Published in final edited form as:

Proteins. 2010 July ; 78(9): 2170–2175. doi:10.1002/prot.22718.

Solution NMR Structure of the ARID Domain of Human AT-rich Interactive Domain-Containing Protein 3A: A Human Cancer Protein Interaction Network Target

Gaohua Liu^{1,*}, Yuanpeng J. Huang¹, Rong Xiao¹, Dongyan Wang¹, Thomas B. Acton¹, and Gaetano T. Montelione^{1,2,*}

¹ Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and Northeast Structural Genomics Consortium (NESG), Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854

² Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854

Abstract

The AT-rich interactive domain (ARID) of human AT-rich interactive domain-containing protein 3A (ARID3A) has been selected for structural characterization by Northeast Structural Genomics Consortium (residues 218-351 NESG ID HR4394C) as part of our Human Cancer Protein Interaction Network (HCPIN) project. Protein ARID3A belongs to the ARID family DNA-binding protein and is known to play important roles in embryonic patterning, cell lineage gene regulation, and cell cycle control, chromatin remodeling and transcriptional regulations. The solution NMR structure of ARID3A ARID domain consists of eight α -helices $\alpha 0$ - $\alpha 7$ and a short β hairpin. Helix $\alpha 0$ and $\alpha 1$ form a V shape, helix $\alpha 2$ - $\alpha 4$ and helix $\alpha 5$ - $\alpha 7$ form two U shapes, and the V and two U shapes packed orthogonal to each other. The NMR structure of the ARID domain of human ARID3A reported here provides a structural basis for elucidating the regulatory mechanisms of ARID3A function, and the molecular mechanism of ARID3A interactions with DNA. It also has potential value in future drug discovery and design.

Keywords

ARID3A; Bright; Drill1; E2FBP1; Transcription factor; DNA-binding protein

INTRODUCTION

The Northeast Structural Genomics Consortium has constructed a Human Cancer Protein Interaction Network (HCPIN), providing structure-function annotations of key proteins associated with human cancer and developmental biology.¹ The long-range goal of the HCPIN project is to provide a comprehensive 3D structure-function database of human-cancer-associated proteins and protein complexes in the context of functional networks, using both experimental structures and high quality homology models (e.g. using protein templates with > 80% sequence identity).¹ The human AT-rich interactive domain-containing protein 3A (ARID3A) is one of these HCPIN proteins. As a part of this large-scale human cancer biology project, the AT-rich interactive domain (ARID) of human

*Correspondence to: Gaohua Liu, CABM Rutgers University, 679 Hoes Lane, Piscataway, NJ 08854, gliu@cabm.rutgers.edu or Gaetano T. Montelione, CABM Rutgers University, 679 Hoes Lane, Piscataway, NJ 08854, guy@cabm.rutgers.edu, Phone: (732) 235-5321, Fax: (732) 235-5779.

ARID3A, residues 218-351 (NESG ID HR4394C) has been selected for structural characterization.

ARID3A, an ARID family transcription factor, also called dead ringer-like protein 1 (Dril1), B-cell regulator of IgH transcription (Bright), and E2F-binding protein 1 (E2FBP1), belongs to the ARID family of DNA-binding proteins known to play important roles in embryonic patterning, cell lineage gene regulation, cell cycle control, chromatin remodeling and transcriptional regulation.²⁻⁵ ARID proteins have been identified in all sequenced higher eukaryotic genomes; the consensus sequence of the ARID domain spans about 100 amino acid residues.³

