a* was recently described as RNA-binding protein (RBP) that can shuttle into the cytoplasm and stabilize mRNAs (Masuda et al. 2013; Nyati et al. 2019). Thus, for the *Arid5* sub-family we currently face highly divergent functions of its two members (see last chapter of this article).

Members of the *Jarid1* sub-family (*Jarid1a-d*) are histone demethylases and thus, transcriptional repressors. They contain a Jumonji C and N (JmjC, JmjN) domain, required for removing H3K4 methyl marks, which is e.g. excellently summarized in (Secombe and Eisenman 2007). A comparable domain architecture is found in the *D. melanogaster* homolog *Kdm5* (Secombe et al. 2007). Current literature controversially discusses, whether the ARID domain in *Jarids* is essential for demethylase activity (Horton et al. 2016; Johansson et al. 2016; Tu et al. 2008; Xiang et al. 2007).

The *Jarid2* protein is a further member of the large Jumonji histone demethylase family (Franci et al. 2014), yet it is catalytically inactive due to sequence alterations (Landeira and Fisher 2011). Still, it fulfills essential gene regulatory functions by interacting with the polycomb repressive complex 2 (*PRC2*) (Li et al. 2010; Peng et al. 2009). (Li et al. 2010) claim its ARID domain supports the C-terminal zinc finger in directing *PRC2* to target sequences.

Phylogenetic correlation between *Arid* proteins

For *Arid* proteins, the highest homologies of full-length sequences are found between the members of one sub-family (Figure 2A). This correlates with a highly similar overall domain architecture (Figure 1) and is reflected e.g. in 52–86% identity between the four *Jarid1* proteins and around 50% between individual members of the sub-families *Arid1* to *Arid5*, respectively. In contrast, homologies for proteins across sub-families are low, with identities of only 10–20% (e.g. 12% identity between *Arid1a* and *Arid4a*), which altogether suggests an early branching

of *Arid* proteins during evolution. In line with that, sequence comparison of the respective ARID domains alone reveals identities of 60–89% between sub-family members (Figure 2B). Interestingly, the similarity of ARID domains across species is remarkably high for *Arids* with a similar overall domain organization, e.g. when comparing the ARID sequences of human *Arid3* members (*a* to *c*) with the *Drosophila* ortholog *Dri* (Figures 2B and 3A). Their ARID sequences are more than 78% identical, and also the co-occurring REKLES domain is shared among human and fly *Arid3* proteins (Kim et al. 2007).

Conservation and variation of ARID domains

The ARID domain is in average 100 amino acids long and shows a broad sequence variety. Yet, there are 13 highly conserved residues, five out of which are identical among all domains (Figure 3A). To our knowledge, 23 ARID structures have been released in the Protein Data Bank (PDB, <https://www.rcsb.org>, as of November 1st, 2021) (Berman et al. 2000) within the last two decades. Alignment of their secondary structure elements shows a conserved arrangement of six helices, H1–H6, which together form the minimal ARID core domain. The ARID core can be extended by either N- or C-terminal helices H0 and H7, respectively, referred to as eARID (Kortschak et al. 2000). Representatives of eARID containing proteins are the *Arid3* sub-family members (Figure 3B).

Interestingly, ARID sequence homology is not directly correlated with the degree of ARID fold conservation (Figure 4A and B). *Z*-scores – derived from ARID structure comparison using the DALI server (Holm 2020) – allow to pairwise examine the 3D convergence of domains (Figure 4B). For example, ARID of *Arid1b* shares 78% sequence identity with *Arid1a* ARID, and – with a *z*-score of 11.2 – their structures are very similar (r.m.s.d. 3 Å). A similar structural homology is found between ARIDs of *Arid1b* and *Jarid1c* (*z*-score 10.7; r.m.s.d. 3.05 Å), despite only 24% sequence homology. Furthermore, *z*-score comparison impressively shows the large deviation of the yeast *Swi1* ARID from all other ARID structures (Figure 4B), while this difference is remarkably less pronounced on the sequence level (Figure 4A).

Between helices H1 and H2, all ARIDs contain an extended loop L1 (Figure 3B). A second loop L2 is located between H3/H4 and H5, which together form the HTH motif responsible for interaction with dsDNA. In comparison to the canonical HTH motif, ARIDs comprise an atypical HTH

Figure 3: Sequence and secondary structure conservation of ARID domains.

(A) Sequence alignment of ARID core domains, based on Clustal Omega (Sievers et al. 2011). Members of *Arid* sub-families are grouped and sequence variations depicted relative to the respective “a”-member. Conserved residues are highlighted (grey – similar; black – identical) according to the 7 groups: GAVLI, FYW, CM, ST, KRH, DENQ, P. Positions of helices H1 to H6 are shown by yellow rectangles, and soft edges comprise the total range of element location. The sequences include ARIDs from the 15 human *Arid* proteins as well as selected non-human *Arids* from *S. cerevisiae* (S.c.), *A. thaliana* (A.t.), *Mus musculus* (M.m.) and *D. melanogaster* (D.m.). (B) Alignment of ARID secondary structural elements from representative structures, comprising the core domain and 18 flanking N-terminal and 20 C-terminal amino acids, respectively. For the latter, their total pI is indicated by a color code explained below. Solid lines cover the sequence range found in a corresponding PDB entry (method given). Core motif helices H1 to H6 are depicted as yellow, extending helices H0 and H7 as light-yellow cylinders, respectively. Significant amino acid insertions or deletions relative to the conserved sequence (see panel A) are indicated by broken lines or dots, respectively.

Figure 4: Sequence and structure conservation in eukaryotic ARID domains.

