The ARID domain of the H3K4 demethylase RBP2 binds to a DNA CCGCCC motif

Shengjiang Tu[1], Yu-Ching Teng[2,3], Chunhua Yuan[4], Ying-Ta Wu[2], Meng-Yu Chan[2], An-Ning Cheng[2], Po-Hsun Lin[2], Li-Jung Juan[2] & Ming-Daw Tsai[1,2,4,5]

The histone H3 lysine 4 demethylase RBP2 contains a DNA binding domain, the AT-rich interaction domain (ARID). We solved the structure of ARID by NMR, identified its DNA binding motif (CCGCCC) and characterized the binding contacts. Immunofluorescence and luciferase assays indicated that ARID is required for RBP2 demethylase activity in cells and that DNA recognition is essential to regulate transcription.

The human tumor suppressor retinoblastoma protein (RB)–interacting protein RBP2 belongs to the JARID1 family of histone H3 lysine 4 (H3K4) demethylases[1,2]. RBP2 participates in cell differentiation and is a candidate trxG protein that regulates homeotic gene expression. RBP2 regulates transcription in conjunction with DNA binding transcription factors and chromatin-associated complexes[3–5]. Importantly, RBP2 also contains an intrinsic DNA binding domain, ARID. ARID-containing proteins participate in various regulatory processes, including embryonic development and tissue-specific gene regulation[6]. Although the founding members—Bright, Dead ringer (Dri) and Mrf-2—bind to AT-rich motifs[7–9], some ARID proteins, such as p270 (also known as SWI1) and Osa, show no sequence preference[10,11]. Interestingly, the ARID domains of Dri, Mrf-2 and p270 all form a similar tertiary fold consisting mainly of helices[12–14]. Furthermore, systematic binding studies of ARID using digested λ-phage DNA suggest that most ARID proteins, including RBP2, bind to DNA nonspecifically[15]. In contrast, PLU-1, another ARID-containing demethylase in the JARID1 family, binds to GC-rich DNA[16]. Thus, the mechanism for diverse DNA recognition by ARID has not been elucidated. Furthermore, the question of whether RBP2 targets specific genes through direct DNA binding also remains unanswered.

To address *in vivo* functions of RBP2 ARID, an ARID-deleted mutant was constructed and examined for its effect on RBP2's demethylase activity. As expected, ectopic expression of full-length RBP2 reduced the cellular levels of di- and trimethylated H3K4 (H3K4me2 and H3K4me3, respectively), but not trimethylated histone H3 lysine 9 (H3K9me3), as seen by immunostaining (**Fig. 1** and data not shown). In contrast, expression of the ARID-deleted mutant did not result in reduced H3K4me3 levels. Similar results were obtained using an RBP2 construct that lacks both the Jumonji N (JmjN) and the ARID domains (data not shown). The data indicate that, as for SMCY (also known as JARID1D)[17] and PLU-1 (ref. 18), intact ARID is necessary for RBP2's demethylase activity.

Next we examined whether RBP2 ARID binds DNA in a sequence-specific manner. A PCR-based selection and amplification method (SELEX, **Supplementary Methods** online) was used to enrich sequences from a pool of random DNA oligomers bound by GST-fused ARID (residues 85–175) of RBP2. In total, 47 sequences were obtained (**Supplementary Table 1** online), with an average of 62% GC content. This is in sharp contrast with preferred binding sequences of Bright, which contain >70% AT[7]. Sequence alignment identified CCGCCC as the most abundant motif. Notably, CCGCCC and close variants with one deviation are enriched in the region 500 base pairs (bp) upstream of the transcription start site of many promoters bound by RBP2, including human genes bromodomain-containing 2 and 8 (*BRD2* and *BRD8*) and *BGLAP* (encoding osteocalcin), and mouse genes *Hoxa1*, *Hoxa5*, *Hoxa7*, *Hoxa11* and *Cxcl12* (also known as *Sdf1*; refs. 1–3 and **Supplementary Table 2** online). For the *Hoxa* and *BRD2*