ARID proteins are partitioned into three structural classes: (i) minimal ARID proteins that consist of a core domain formed by six α -helices; (ii) ARID proteins that supplement the core domain with an N-terminal α -helix; and (iii) extended-ARID proteins, which contain the core domain and additional α -helices at their N- and C-termini.⁶ The 15 distinct human ARID family proteins can be divided into seven subfamilies based on the degree of sequence identity between individual members.³ The majority of ARID subfamilies (i.e. five out of seven) bind DNA without obvious sequence preference, though DNA-binding affinity varies somewhat between subfamilies. Structural studies have identified the DNA major groove contact site as a modified helix-turn-helix motif.^{3,4} The third mammalian ARID subfamily, ARID3, contains three human proteins (ARID3A, ARID3B and ARID3C). These paralogs are the most direct mammalian counterparts of the drosophila “dead ringer” protein Dri.³ Like drosophila Dri, these human ARID3 proteins have both N- and C- terminal extensions beyond the consensus ARID core sequence, which does not occur in other human ARID family members.^{3,4}

ARID domains generally function as dsDNA-binding domains. The ARID domain of human ARID3A, the subject of this study, has ~ 78% sequence identity with drosophila “dead ringer” protein Dri, an essential ARID-containing transcription factor. ARID3A is also an ARID family transcription factor, ubiquitously expressed in all tissues examined.⁷ It is a direct p53 target gene.² It can immortalize primary mouse fibroblasts, bypass RAS^{V12}-induced cellular senescence, and collaborates with RAS^{V12} or MYC in mediating oncogenic transformation.^{2,5} ARID3A also activates immunoglobulin heavy chain transcription and engages in heterodimer formation with E2F to stimulate E2F-dependent transcription.^{5,8} While the regulatory mechanisms of ARID3A function remain largely unknown, SUMO (Small Ubiquitin-related Modifier) modification of ARID3A was recently revealed to play crucial roles in ARID3A transcriptional activity.⁵ ARID3A was found to interact with the SUMO-conjugating enzyme Ubc9, and is sumoylated both in *vitro* and in *vivo* at Lys³⁹⁸. PIASy, a member of the Protein Inhibitor of Activated STAT (PIAS) family, functions as a specific SUMO E3-ligase for ARID3A, and promotes its sumoylation both in *vitro* and in *vivo*.⁵

The ARID DNA-binding domain of ARID3A was selected as a NESG target as part of our HCPIN human cancer biology theme project.¹ The NMR structure of the ARID domain of human ARID3A reported here provides a structural basis for elucidating the regulatory mechanisms of ARID3A function, and the molecular mechanism of ARID3A interactions with DNA. It also has potential value in future drug discovery and design.

MATERIALS AND METHODS

The ARID domain of ARID3A from *Homo sapiens* (UniProtKB/Swiss-Prot ID Q99856/ARI3A_HUMAN, residues 218-351) was cloned, expressed and purified following standard, largely-automated NESG protocols, to produce a uniformly ¹³C, ¹⁵N-enriched

protein sample.⁹ Briefly, the truncated ARI3A_HUMAN (218-351) gene was cloned into a modified pET14-15C (Novagen), as described in reference⁹, yielding the plasmid pHR4394C-14.6. The resulting construct contains 11 nonnative residues at the N-terminus (MGHHHHHHSHM) that facilitate protein purification. *Escherichia coli* BL21 (DE3) pMGK cells were transformed with pHR4394C-14.6, and cultured in MJ9 minimal medium¹⁰ containing (U -¹⁵NH₄)₂SO₄ and U -¹³C-glucose as sole nitrogen and carbon sources, respectively. U -¹³C, ¹⁵N ARID3A was purified using an ÄKTExpress™ (GE Healthcare) two-step protocol consisting of IMAC (HisTrap HP) and gel filtration (HiLoad 26/60 Superdex 75) chromatographies. The final yield of purified U -¹³C, ¹⁵N ARID3A (>98% homogeneous by SDS-PAGE; 18.2 kDa by MALDI-TOF mass spectrometry) was ~38 mg/L. In addition, a U -¹⁵N and 5% ¹³C-enriched sample was generated for stereo-specific assignment of isopropyl methyl groups. Both U -¹³C, ¹⁵N and 5% ¹³C, U -¹⁵N ARID3A were dissolved at concentrations of ~0.9 mM in 95% H₂O/5% ²H₂O buffer containing 20 mM MES, 200 mM NaCl, 10 mM DTT, 5 mM CaCl₂, and 0.02% NaN₃, at pH 6.5. Static light scattering data⁹ and rotational correlation time measurements (Supplementary Figures S3, S4) indicate that the protein is monomeric in solution under the conditions used for these NMR studies.