(A) Phylogenetic tree and heatmap of sequence conservation within representative PDB entries across human and selected non-human ARID domains. Colors refer to *Arid* sub-families as introduced in Figure 2. Sequence alignments were performed with Clustal Omega (Sievers et al. 2011). (B) Phylogenetic tree and heatmap of structure conservation within the PDB entries as in A based on the z-score provided by DALI (Holm 2020). (C) The ARID HTH motif (blue) as taken from PDB entry 2EH9. Additional helices H2 and H3 are included. (D) Sidechains of least and most conserved residues from ARID sequences within proteins from Table 1 are shown as spheres and conservation plotted on the *Arid1b* ARID domain (2EH9). Conservation plots have been performed with the ConSurf server (Ashkenazy et al. 2016). (E) The HTH motif from panel C as embedded within the ARID structure. Sidechains of the most conserved residues within the hydrophobic core (see D) are shown as sticks.

with an unusually long “turn” L2 (Iwahara and Clubb 1999) (Figure 4C). The HTH motif is stabilized by helix H2, which is important for the compact globular fold (Huffman and Brennan 2002). In analogy to other HTHs, H5 represents the so-called DNA recognition helix, which mediates contacts with the DNA major groove (Brennan and Matthews 1989; Iwahara and Clubb 1999).

Comparative mapping of ARID residues with the highest versus those with the lowest conservation reveals an evident feature: while non-conserved residues are predominantly located on the surface of the structure, highly conserved residues show a side chain orientation towards the hydrophobic core of the domain, formed between helices H2 and H5 (Figure 4D and E). (Chandler et al. 2013) suggested the preservation of this hydrophobic pocket to be an important aspect of ARID functionality. In their work, a centrally located valine (V1067) within ARID of *Arid1a* was found essential for DNA-binding. Recent work

indicates that the hydrophobicity of valine at this position significantly supports thermal stability and its mutation to glycine (V1067G) enhances unfolding of the ARID domain (Sandhya et al. 2018).

Interestingly, residues directly within the DNA-binding interface of the HTH motif are little conserved across *Arids* in general. Yet, they show high conservation among members of each sub-family (Figure 3A). Similarly, least-conserved residues among ARIDs show high conservation within individual sub-families (Figure 3A). This is in accordance with the suggested co-evolution of domains in *Arids*. Sandhya et al. (2018) claim certain surface exposed ARID residues are conserved only within proteins of an *Arid1*-like domain architecture, arguing for a co-evolution with the BAF250_C domain. As observed early for other DBDs (D’Elia et al. 2001), evolution might have impacted the tertiary fold of the ARID domain, rather than its primary sequence.

Different flavors of ARID domains

On a structural level, the helices H1–H6 together with the embedded loops L1 and L2 (Figure 5A) form the minimal core domain in all ARIDs (Figure 5B), which can e.g. be found in *Jarid1d* (Figure 5C). Most ARID structures reveal deviations from this prototypic organization, which appear either in their N- and C-termini or within the loop regions and can occur alone or in combination (Figure 5D–H). In this context, a prevalent feature found in more than 60% of ARIDs is a β -hairpin in L1, that has first been described for *Dri* (Iwahara and Clubb 1999). The two short antiparallel β -strands significantly reduce flexibility of loop L1, which is involved in DNA contacts (Cai et al. 2007; Iwahara and Clubb 1999). The eARID includes additional helices at the N- (H0), the C- (H7), or both termini (Figure 5D–F). H7 allows additional contacts with the target DNA, thereby enlarging the binding surface (Iwahara and Clubb 1999).

H0 was described to fold back onto the core domain, and very recently for *Arid4a*, to auto-inhibit through mimicking DNA and to play a key role in specific chromatin interaction (Gong et al. 2021; Liu et al. 2010). Strikingly, the termini flanking the ARID core domains differ immensely in their net charge. E.g., a highly acidic N-terminal stretch is common to all ARIDs with an additional N-terminal H0 (with exception of *Jarid1c*) (Figure 3B). This hints to similarly important roles for these extending regions as described for *Arid4a*. Further ARID variations are manifested in significantly altered loop regions. L1 and L2 of the *A. thaliana* *Arid4* ARID domain are prolonged, while L2 of *S. cerevisiae* *Swi1* is shorter compared to the typical core domains.

Of importance, most of the currently available ARID structures are derived from isolated domains or found within tandems (like ARID-PHD or PWWP-ARID) (Table 1). Yet, high resolution crystal structures of the amino-terminal half of *Jarid1a* impressively show the integration

Figure 5: Variability in ARID structures.

(A) Secondary structure of the minimal ARID core motif, the optional helices H0 and H7, and an optional β -hairpin in L1. (B) Schematic 3D arrangement of ARID core elements from A. (C) Representative minimal ARID core domain structure from human *Jarid1d* (2YQE). (D–H) Representative ARID structures with known types of variations. Structures are presented in the same orientation. Additional elements, highlighted in magenta, show: (D) N-terminal helix H0 and β -hairpin in L1 from human *Arid1b* (2EH9), (E) C-terminal helix H7 and β -hairpin in L1 from murine *Jarid2* (2RQ5), (F) both N- and C-terminal helices H0 and H7, and β -hairpin in L1 from human *Arid3a* (2KK0), (G) an extended L1 (by 10 aa) and L2 (by 3 aa), respectively, and β -hairpin in L1 from *Arid4* from *A. thaliana* (6LQF), and (H) a shortened L2 (by 3 aa) from *Swi1* from *S. cerevisiae* (2LI6).

of ARID into a multidomain context (Labadie et al. 2016; Liang et al. 2017; Vinogradova et al. 2016). In 2016, Vinogradova and colleagues reported the isolated ARID structure to be well superimposable with the domain in its *Jarid1a* embedment with only minor differences in loop conformations and an r.m.s.d value of 1.6 Å (Vinogradova et al. 2016). This supports the autonomy and stability of the ARID fold (Labadie et al. 2016; Vinogradova et al. 2016). However, structural data for ARIDs within an extended protein or even multi-protein complex is rare and, astonishingly, ARIDs are unresolved in available cryo-EM and crystal structures of yeast and human chromatin remodeling complexes, respectively (Grau et al. 2021; Han et al. 2020; He et al. 2020). In fact, Reyes, Marcum and He very recently suggested this observation to be based on a high conformational flexibility, even upon DNA-binding (Reyes et al. 2021).