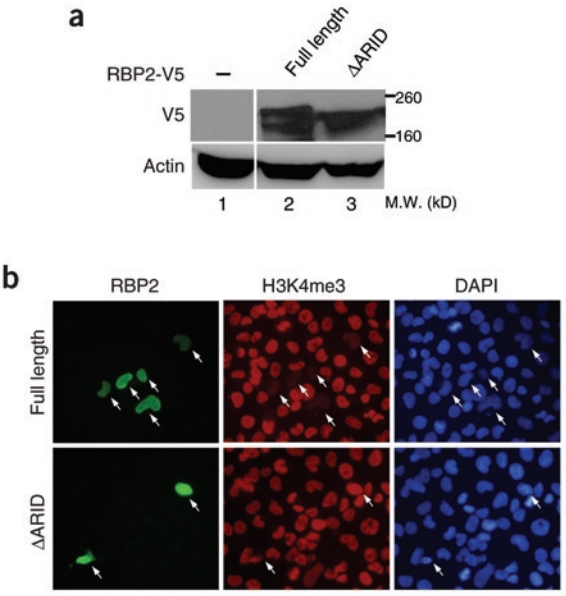

Figure 1 ARID is required for RBP2 demethylase activity. H1299 cells were transfected with a plasmid encoding V5-tagged full-length RBP2 or an RBP2 mutant in which the ARID domain was deleted. (**a**) Western blot of cell lysates using an antibody against V5 (Invitrogen), with actin as a loading control. (**b**) Double-immunostaining assays using antibodies against V5 (for RBP2, left) and H3K4me3 (Abcam; middle). Nuclei were stained with DAPI (right).

[1]Department of Chemistry, Ohio State University, 100 West 18th Ave., Columbus, Ohio 43210, USA. [2]Genomics Research Center, Academia Sinica, 128 Academia Rd. Sec. 2, Nankang, Taipei 115, Taiwan. [3]Institute of Biochemistry and Molecular Biology, National Yang-Ming University, 115, Sec. 2, Linong St., Beitou, Taipei 112, Taiwan. [4]Campus Chemical Instrument Center, 176 West 19th Ave., Ohio State University, Columbus, Ohio 43210, USA. [5]Institute of Biological Chemistry, Academia Sinica, 128 Academia Rd., Sec. 2, Nankang, Taipei 115, Taiwan. Correspondence should be addressed to M.-D.T. (mdtsai@gate.sinica.edu.tw) or L.-J.J. (ljjuan@gate.sinica.edu.tw).

Received 20 November 2007; accepted 6 February 2008; published online 12 February 2008; doi:10.1038/nsmb.1400