All NMR spectra were recorded at 25 °C, using cryogenic NMR probes. Triple resonance NMR data were collected on Varian INOVA 600 MHz spectrometer, while simultaneous 3D ¹⁵N/¹³C_{aliphatic}/¹³C_{aromatic}-edited NOESY¹¹ (mixing time 100 ms) in H₂O and 3D ¹³C-edited NOESY (mixing time 100 ms) in ²H₂O spectra were acquired on a Bruker AVANCE 800 MHz spectrometer. 2D constant-time [¹³C, ¹H]-HSQC spectra, with 28 ms and 42 ms constant-time delays, were recorded for the 5% ¹³C-enriched sample on the Varian INOVA 600 MHz spectrometer in order to obtain stereo-specific assignments for isopropyl groups of valines and leucines.¹² All NMR data were processed using the program NMRPipe¹³ and analyzed using the program XEASY.¹⁴ Spectra were referenced to external DSS. Sequence-specific resonance assignments were determined as described previously.¹⁵ Backbone assignments ($H^N/N/C'$, H^α/C^α , and H^β/C^β) were obtained in largely automated fashion with the program AUTOASSIGN.¹⁶ These assignments, together with random coil sidechain chemical shift values, were then used to simulate peak lists which facilitated manual analysis of side-chain resonance assignments. Simultaneous 3D ¹⁵N/¹³C_{aliphatic}/¹³C_{aromatic}-NOESY and CCH-TOCSY were then analyzed manually to obtain nearly complete side-chain assignments. Assignments were obtained for 93% of backbone and side-chain chemical shifts assignable with the NMR experiments listed above (excluding N-terminal NH₃⁺, Lys NH₃⁺, Arg NH₂, OH of Ser, Thr and Tyr, ¹³C $^\gamma$ of Asp and Asn, ¹³C $^\delta$ of Glu and Gln, and aromatic ¹³C $^\gamma$ shifts, Supplementary Table S1). Chemical shifts were deposited in the BioMagResBank on 06/14/2009 with accession code 16348.

The locations of regular secondary structure elements were next identified based on backbone chemical shift data.¹⁷ A NOESY peak list containing expected intra-residue, sequential, and α -helical medium-range NOE peaks was initially generated and then manually edited by visual inspection of the simultaneous-NOESY spectrum. Subsequent manual peak picking was then used to identify remaining, primarily long-range NOEs.¹⁵ The programs CYANA^{18,19} and AUTOSTRUCTURE^{20,21} were used in parallel to automatically assign long-range NOEs. Assignments identically obtained by both programs ('consensus assignments')¹⁵ were retained and established the starting point for iterative cycles of noise/artifact peak removal, peak picking and NOE assignment. Statistical summaries of ¹H - ¹H upper-bound distance constraints used for structure calculations are summarized in Table I. In addition, backbone dihedral angle constraints were derived from chemical shifts using the program TALOS+²² for residues located in well-defined secondary structure elements (Table I). No pre-determined distance constraint was used throughout the structure determination. The final structure calculation was performed with CYANA 3.0,

and the 20 conformers with the lowest target function value were refined in an 'explicit water bath'²³ using the program CNS.²⁴ The coordinates were deposited in the Protein Data Bank on 06/14/2009 (accession code 2KK0). Structural statistics and global structure quality factors including Verify3D,²⁵ ProsaII,²⁶ PROCHECK,²⁷ and MolProbity²⁸ raw and statistical Z-scores were computed using the PSVS 1.3 software package.²⁹ The global goodness-of-fit of the final structure ensembles with the NOESY peak list data were determined using the RPF analysis program.³⁰ All structure figures were made using MOLMOL³¹ or PyMOL 1.1.³²