Structures of ARID domains bound to DNA

The DBD character of ARIDs was identified in structural studies, which revealed a modified HTH motif (Figure 4C) (Cai et al. 2007; Iwahara et al. 2002; Kim et al. 2004; Zhu et al. 2001). Only three experimental high-resolution structures of ARIDs in complex with DNA exist to date (Cai et al. 2007; Iwahara et al. 2002; Kim et al. 2004; Zhu et al. 2001), which comprise an ARID core, an eARID (both determined by NMR) and a crystal structure of a plant ARID domain in concert with a PHD motif (Figure 6). In all three of them, we find very similar or identical interfaces of ARIDs with dsDNA as visible after a structural alignment of both protein and DNA components, and a renumbering of respective sequences for convenient comparison (Figure 6A–D). While the three complexes reveal different extents of nucleotide-specific interactions, those are consistently formed between the DNA major groove AT base pairs and protein sidechains in L2 and H5, which are part of the HTH motif (Figure 6A). Additional non-specific interactions are found between the ribose-phosphate backbone of the adjacent minor groove and ARID residues in L1 and (optionally) the C-terminal region (Iwahara et al. 2002; Zhu et al. 2001) (Figure 6B–D). Within *Dri*, two additional specific contacts are mediated by an arginine within helix H7, a unique feature of eARID domains. This extension of the DNA-binding interface is suggested to result in higher affinities for its target DNA compared to core ARID domains (Table 2) (Iwahara and Clubb 1999) and interestingly involves a GC-TA tandem base pair

(Figure 6C). Within the *Arid4* ARID-PHD complex structure, one additional non-specific contact from the PHD domain contributes to DNA-binding (Figure 6D), demonstrating possible combinatorial interactions of ARID along with co-occurring domains and likely not the single case.

In each of the three ARID-DNA complex structures, specific contacts are formed with at least one central AT base pair. This contact is centered around the crucial ARID residue Thr71 (numbering with respect to normalized sequences as shown in Figure 6A), which is located at the transition from L2 to H5 and forms a hydrogen bond with either Ade or Thy (Figure 6B–D). In 2007, Cai and colleagues solved the NMR structure of human *Arid5b* in complex with an AT-rich dsDNA supported by paramagnetic relaxation enhancement caused by spin-labeled DNA (Cai et al. 2007). This elegant approach revealed an ensemble of complexes where Thr71 can specifically interact with one of two neighboring Thy bases of an AT-tandem base pair (Figure 6E), which at this time underlined the preference of the ARID core for AT-DNAs.

AT or not AT: sequence preferences and affinities across *Arids*

(Tu et al. 2008) recently emphasized the correlation of threonine at position 71 within ARIDs (numbering with respect to normalized sequences as shown in Figure 6A) with a preference for AT-rich sequences. ARID domains that share a conserved lysine at the equivalent position were found to target GC-rich sequences, whereas a serine is correlated with non-specific DNA-binding (Figure 6F). However, we have to note that those findings are to date not corroborated by experimental structural data, which uniquely would allow to derive atom-resolved correlations of specificity between ARID sidechains and multiple non-AT base pairs.

In general, the available data on target sequences reveals our limited in-depth knowledge about preferences, specificities and determinants of ARID-DNA interactions. As shown in Table 2, a large fraction of available data refers to SELEX- or ChIP (chromatin immunoprecipitation)-based approaches. Here, the enrichment of DNAs specifically bound by *Arid* proteins has been carried out with full-length proteins or protein fragments. As a result, it is difficult to draw conclusions on DNA-recognition that are transferable to the isolated ARID domain. In 2005, Moren and colleagues performed a systematic analysis of DNA-binding across all human *Arids* and found no obvious

Figure 6: The DNA-binding specificity of ARID domains. (A) Structure-guided sequence alignment of ARID residues as shown in PDB entries for the three experimental structures with DNA. Helices and the HTH are indicated. The numbering refers to the isolated ARID domain of human *Arid5b* (PDB 20EH) and analogous positions in the other two PDBs are used for better comparison in panels B–D. Orange positions and asterisks highlight residues with specific recognition of DNA. The violet area indicates the C-terminal extension of the ARID core in 1KQQ, which provides an additional specific contact. The sequence of the ARID-flanking PHD domain in 6LQF is indicated with a green box. (B–D) Top, PDB entries from A showing the three ARID domains with highlighted residues (spheres) relative to the DNA. Only the relevant fraction of DNA is shown as ribbon. Bottom, schematic interactome of contacts as present in structures above. Grey indicates non-sequence specific interactions with the DNA backbone, while orange residues, positioned within the WC-scheme indicate specific interactions. Note the referenced sequence numbering of proteins shown in A. (E) Zoom-in of protein-DNA interface in PDB 20EH highlighting the interactions of three conserved residues with the tandem AT base pair (see panel B). (F) Comparison of ARID domain HTH motifs with the respective residue 71 and the suggested DNA base pair preference. Two representative PDBs are superimposed for the region of interest accompanied by a simplified scheme below.

sequence specificity in five out of the seven sub-families (Patsialou et al. 2005). However, their data derive from a mere qualitative approach using pulldown experiments and do not yield a quantitative measure for specificity. It also ignores DNA length, which has a significant influence on stoichiometry and thus, affinities (Maulik et al. 2019). In contrast, two groups have suggested a GC preference within the *Jarid1* sub-family (Scibetta et al. 2007; Tu et al. 2008). In other studies, the highly divergent methodological setups have led to experimentally observed K_D values over more than the orders of magnitude, i.e. low nanomolar to micromolar (Table 2).

The latter is best exemplified by the controversial findings for the ARID domain of *Arid1a* (Table 2), where a large experimental variance in both DNA lengths and respective ARID domain boundaries complicate data interpretation. The ARID domains of *Arid1a/b* and *Arid2* have been described to bind DNA non-specifically and with inconsistent affinities (Kim et al. 2004; Patsialou et al. 2005; Wang et al. 2004). Their individual incorporation as subunits into SWI/SNF or PBAF complexes resulted in distinct roles in cell cycle control (Nagl et al. 2007). This hints towards an intrinsic target DNA-specificity of each *Arid*, emerging only in combination with further domains. Possibly, ARID domains are only able to exhibit their subtle preferences for certain DNAs when properly positioned with the help of additional SWI/SNF components. Important *Arid*-mediated PPIs are also suggested in the growing number of cryo-EM structures of nucleosomes (Han et al. 2020; He et al. 2020; Kasinath et al. 2021).