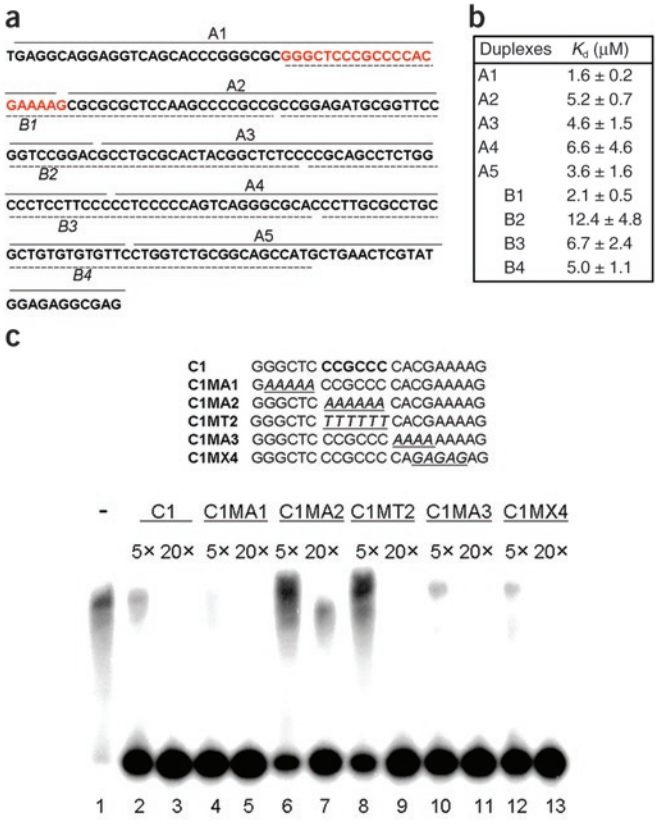

Figure 2 RBP2 ARID specifically binds to the CCGCCC motif in the *BRD2* promoter. (**a**) Duplexes covering the *BRD2* promoter region co-immunoprecipitated with RBP2 (–246 bp to –20 bp relative to the transcription start site)[3]. The 'A' and 'B' series are labeled with solid and dash lines, respectively. The sequences shared by A1 and B1 are shown in red and designated as C1. (**b**) Dissociation constants (K_d) of each duplex, as determined by EMSA, with s.d. reported. (**c**) Cold competition. The intact core CCGCCC motif in C1 is highlighted in bold. The mutated bases are underlined. Unlabeled competitors were added at 5-fold or 20-fold concentrations to a fixed amount of labeled C1-ARID complex.

genes, at least four CCGCCC-related motifs occur within the proximal promoter region. We asked whether RBP2 ARID selectively binds to the *BRD2* promoter via the CCGCCC motifs. By electrophoretic mobility shift assays (EMSA), we determined the dissociation constants (K_d) of two series of consecutive duplexes covering the entire region of the *BRD2* promoter known to be co-immunoprecipitated by an RBP2 antibody[3] (**Fig. 2a**,**b**). The tightest-binding fragments (A1 and B1) shared a CCGCCC motif. To further determine the binding specificity, the CCGCCC-containing sequence C1 was radioactively labeled and subjected to cold competition in EMSA. The C1-ARID complex was disrupted by duplexes containing intact CCGCCC, but not by those with the motif mutated to A or T, confirming that RBP2 ARID preferentially binds to C1 through CCGCCC (**Fig. 2c**). Additional competition experiments suggested that mutations of three or more bases within the CCGCCC motif are sufficient to disrupt binding (**Supplementary Fig. 1** online), whereas single-base mutations had minimal effects on binding to the RBP2 ARID domain (data not shown). Taken together, the DNA binding results support the hypothesis that RBP2 ARID selectively binds DNA through the motif CCGCCC.

To understand the structural basis of GC preference by RBP2 ARID, we solved its solution structure by NMR (**Supplementary Fig. 2a**,**b** and **Supplementary Table 3** online). The tertiary fold obtained is similar to the core structure of the other ARID domains determined so far[12–14]. Briefly, the domain comprises six helices (H1–H6) and two loops (L1 and L2), with three of the helices (H2, H3 and H4) forming a salient U shape. NMR HSQC chemical shift perturbation upon DNA binding indicates that the binding interface is likely to include L1 and a ubiquitous helix-turn-helix (HTH) DNA binding motif formed by H4, L2 and H5 (**Fig. 3a** and **Supplementary Fig. 3** online). Of these regions, residues around the L2–H5 joint showed the largest chemical shift changes. Overall, the binding features are similar to those of Dri and Mrf-2, which recognize specific sequences[12,19]. In contrast, p270, which binds DNA nonspecifically, does not show chemical shift perturbation in loop L2[14].