RESULTS AND DISCUSSIONS

The solution NMR structure of ARID3A ARID domain consists of eight α -helices $\alpha 0$ - $\alpha 7$ (residues 231-234, 239-254, 272-282, 285-291, 294-300, 310-321, 324-329, 335-346) and a short β hairpin consisting of anti-parallel strands $\beta 1$ and $\beta 2$ (residues 264-265, 268-269) (Fig. 1a & 1b). Helices $\alpha 0$ and $\alpha 1$ form a V shape, helices $\alpha 2$ - $\alpha 4$ and helices $\alpha 5$ - $\alpha 7$ form two U shapes, and the V and two U shapes pack orthogonal to each other for form the protein fold. Structural statistics for this largely helical domain are given in Table I; this solution NMR structure exhibits good quality assessment scores. The ARID core structure, including helices $\alpha 1$ - $\alpha 6$ and the short β hairpin, is well defined. Backbone amides of residues Asn²⁶⁰, Val²⁶⁹, Arg²⁹⁵ and N-terminal residues 224-230 were not observed in [¹⁵N-¹H]-HSQC spectra due to line broadening; these apparently exchange-broadened residues are located in the same region of the ARID core structure to which the extended N-terminal helix $\alpha 0$ is anchored. The N-terminal region, from residues 218-230, is flexible in solution. In contrast, helix $\alpha 7$, formed by the C-terminal extended sequence characteristic of the ARID3 subfamily, is well defined and packs against the ARID core. Most residues buried in the helical bundle are hydrophobic, and the structure is mainly stabilized by a hydrophobic core that is well defined by the extensive inter-helical NOE interaction network.

Some 3D structures of ARID family members are available in PDB, both with and without bound dsDNA.^{6,33} A search for structurally-similar proteins in the PDB using the DALI³⁴ server revealed significant structural similarity between ARID3A and these other ARID domains, including Dri ARID from *Drosophila melanogaster* in the presence of dsDNA³³ (Fig. 1c, PDB code 1kqj with DALI Z score 17.6, RMSD 2.1 Å, 79% SeqID), Dri ARID without dsDNA⁶ (Fig. 1c, PDB code 1c20 with DALI Z score 12.3, RMSD 3.1 Å, 79% SeqID), ARID1B (PDB code 2eh9 with DALI Z score 12.6, RMSD 2.0 Å, SeqID 43%), the ARID domain from the human JARD1C protein³⁵ (PDB code 2jrz with DALI Z score 11.5, RMSD 2.5 Å, SeqID 25%), and the ARID domain from the histone H3K4 demethylase RBP2³⁶ (PDB code 2jxj with DALI Z score 9.8, RMSD 2.4 Å, SeqID 32%).

The ARID domain structures of ARID3A and Dri have similar global folds, as expected from their high sequence similarity (79% SeqID, Supplementary Figure S5), and they also have significant structural differences (as described below). ARID3A has a large basic charged surface distribution on one face that may facilitate recognition of the negatively charged DNA duplex (Fig. 1d), and a mainly negative charged distribution on the opposite side of the molecule. Similar surface electrostatic distributions are observed for the drosophila Dri.³³ The complex structure of Dri bound to dsDNA³³ shows that side-chains of 16 residues of Dri have close contact with dsDNA, including (i) the modified helix-turn-helix motif ($\alpha 4$ -Loop2- $\alpha 5$ region) that directly contacts the DNA major groove, and (ii) the β hairpin loop L1 and C-terminal extended helix $\alpha 7$ regions that have contacts with the DNA regions outside the major groove.^{3,33} All of these residues are solvent exposed, and 15 out of the 16 residues are identical between ARID3A and Dri (Supplementary Figure S5); hence ARID3A likely binds to dsDNA using a similar mechanism as Dri.