Of note, a number of studies have taken into account the potential role of molecular dynamics in DNA-recognition by ARIDs (Iwahara et al. 2002; Kim et al. 2004; Kusunoki et al. 2009; Maulik et al. 2019) using solution NMR spectroscopy. Overcoming failure to co-crystallize, the unique and valuable information of chemical shifts has allowed to model or experimentally determine complexes of ARIDs with their suggested DNA motifs (Cai et al. 2007; Maulik et al. 2019), and NMR has further helped in identifying the role of flexibility in loops within ARIDs (Kim et al. 2004).

Molecular dynamics (MD) are also possible to target with computational simulations. For *Arid3a*, (Invernizzi et al. 2014) used MD to postulate specific “hub” residues in the ARID domain involved in DNA-binding. They showed that Tyr119 plays a key role in long-range communication to conserved loop residues in the DNA interface, which allowed speculating about coupling DNA-binding to intramolecular communication.

Why only DNA: RNA-binding by *Arid5a*

As a unique occurrence among *Arids*, in 2013 Kishimoto and colleagues described *Arid5a* as a novel RBD (Masuda et al. 2013). In this and a number of later studies the protein was found to stabilize mRNAs by specifically binding to short 3'-untranslated region stem-loops with no particular consensus sequence (Hanieh et al. 2018; Masuda et al. 2016; Zaman et al. 2016). Those publications suggested that *Arid5a* blocks binding of the immune-suppressive RBPs Regnase and Roquin, both known to tightly engage with identical stem-loops or overlapping regions in mRNAs (Janowski et al. 2016; Jeltsch et al. 2014; Mino et al. 2015; Schlundt et al. 2014). Thus, *Arid5a* was categorized as proinflammatory RBP [e.g. reviewed in (Nyati et al. 2020)], with evidences for an ARID-mediated interaction with particular RNAs (Hanieh et al. 2018; Masuda et al. 2013, 2016). Of note, to our knowledge there is no study that provides binding data of the isolated *Arid5a* ARID domain to one of the designated RNA elements.

As of now, it remains to clearly disentangle the (simultaneous) DBP- from RBP functions of *Arid5a*, especially in the upregulation of genes during immune responses, as e.g. for IL-6 (Ikeuchi et al. 2021; Masuda et al. 2013). However, we have to speculate about the general participation of *Arids* in dual-functioning as DNA/RNA-binding proteins, which occur increasingly abundant at the chromatin-RNA interface (Hudson and Ortlund 2014; Huo et al. 2020; Xiao et al. 2019).

Conclusion and future perspectives

As summarized in this and other overview articles, *Arids* fulfill cellular functions centered around gene regulation including potential roles in tumorigenesis (Kortschak et al. 2000; Lin et al. 2014; Patsialou et al. 2005; Wilsker et al. 2002). We need to revalidate ARIDs with respect to their original definition as pure AT-rich dsDNA binding domains, and rather conceive their core shape as *bona fide* dsDNA-binding domain (D'Elia et al. 2001; Gregory et al. 1996; Wilsker et al. 2005). In this, we also suggest to consider the role of DNA tertiary structure, e.g. its bending, looping, supercoiling in nucleosomes and temporary triplex structures (Neidle 2021; Wolberger 2021). Certain DNA stretches might experience an unexpected preference

given by a pre-organized geometry that would support an increased on-rate in ARID binding.

A systematic biochemical and structural analysis of ARID-DNA complexes will improve our understanding of target preference and predictability. Perspectively, it allows engineering of ARIDs against DNA sequences of choice as shown for other NA-binding domains (Chen and Varani 2013; Inamoto et al. 2021; Sera and Uranga 2002; Zhou et al. 2021). A combination of SELEX and ChIP-seq will relate high-affine motifs to natural sequence contexts (Kharchenko et al. 2008; Narlikar and Jothi 2012). Also, this approach will disentangle contributions of ARIDs to DNA targets in MD-DBPs.

Much light will be shed on the atomic basis of specific ARID-DNA interactions via additional high-resolution complex structures. In addition, the systematic analysis of permuted DNA motifs against an ARID, e.g. by NMR-based principal component analysis (Collins et al. 2015) will explain subtle differences in affinities between motifs. Additional sources of structure-providing methods are to be exploited, including homology-based modelling, docking and MD simulations (Koukos and Bonvin 2020; Trnka et al. 2019). E.g. solution-oriented integrated structural biology (Rout and Sali 2019) was recently applied to a bacterial promoter DNA-protein complex (Schlundt et al. 2017).

Consequently, future work will explore larger DNA-Arid complexes that to date do not provide an unambiguous picture of the ARID domain, e.g. in the recently determined cryo-EM structures of nucleosomes (Han et al. 2020; He et al. 2020; Kasinath et al. 2021). Also, for most *Arid* proteins we still lack the meaning of the extended unstructured, but highly-conserved regions (*Arid5a*). In particular for TFs, the recent years have revolutionized our view on IDRs as potent contributors to specific DNA interactions (Brodsky et al. 2020; Guo et al. 2012; Vuzman and Levy 2012). In *Arids*, such IDRs might as well act in concerted action with the ARID domain.

Finally, systematic research is needed to understand mammalian ARIDs in their full capacity including potentially hidden capabilities of protein interaction domains as known from plants (Tan et al. 2020; Zhu et al. 2008). Altogether, despite their relatively early discovery, ARID domains have left us with many unknowns and atom-resolving technology will remarkably help to fill the many gaps in understanding functions and mechanisms of *Arids* mediated by their ARID domain.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: This work was funded by Johanna Quandt Young Academy at Goethe (2019/AS01) and Deutsche Forschungsgemeinschaft (SCHL2062/2-1, SFB902-3/B16).