Although ARIDs of ARID3A and Dri have similar overall structures, there are also some structural differences between these structures. Interestingly, the ARID3A structure is more similar to the Dri ARID domain in the presence of dsDNA than to the structure of Dri determined in the absence of dsDNA, especially the β -hairpin loop L1 region. The backbone RMSD values to dsDNA-bound and unbound Dri are 2.1 Å and 3.1 Å, respectively (Fig. 1c).⁶ In the ARID domain of ARID3A, the N-terminal extension sequence is three residues shorter than that of *Drosophila* Dri; as the resonances of several residues in this N-terminal segment are too weak to be observed, presumably due to exchange broadening, we cannot accurately characterize the length of helix $\alpha 0$ in this human ARID domain. These exchange-broadened residues are helical in a CS_Rosetta³⁷ structure of this same construct (unpublished results). The two loops, L1 between $\beta 1$ and $\beta 2$, and L2 between $\alpha 4$ and $\alpha 5$, in ARID3A are better defined in the solution NMR structure of ARID3 than those in the unbound Dri, loop L2 is actually disordered in the unbound Dri.⁶ As RMSD values across an ensemble of NMR structures should only be used qualitatively in estimating the precision and/or flexibility of a NMR structure,³⁸ the higher degree of order in loops L1 and L2 of the human ARID domain compared with Dri was carefully investigated, and is supported by extensive NOESY and heteronuclear NOE data summarized in Supplementary Figure S2. Besides unobserved backbone amides of Asn²⁶⁰ and Val²⁶⁹ that located at where the N-terminal helix $\alpha 0$ is anchored, the backbone amide resonances of nearby residues Ile²⁶², Met²⁶⁵, Ala²⁶⁶, and Leu²⁷⁰ are also slightly broadened, suggesting there may be some intermediate conformational exchange of loop L1 or exchange between an ordered loop L1 contacting and not contacting the helix $\alpha 0$. Such exchanges can result in inconsistent distance constraints due to NOEs arising from multiple members of the conformational ensemble; satisfying these constraints can result in a tighter bundle than one might expect for a loop with modest flexibility.³⁸ Though this loop is much more ordered in ARID3A ARID domain than in the Dri domain in the absence of DNA, we cannot rule out the possibility that this DNA-binding loop has modest conformational flexibility in the ARID domain structure. The 3D structure of ARID3A shows that side-chains of Met²⁶⁵ and Ala²⁶⁶ are packed into a hydrophobic pocket formed by the side-chains of Ala³¹⁰, Thr³¹³ and Leu³¹⁴ from helix $\alpha 5$ and Leu³⁰³, Pro³⁰⁴ and Ile³⁰⁷ from loop L2. This is supported by observed NOE distance constraints (Supplementary Table S2 and Figure S6). This feature is also observed in the Dri structure in the presence of dsDNA³³, but not in Dri in the absence of dsDNA.⁶

It has been suggested that dsDNA binding drives the rotation of the N-terminal sub-domain (comprised of helices $\alpha 0$, $\alpha 1$ and the β hairpin) relative to the remainder of the polypeptide.³³ However, the similarity between human ARID3A ARID domain in the absence of dsDNA and the Dri ARID domain bound to DNA indicates less conformational rearrangement upon DNA-binding by human ARID3A than is observed for the *Drosophila* “dead ringer” Dri protein, and/or that the equilibrium distribution between apo and dsDNA-binding forms of the domain are different in these human and *Drosophila* ARID domain constructs. Interestingly, of the 16 residues that contact the DNA in the Dri-dsDNA complex, only one is different between the Dri and human ARID3A; i.e. Gln³³⁸ in Dri is replaced by basic Arg²⁹⁵ in human ARID3A, which may affect the dsDNA binding affinity. Comparisons of DNA-binding by human ARID3A ARID and Dri ARID domains may thus provide useful insights into the functional role of conformational rearrangements in the process of dsDNA recognition by this important class of eukaryotic transcription factors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by a grant from the Protein Structure Initiative of National Institute of General Medical Sciences, U54-GM074958.

References

- Huang YJ, Hang D, Lu LJ, Tong L, Gerstein MB, Montelione GT. Targeting the human cancer pathway protein interaction network by structural genomics. *Mol Cell Proteomics*. 2008; 7(10):2048–2060. [PubMed: 18487680]
- Peeper DS, Shvarts A, Brummelkamp T, Douma S, Koh EY, Daley GQ, Bernards R. A functional screen identifies hDRIL1 as an oncogene that rescues RAS-induced senescence. *Nat Cell Biol*. 2002; 4(2):148–153. [PubMed: 11812999]
- Wilsker D, Probst L, Wain HM, Maltais L, Tucker PW, Moran E. Nomenclature of the ARID family of DNA-binding proteins. *Genomics*. 2005; 86(2):242–251. [PubMed: 15922553]
- Patsialou A, Wilsker D, Moran E. DNA-binding properties of ARID family proteins. *Nucleic Acids Res*. 2005; 33(1):66–80. [PubMed: 15640446]
- Prieur A, Nacerddine K, van Lohuizen M, Peeper DS. SUMOylation of DRIL1 directs its transcriptional activity towards leukocyte lineage-specific genes. *PLoS One*. 2009; 4(5):e5542. [PubMed: 19436740]
- Iwahara J, Clubb RT. Solution structure of the DNA binding domain from Dead ringer, a sequence-specific AT-rich interaction domain (ARID). *EMBO J*. 1999; 18(21):6084–6094. [PubMed: 10545119]
- Ma K, Araki K, Ichwan SJ, Suganuma T, Tamamori-Adachi M, Ikeda MA. E2FBP1/DRIL1, an AT-rich interaction domain-family transcription factor, is regulated by p53. *Mol Cancer Res*. 2003; 1(6):438–444. [PubMed: 12692263]
- Lin D, Ippolito GC, Zong RT, Bryant J, Koslovsky J, Tucker P. Bright/ARID3A contributes to chromatin accessibility of the immunoglobulin heavy chain enhancer. *Mol Cancer*. 2007; 6:23. [PubMed: 17386101]
- Acton TB, Gunsalus KC, Xiao R, Ma LC, Aramini J, Baran MC, Chiang YW, Climent T, Cooper B, Denissova NG, Douglas SM, Everett JK, Ho CK, Macapagal D, Rajan PK, Shastry R, Shih LY, Swapna GVT, Wilson M, Wu M, Gerstein M, Inouye M, Hunt JF, Montelione GT. Robotic cloning and Protein Production Platform of the Northeast Structural Genomics Consortium. *Method Enzymol*. 2005; 394:210–243.
- Jansson M, Li YC, Jendeborg L, Anderson S, Montelione GT, Nilsson B. High-level production of uniformly N-15- and C-13-enriched fusion proteins in *Escherichia coli*. *Journal of Biomolecular Nmr*. 1996; 7(2):131–141. [PubMed: 8616269]
- Shen Y, Atreya HS, Liu GH, Szyperski T. G-matrix Fourier transform NOESY-based protocol for high-quality protein structure determination. *Journal of the American Chemical Society*. 2005; 127(25):9085–9099. [PubMed: 15969587]
- Neri D, Szyperski T, Otting G, Senn H, Wuthrich K. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ¹³C labeling. *Biochemistry*. 1989; 28(19):7510–7516. [PubMed: 2692701]
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *Journal of Biomolecular Nmr*. 1995; 6(3):277–293. [PubMed: 8520220]
- Bartels C, Xia TH, Billeter M, Guntert P, Wuthrich K. The Program Xeasy for Computer-Supported Nmr Spectral-Analysis of Biological Macromolecules. *Journal of Biomolecular Nmr*. 1995; 6(1):1–10.
- Liu GH, Shen Y, Atreya HS, Parish D, Shao Y, Sukumaran DK, Xiao R, Yee A, Lemak A, Bhattacharya A, Acton TA, Arrowsmith CH, Montelione GT, Szyperski T. NMR data collection and analysis protocol for high-throughput protein structure determination. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102(30):10487–10492. [PubMed: 16027363]