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

- Animireddy, S., Kavadiyula, P., Kotapalli, V., Gowrishankar, S., Rao, S., and Bashyam, M.D. (2021). Aberrant cytoplasmic localization of ARID1B activates ERK signaling and promotes oncogenesis. *J. Cell Sci.* 134: jcs251637.
- Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M., and Iyer, L.M. (2005). The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.* 29: 231–262.
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 44: W344–W350.
- Bateman, N.W., Shoji, Y., Conrads, K.A., Stroop, K.D., Hamilton, C.A., Darcy, K.M., Maxwell, G.L., Risinger, J.I., and Conrads, T.P. (2016). Identification and functional characterization of a novel bipartite nuclear localization sequence in ARID1A. *Biochem. Biophys. Res. Commun.* 469: 114–119.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The protein data bank. *Nucleic Acids Res.* 28: 235–242.
- Bluemn, T., Schmitz, J., Zheng, Y., and Zhu, N. (2019). Both Arid1b and Arid2 are tumor suppressors in MLL-AF9 leukemogenesis. *Blood* 134: 1248.
- Bobbs, A., Gellerman, K., Hallas, W.M., Joseph, S., Yang, C., Kurkewich, J., and Cowden Dahl, K.D. (2015). ARID3B directly regulates ovarian cancer promoting genes. *PLoS One* 10: e0131961.
- Brennan, R.G. and Matthews, B.W. (1989). Structural basis of DNA-protein recognition. *Trends Biochem. Sci.* 14: 286–290.
- 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.
- 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.
- Chandler, R.L., Brennan, J., Schisler, J.C., Serber, D., Patterson, C., and Magnuson, T. (2013). ARID1a-DNA interactions are required for promoter occupancy by SWI/SNF. *Mol. Cell Biol.* 33: 265–280.
- Chang, S., Yim, S., and Park, H. (2019). The cancer driver genes IDH1/2, JARID1C/KDM5C, and UTX/KDM6A: crosstalk between histone demethylation and hypoxic reprogramming in cancer metabolism. *Exp. Mol. Med.* 51: 1–17.
- Chen, Y. and Varani, G. (2013). Engineering RNA-binding proteins for biology. *FEBS J.* 280: 3734–3754.
- Collins, K.M., Oregioni, A., Robertson, L.E., Kelly, G., and Ramos, A. (2015). Protein-RNA specificity by high-throughput principal component analysis of NMR spectra. *Nucleic Acids Res.* 43: e41.

- D'Elia, A.V., Tell, G., Paron, I., Pellizzari, L., Lonigro, R., and Damante, G. (2001). Missense mutations of human homeoboxes: a review. *Hum. Mutat.* 18: 361–374.
- Dallas, P.B., Pacchione, S., Wilsker, D., Bowrin, V., Kobayashi, R., and Moran, E. (2000). The human SWI-SNF complex protein p270 is an ARID family member with non-sequence-specific DNA binding activity. *Mol. Cell Biol.* 20: 3137–3146.
- Douclevy, M. and Clore, G.M. (2008). Global jumping and domain-specific intersegment transfer between DNA cognate sites of the multidomain transcription factor Oct-1. *Proc. Natl. Acad. Sci. U.S.A.* 105: 13871–13876.
- Euskirchen, G., Auerbach, R.K., and Snyder, M. (2012). SWI/SNF chromatin-remodeling factors: multiscale analyses and diverse functions. *J. Biol. Chem.* 287: 30897–30905.
- Franci, G., Ciotta, A., and Altucci, L. (2014). The Jumonji family: past, present and future of histone demethylases in cancer. *Biomol. Concepts* 5: 209–224.
- Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B., and Burley, S.K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* 403: 916–921.
- Gong, W., Liang, Q., Tong, Y., Perrett, S., and Feng, Y. (2021). Structural insight into chromatin recognition by multiple domains of the tumor suppressor RBBP1. *J. Mol. Biol.* 433: 167224.
- Gong, W., Wang, J., Perrett, S., and Feng, Y. (2014). Retinoblastoma-binding protein 1 has an interdigitated double tudor domain with DNA binding activity. *J. Biol. Chem.* 289: 4882–4895.
- Gong, W., Zhou, T., Mo, J., Perrett, S., Wang, J., and Feng, Y. (2012). Structural insight into recognition of methylated histone tails by retinoblastoma-binding protein 1. *J. Biol. Chem.* 287: 8531–8540.
- Grau, D., Zhang, Y., Lee, C.H., Valencia-Sanchez, M., Zhang, J., Wang, M., Holder, M., Svetlov, V., Tan, D., Nudler, E., et al. (2021). Structures of monomeric and dimeric PRC2:EZH1 reveal flexible modules involved in chromatin compaction. *Nat. Commun.* 12: 714.
- 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.
- Gribskov, M. and Burgess, R.R. (1986). Sigma factors from *E. coli*, *B. subtilis*, phage SP01, and phage T4 are homologous proteins. *Nucleic Acids Res.* 14: 6745–6763.
- Guan, B., Gao, M., Wu, C.H., Wang, T.L., and Shih, Ie, M. (2012). Functional analysis of in-frame indel ARID1A mutations reveals new regulatory mechanisms of its tumor suppressor functions. *Neoplasia* 14: 986–993.
- Guo, X., Bulyk, M.L., and Hartemink, A.J. (2012). Intrinsic disorder within and flanking the DNA-binding domains of human transcription factors. *Pac. Symp. Biocomput.* 1: 104–115.
- Han, Y., Reyes, A.A., Malik, S., and He, Y. (2020). Cryo-EM structure of SWI/SNF complex bound to a nucleosome. *Nature* 579: 452–455.
- Hanieh, H., Masuda, K., Metwally, H., Chalise, J.P., Mohamed, M., Nyati, K.K., Standley, D.M., Li, S., Higa, M., Zaman, M.M., 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.
- Hansen, F.T., Madsen, C.K., Nordland, A.M., Grasser, M., Merkle, T., and Grasser, K.D. (2008). A novel family of plant DNA-binding proteins containing both HMG-box and AT-rich interaction domains. *Biochemistry* 47: 13207–13214.
- Hatayama, M. and Aruga, J. (2010). Characterization of the tandem CWCH2 sequence motif: a hallmark of inter-zinc finger interactions. *BMC Evol. Biol.* 10: 53.
- He, S., Wu, Z., Tian, Y., Yu, Z., Yu, J., Wang, X., Li, J., Liu, B., and Xu, Y. (2020). Structure of nucleosome-bound human BAF complex. *Science* 367: 875–881.
- 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.
- 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.
- Holm, L. (2020). Using dali for protein structure comparison. *Methods Mol. Biol.* 2112: 29–42.
- Horton, J.R., Engstrom, A., Zoeller, E.L., Liu, X., Shanks, J.R., Zhang, X., Johns, M.A., Vertino, P.M., Fu, H., and Cheng, X. (2016). Characterization of a linked Jumonji domain of the KDM5/JARID1 family of histone H3 lysine 4 demethylases. *J. Biol. Chem.* 291: 2631–2646.
- 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.
- Huffman, J.L. and Brennan, R.G. (2002). Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Curr. Opin. Struct. Biol.* 12: 98–106.
- Huo, X., Ji, L., Zhang, Y., Lv, P., Cao, X., Wang, Q., Yan, Z., Dong, S., Du, D., Zhang, F., 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.
- Ikeuchi, W., Wakita, Y., Zhang, G., Li, C., Itakura, K., and Yamakawa, T. (2021). AT-rich interaction domain 5A regulates the transcription of interleukin-6 gene in prostate cancer cells. *Prostate* 82: 97–106.
- Inamoto, I., Sheoran, I., Popa, S.C., Hussain, M., and Shin, J.A. (2021). Combining rational design and continuous evolution on minimalist proteins that target the E-box DNA site. *ACS Chem. Biol.* 16: 35–44.
- Invernizzi, G., Tiberti, M., Lambrugh, M., Lindorff-Larsen, K., and Papaleo, E. (2014). Communication routes in ARID domains between distal residues in helix 5 and the DNA-binding loops. *PLoS Comput. Biol.* 10: e1003744.
- Iwahara, J. and Clubb, R.T. (1999). Solution structure of the DNA binding domain from dead ringer, a sequence-specific AT-rich interaction domain (ARID). *EMBO J.* 18: 6084–6094.
- 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.
- Janowski, R., Heinz, G.A., Schlundt, A., Wommelsdorf, N., Brenner, S., Gruber, A.R., Blank, M., Buch, T., Buhmann, R., Zavolan, M., et al. (2016). Roquin recognizes a non-canonical hexaloop structure in the 3'-UTR of Ox40. *Nat. Commun.* 7: 11032.
- Jeltsch, K.M., Hu, D., Brenner, S., Zoller, J., Heinz, G.A., Nagel, D., Vogel, K.U., Rehage, N., Warth, S.C., Edelmann, S.L., 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.