16. Moseley HNB, Monleon D, Montelione GT. Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. *Nuclear Magnetic Resonance of Biological Macromolecules*, Pt B. 2001; 339:91–108.
17. Wishart DS, Sykes BD. The C-13 Chemical-Shift Index - a Simple Method for the Identification of Protein Secondary Structure Using C-13 Chemical-Shift Data. *Journal of Biomolecular Nmr*. 1994; 4(2):171–180. [PubMed: 8019132]
18. Guntert P, Mumenthaler C, Wuthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *Journal of Molecular Biology*. 1997; 273(1):283–298. [PubMed: 9367762]
19. Herrmann T, Guntert P, Wuthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *Journal of Molecular Biology*. 2002; 319(1):209–227. [PubMed: 12051947]
20. Huang YJ, Tejero R, Powers R, Montelione GT. A topology-constrained distance network algorithm for protein structure determination from NOESY data. *Proteins*. 2006; 62(3):587–603. [PubMed: 16374783]
21. Huang YPJ, Moseley HNB, Baran MC, Arrowsmith C, Powers R, Tejero R, Szyperski T, Montelione GT. An integrated platform for automated analysis of protein NMR structures. *Nuclear Magnetic Resonance of Biological Macromolecules*, Part C. 2005; 394:111.
22. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR*. 2009
23. Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M. Refinement of protein structures in explicit solvent. *Proteins*. 2003; 50(3):496–506. [PubMed: 12557191]
24. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D-Biological Crystallography*. 1998; 54:905–921.
25. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. *Nature*. 1992; 356(6364):83–85. [PubMed: 1538787]
26. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. *Proteins*. 1993; 17(4):355–362. [PubMed: 8108378]
27. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography*. 1993; 26:283–291.
28. Word JM, Bateman RC, Presley BK, Lovell SC, Richardson DC. Exploring steric constraints on protein mutations using MAGE/PROBE. *Protein Science*. 2000; 9(11):2251–2259. [PubMed: 11152136]
29. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. *Proteins*. 2007; 66(4):778–795. [PubMed: 17186527]
30. Huang YJ, Powers R, Montelione GT. Protein NMR recall, precision, and F-measure scores (RPF scores): Structure quality assessment measures based on information retrieval statistics. *Journal of the American Chemical Society*. 2005; 127(6):1665–1674. [PubMed: 15701001]
31. Koradi R, Billeter M, Wuthrich K. MOLMOL: A program for display and analysis of macromolecular structures. *Journal of Molecular Graphics*. 1996; 14(1):51. [PubMed: 8744573]
32. DeLano, WL. The pymol manual. San Carlos, CA: DeLano Scientific; 2002.
33. Iwahara J, Iwahara M, Daughdrill GW, Ford J, Clubb RT. The structure of the Dead ringer-DNA complex reveals how AT-rich interaction domains (ARIDs) recognize DNA. *EMBO J*. 2002; 21(5):1197–1209. [PubMed: 11867548]
34. Holm L, Sander C. Dali: a network tool for protein structure comparison. *Trends Biochem Sci*. 1995; 20(11):478–480. [PubMed: 8578593]
35. Koehler C, Bishop S, Dowler EF, Schmieder P, Diehl A, Oschkinat H, Ball LJ. Backbone and sidechain 1H, 13C and 15N resonance assignments of the Bright/ARID domain from the human JARID1C (SMCX) protein. *Biomol NMR Assign*. 2008; 2(1):9–11. [PubMed: 19636912]

36. Tu S, Teng YC, Yuan C, Wu YT, Chan MY, Cheng AN, Lin PH, Juan LJ, Tsai MD. The ARID domain of the H3K4 demethylase RBP2 binds to a DNA CCGCCC motif. *Nat Struct Mol Biol.* 2008; 15(4):419–421. [PubMed: 18270511]
37. Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu GH, Eletsky A, Wu YB, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A. Consistent blind protein structure generation from NMR chemical shift data. *Proceedings of the National Academy of Sciences of the United States of America.* 2008; 105(12):4685–4690. [PubMed: 18326625]
38. Snyder DA, Bhattacharya A, Huang YPJ, Montelione GT. Assessing precision and accuracy of protein structures derived from NMR data. *Proteins-Structure Function and Bioinformatics.* 2005; 59(4):655–661.