- Johansson, C., Velupillai, S., Tumber, A., Szykowska, A., Hookway, E.S., Nowak, R.P., Strain-Damerell, C., Gileadi, C., Philpott, M., Burgess-Brown, N., et al. (2016). Structural analysis of human KDM5B guides histone demethylase inhibitor development. *Nat. Chem. Biol.* 12: 539–545.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596: 583–589.
- Kasinath, V., Beck, C., Sauer, P., Poepsel, S., Kosmatka, J., Faini, M., Toso, D., Aebersold, R., and Nogales, E. (2021). JARID2 and AEBP2 regulate PRC2 in the presence of H2AK119ub1 and other histone modifications. *Science* 371: eabc3393.
- Kharchenko, P.V., Tolstorukov, M.Y., and Park, P.J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.* 26: 1351–1359.
- Kim, D., Probst, L., Das, C., and Tucker, P.W. (2007). REKLES is an ARID3-restricted multifunctional domain. *J. Biol. Chem.* 282: 15768–15777.
- Kim, D. and Tucker, P.W. (2006). A regulated nucleocytoplasmic shuttle contributes to Bright's function as a transcriptional activator of immunoglobulin genes. *Mol. Cell Biol.* 26: 2187–2201.
- 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.
- Koehler, C., Bishop, S., Dowler, E.F., Schmieder, P., Diehl, A., Oschkinat, H., and Ball, L.J. (2008). Backbone and sidechain ¹H, ¹³C and ¹⁵N resonance assignments of the Bright/ARID domain from the human JARID1C (SMCX) protein. *Biomol. NMR Assignments* 2: 9–11.
- Kortschak, R.D., Tucker, P.W., and Saint, R. (2000). ARID proteins come in from the desert. *Trends Biochem. Sci.* 25: 294–299.
- Koukos, P.I. and Bonvin, A. (2020). Integrative modelling of biomolecular complexes. *J. Mol. Biol.* 432: 2861–2881.
- Kusunoki, H., Takeuchi, T., and Kohno, T. (2009). Solution structure of the AT-rich interaction domain of Jumonji/JARID2. *Proteins* 76: 1023–1028.
- Labadie, S.S., Dragovich, P.S., Cummings, R.T., Deshmukh, G., Gustafson, A., Han, N., Harmange, J.C., Kiefer, J.R., Li, Y., Liang, J., et al. (2016). Design and evaluation of 1,7-naphthyridones as novel KDM5 inhibitors. *Bioorg. Med. Chem. Lett.* 26: 4492–4496.
- Lai, A., Kennedy, B.K., Barbie, D.A., Bertos, N.R., Yang, X.J., Theberge, M.C., Tsai, S.C., Seto, E., Zhang, Y., Kuzmichev, A., et al. (2001). RBP1 recruits the mSIN3-histone deacetylase complex to the pocket of retinoblastoma tumor suppressor family proteins found in limited discrete regions of the nucleus at growth arrest. *Mol. Cell Biol.* 21: 2918–2932.
- Landeira, D. and Fisher, A.G. (2011). Inactive yet indispensable: the tale of Jarid2. *Trends Cell Biol.* 21: 74–80.
- Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev.* 24: 368–380.
- Liang, J., Labadie, S., Zhang, B., Ortwein, D.F., Patel, S., Vinogradova, M., Kiefer, J.R., Mauer, T., Gehling, V.S., Harmange, J.C., et al. (2017). From a novel HTS hit to potent, selective, and orally bioavailable KDM5 inhibitors. *Bioorg. Med. Chem. Lett.* 27: 2974–2981.
- Lin, C., Song, W., Bi, X., Zhao, J., Huang, Z., Li, Z., Zhou, J., Cai, J., and Zhao, H. (2014). Recent advances in the ARID family: focusing on roles in human cancer. *Oncotargets Ther.* 7: 315–324.
- 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.
- Liu, G., Huang, Y.J., Xiao, R., Wang, D., Acton, T.B., and Montelione, G.T. (2010). Solution NMR structure of the ARID domain of human AT-rich interactive domain-containing protein 3A: a human cancer protein interaction network target. *Proteins* 78: 2170–2175.
- Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V.N., and Dunker, A.K. (2006). Intrinsic disorder in transcription factors. *Biochemistry* 45: 6873–6888.
- Luchini, C., Veronese, N., Solmi, M., Cho, H., Kim, J.H., Chou, A., Gill, A.J., Faraj, S.F., Chaux, A., Netto, G.J., et al. (2015). Prognostic role and implications of mutation status of tumor suppressor gene ARID1A in cancer: a systematic review and meta-analysis. *Oncotarget* 6: 39088–39097.
- Luscombe, N.M., Austin, S.E., Berman, H.M., and Thornton, J.M. (2000). An overview of the structures of protein-DNA complexes. *Genome Biol* 1: REVIEWS001.
- Masuda, K., Ripley, B., Nishimura, R., Mino, T., Takeuchi, O., Shioi, G., Kiyonari, H., and Kishimoto, T. (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.
- Masuda, K., Ripley, B., Nyati, K.K., Dubey, P.K., Zaman, M.M., Hanieh, H., Higa, M., Yamashita, K., Standley, D.M., Mashima, T., et al. (2016). Arid5a regulates naive CD4+ T cell fate through selective stabilization of Stat3 mRNA. *J. Exp. Med.* 213: 605–619.
- 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.
- Minezaki, Y., Homma, K., Kinjo, A.R., and Nishikawa, K. (2006). Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J. Mol. Biol.* 359: 1137–1149.
- Mino, T., Murakawa, Y., Fukao, A., Vandenbon, A., Wessels, H.H., Ori, D., Uehata, T., Tarte, S., Akira, S., Suzuki, Y., et al. (2015). Regnase-1 and roquin regulate a common element in inflammatory mRNAs by spatiotemporally distinct mechanisms. *Cell* 161: 1058–1073.
- Nagl, N.G., Jr., Wang, X., Patsialou, A., Van Scoy, M., and Moran, E. (2007). Distinct mammalian SWI/SNF chromatin remodeling complexes with opposing roles in cell-cycle control. *EMBO J.* 26: 752–763.
- Narlikar, L. and Jothi, R. (2012). ChIP-Seq data analysis: identification of protein-DNA binding sites with SISR peak-finder. *Methods Mol. Biol.* 802: 305–322.
- Neidle, S. (2021). Beyond the double helix: DNA structural diversity and the PDB. *J. Biol. Chem.* 296: 100553.
- 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.
- 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.
- Nyati, K.K., Zaman, M.M., Sharma, P., and Kishimoto, T. (2020). Arid5a, an RNA-binding protein in immune regulation: RNA

- stability, inflammation, and autoimmunity. *Trends Immunol.* 41: 255–268.
- Ogata, K., Kanei-Ishii, C., Sasaki, M., Hatanaka, H., Nagadoi, A., Enari, M., Nakamura, H., Nishimura, Y., Ishii, S., and Sarai, A. (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.
- Patsialou, A., Wilsker, D., and Moran, E. (2005). DNA-binding properties of ARID family proteins. *Nucleic Acids Res.* 33: 66–80.
- Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* 139: 1290–1302.
- Reyes, A.A., Marcum, R.D., and He, Y. (2021). Structure and function of chromatin remodelers. *J. Mol. Biol.* 433: 166929.
- Rhee, C., Lee, B.K., Beck, S., Anjum, A., Cook, K.R., Popowski, M., Tucker, H.O., and Kim, J. (2014). Arid3a is essential to execution of the first cell fate decision via direct embryonic and extraembryonic transcriptional regulation. *Genes Dev.* 28: 2219–2232.
- Rout, M.P. and Sali, A. (2019). Principles for integrative structural biology studies. *Cell* 177: 1384–1403.
- 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.
- Schlundt, A., Buchner, S., Janowski, R., Heydenreich, T., Heermann, R., Lassak, J., Geerlof, A., Stehle, R., Niessing, D., Jung, K., et al. (2017). Structure-function analysis of the DNA-binding domain of a transmembrane transcriptional activator. *Sci. Rep.* 7: 1051.
- Schlundt, A., Heinz, G.A., Janowski, R., Geerlof, A., Stehle, R., Heissmeyer, V., Niessing, D., and Sattler, M. (2014). Structural basis for RNA recognition in roquin-mediated post-transcriptional gene regulation. *Nat. Struct. Mol. Biol.* 21: 671–678.
- Schmidt, C., Kim, D., Ippolito, G.C., Naqvi, H.R., Probst, L., Mathur, S., Rosas-Acosta, G., Wilson, V.G., Oldham, A.L., Poenie, M., et al. (2009). Signalling of the BCR is regulated by a lipid rafts-localised transcription factor. *Bright. EMBO J* 28: 711–724.
- Scibetta, A.G., Santangelo, S., Coleman, J., Hall, D., Chaplin, T., Copier, J., Catchpole, S., Burchell, J., and Taylor-Papadimitriou, J. (2007). Functional analysis of the transcription repressor PLU-1/JARID1B. *Mol. Cell Biol.* 27: 7220–7235.
- Secombe, J. and Eisenman, R.N. (2007). The function and regulation of the JARID1 family of histone H3 lysine 4 demethylases: the Myc connection. *Cell Cycle* 6: 1324–1328.
- Secombe, J., Li, L., Carlos, L., and Eisenman, R.N. (2007). The Trithorax group protein Lid is a trimethyl histone H3K4 demethylase required for dMyc-induced cell growth. *Genes Dev.* 21: 537–551.
- Sera, T. and Uranga, C. (2002). Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table. *Biochemistry* 41: 7074–7081.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7: 539.
- Slattery, M., Zhou, T., Yang, L., Dantas Machado, A.C., Gordan, R., and Rohs, R. (2014). Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* 39: 381–399.
- Stec, I., Nagl, S.B., van Ommen, G.J., and den Dunnen, J.T. (2000). The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation?. *FEBS Lett.* 473: 1–5.
- Tan, L.M., Liu, R., Gu, B.W., Zhang, C.J., Luo, J., Guo, J., Wang, Y., Chen, L., Du, X., Li, S., et al. (2020). Dual recognition of H3K4me3 and DNA by the ISWI component ARID5 regulates the floral transition in Arabidopsis. *Plant Cell* 32: 2178–2195.
- Tidwell, J.A., Schmidt, C., Heaton, P., Wilson, V., and Tucker, P.W. (2011). Characterization of a new ARID family transcription factor (Brightlike/ARID3C) that co-activates Bright/ARID3A-mediated immunoglobulin gene transcription. *Mol. Immunol.* 49: 260–272.
- Trnka, M.J., Pellarin, R., and Robinson, P.J. (2019). Role of integrative structural biology in understanding transcriptional initiation. *Methods* 159–160: 4–22.
- Tu, S., Teng, Y.C., Yuan, C., Wu, Y.T., Chan, M.Y., Cheng, A.N., Lin, P.H., Juan, L.J., and Tsai, M.D. (2008). The ARID domain of the H3K4 demethylase RBP2 binds to a DNA CCGCCC motif. *Nat. Struct. Mol. Biol.* 15: 419–421.
- Tu, X., Wu, J., Xu, Y., and Shi, Y. (2001). ¹H, ¹³C and ¹⁵N resonance assignments and secondary structure of ADR6 DNA-binding domain. *J. Biomol. NMR* 21: 187–188.
- Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Zidek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. *Nature* 596: 590–596.
- UniProt, C. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49: D480–D489.
- van Zundert, G.C.P., Rodrigues, J., Trellet, M., Schmitz, C., Kastiris, P.L., Karaca, E., Melquiond, A.S.J., van Dijk, M., de Vries, S.J., and Bonvin, A. (2016). The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* 428: 720–725.
- Vinogradova, M., Gehling, V.S., Gustafson, A., Arora, S., Tindell, C.A., Wilson, C., Williamson, K.E., Guler, G.D., Gangurde, P., Manieri, W., et al. (2016). An inhibitor of KDM5 demethylases reduces survival of drug-tolerant cancer cells. *Nat. Chem. Biol.* 12: 531–538.
- Vuzman, D. and Levy, Y. (2012). Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Mol. Biosyst.* 8: 47–57.
- Wang, T., Zhang, J., Zhang, X., and Tu, X. (2012). Solution structure of SWI1 AT-rich interaction domain from *Saccharomyces cerevisiae* and its nonspecific binding to DNA. *Proteins* 80: 1911–1917.
- Wang, X., Nagl, N.G., Wilsker, D., Van Scoy, M., Pacchione, S., Yaciuk, P., Dallas, P.B., and Moran, E. (2004). Two related ARID family proteins are alternative subunits of human SWI/SNF complexes. *Biochem. J.* 383: 319–325.
- 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.
- 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.
- Wilsker, D., Probst, L., Wain, H.M., Maltais, L., Tucker, P.W., and Moran, E. (2005). Nomenclature of the ARID family of DNA-binding proteins. *Genomics* 86: 242–251.
- Wolberger, C. (2021). How structural biology transformed studies of transcription regulation. *J. Biol. Chem.* 296: 100741.