Fig. 1. NMR structures of ARID3A. **(a)** Backbone trace of residues 218-351 of 20 representative CNS conformers of ARID3A, after superposition of backbone N, C α and C' atoms of the regular secondary structure elements for minimal root-mean-square deviation (RMSD). The N- and C-termini are labeled as "N" and "C". **(b)** Ribbon drawing of residues 218-351 of the ARID3A conformer with the lowest CNS energy. α -helices $\alpha 0$ - $\alpha 7$, β strands $\beta 1$ - $\beta 2$, and loops L1-L2 are labeled and colored; the N- and C-termini are labeled as "N" and "C". **(c)** Comparison of the ARID domain of human ARID3A (red) with drosophila "dead ringer" Dri in the presence of dsDNA (blue, PDB code 1kqq) and in the absence of dsDNA (green, PDB code 1c20) after superposition of backbone N, C α and C' atoms of regular secondary structure elements for minimal root-mean-square deviation (RMSD). Side-chains of the 16 residues in ARID3A corresponding to those involved in DNA interactions in the Dri-DNA complex are colored red **(d)** Surface representation of the ARID3A ARID domain conformer with the lowest CNS energy. The structure shown on the right is rotated by 180° about the vertical axis. Surface colors represent the electrostatic potential. All figures were prepared with the programs MOLMOL³¹ or PYMOL³².

Table I**NMR Structure Quality Statistics for ARID3A ARID Domain**

Conformationally restricting distance constraints^a	
Intra-residue [$i = j$]	767
Sequential [$ i-j = 1$]	1028
Medium range [$1 < i-j < 5$]	1211
Long range [$ i-j \geq 5$]	1239
Total	4245
Hydrogen bond constraints	84
Dihedral angle constraints	
ϕ	92
ψ	93
Number of constraints per residue ^b	34.2
Number of long range constraints per residue ^b	9.4
CYANA target function [$\text{\AA}^2$]	1.43 ± 0.17
Average r.m.s.d. to the mean CNS coordinates [$\text{\AA}$]	
regular secondary structure elements ^c backbone heavy atoms N, C $^\alpha$, C $^\gamma$	0.42 ± 0.05
regular secondary structure, ^c all heavy atoms	0.76 ± 0.08
Residues 230-348, backbone heavy atoms N, C $^\alpha$, C $^\gamma$	0.51 ± 0.05
Residues 230-348, all heavy atoms	0.87 ± 0.09
RPF R/P/F/DP ³¹	95/95/95/82
Ramachandran plot summary for all ordered (by PSVS) ^d	
most favorable regions	89.70
additionally allowed regions	10.3
generously allowed regions	0.0
disallowed regions	0.0
Global quality scores (raw/Z - score) ^d	
VERIFY3D ²⁵	0.42/−0.64
ProsaII ²⁶	0.74/0.37
PROCHECK G-factor ϕ and ψ ²⁷	0.05/0.51
PROCHECK G-factor all dihedral angles	0.00/0.00
MOLPROBITY ²⁸ Clash Score	17.98/−1.56

^aAnalysed for all residues including HisTag.

^bThere are 132 residues with conformationally restricting constraints,

^cRegular secondary elements: alpha helices 231-234, 239-254, 272-282, 285-291, 294-300, 310-321, 324-329, 335-346; beta strands: 264-265, 268-269.

^dOrder residues 230 - 307, 310 - 329, 332 - 347, defined based on dihedral angle order parameters $S(\phi)$ and $S(\psi) > 0.90$. Z-scores were computed relative to corresponding structure quality measures for high resolution X-ray crystal structures²⁹.