- Wu, R.C., Jiang, M., Beaudet, A.L., and Wu, M.Y. (2013). ARID4A and ARID4B regulate male fertility, a functional link to the AR and RB pathways. *Proc. Natl. Acad. Sci. U.S.A.* 110: 4616–4621.
- Xiang, Y., Zhu, Z., Han, G., Ye, X., Xu, B., Peng, Z., Ma, Y., Yu, Y., Lin, H., Chen, A.P., et al. (2007). JARID1B is a histone H3 lysine 4 demethylase up-regulated in prostate cancer. *Proc. Natl. Acad. Sci. U.S.A.* 104: 19226–19231.
- Xiao, R., Chen, J.Y., Liang, Z., Luo, D., Chen, G., Lu, Z.J., Chen, Y., Zhou, B., Li, H., Du, X., et al. (2019). Pervasive chromatin-RNA binding protein interactions enable RNA-based regulation of transcription. *Cell* 178: 107–121.e118.
- Xu, F., Flowers, S., and Moran, E. (2012). Essential role of ARID2 protein-containing SWI/SNF complex in tissue-specific gene expression. *J. Biol. Chem.* 287: 5033–5041.
- Yan, Z., Cui, K., Murray, D.M., Ling, C., Xue, Y., Gerstein, A., Parsons, R., Zhao, K., and Wang, W. (2005). PBAF chromatin-remodeling complex requires a novel specificity subunit, BAF200, to regulate expression of selective interferon-responsive genes. *Genes Dev.* 19: 1662–1667.
- 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.
- Yao, W., Peng, Y., and Lin, D. (2010). The flexible loop L1 of the H3K4 demethylase JARID1B ARID domain has a crucial role in DNA-binding activity. *Biochem. Biophys. Res. Commun.* 396: 323–328.
- Zaman, M.M., Masuda, K., Nyati, K.K., Dubey, P.K., Ripley, B., Wang, K., Chalise, J.P., Higa, M., Hanieh, H., and Kishimoto, T. (2016). Arid5a exacerbates IFN-gamma-mediated septic shock by stabilizing T-bet mRNA. *Proc. Natl. Acad. Sci. U.S.A.* 113: 11543–11548.
- Zhou, W., Melamed, D., Banyai, G., Meyer, C., Tuschl, T., Wickens, M., Cao, J., and Fields, S. (2021). Expanding the binding specificity for RNA recognition by a PUF domain. *Nat. Commun.* 12: 5107.
- Zhu, H., Chen, T., Zhu, M., Fang, Q., Kang, H., Hong, Z., and Zhang, Z. (2008). A novel ARID DNA-binding protein interacts with SymRK and is expressed during early nodule development in *Lotus japonicus*. *Plant Physiol.* 148: 337–347.
- 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.
- Zinzalla, G. (2016). A new way forward in cancer drug discovery: inhibiting the SWI/SNF chromatin remodelling complex. *Chembiochem* 17: 677–682.