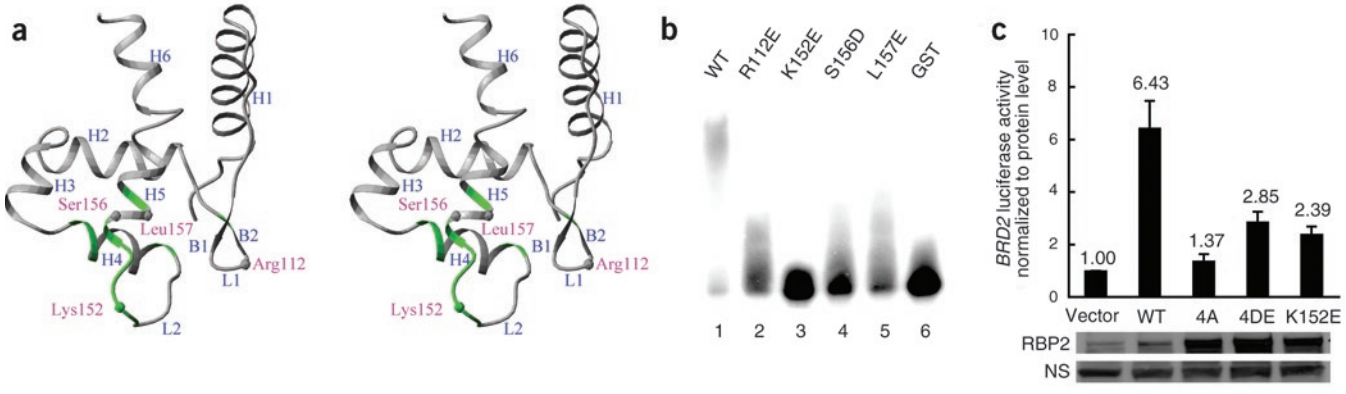

Figure 3 Structure and function of the ARID domain of RBP2. (**a**) Stereoview of RBP2 ARID structure as a ribbon diagram with the secondary elements labeled; the residues with relatively large chemical shift changes upon DNA binding are highlighted in green, and key DNA binding residues are labeled. (**b**) Mutations of RBP2 ARID in GST fusions affect C1 binding in EMSA. Weaker binding by the mutants led to smearing (lanes 2, 4 and 5) or no shift (lane 3). GST alone was used as a control (lane 6). (**c**) Requirement of DNA binding to regulate specific gene transcription. SAOS-2 cells were transfected with a *BRD2* promoter–fused luciferase reporter construct in the absence or presence of VSVG-tagged wild-type RBP2 (WT) or RBP2 mutant (4A (R112 K152 S156 L157A), 4DE (R112 K152 L157E S156D) or K152E) followed by luciferase assays. The expression of RBP2 and the mutants was confirmed by western blotting using anti-RBP2 (shown) or anti-VSVG antibodies (**Supplementary Fig. 5**). NS denotes a nonspecific band used as an internal loading control. The luciferase intensity was normalized to the corresponding protein level (quantified with AlphaEaseFC). Error bars indicate s.d. from triplicates.

To verify the role of these regions in DNA sequence recognition, several RBP2 ARID mutants were tested for C1 binding by EMSA (**Fig. 3b**). The mutants showed reduced binding in the order K152E > S156D > L157E or R112E (with DNA binding of the K152E mutant being the most reduced). These results suggest that L2 and H5, especially Lys152 in L2, make a major contribution toward sequence-dependent DNA binding. Notably, the residue equivalent to Lys152 (position X in **Supplementary Fig. 4** online) in the three classes of ARID is threonine for AT-rich binding ARID, lysine in GC-rich binding ARID and serine in nonspecific ARID, which may partially explain the DNA binding preferences: threonine prefers thymine; lysine prefers guanine; and serine might not contribute to specificity[20].

We examined whether DNA contact is essential for RBP2 to regulate transcription. The *BRD2* promoter was fused to a luciferase reporter, and the promoter's activity was assayed in the absence or presence of wild-type RBP2 or DNA contact mutants (**Fig. 3c**). Wild-type RBP2 activated the *BRD2* promoter, as shown previously[3]. In contrast, RBP2-mediated activation of *BRD2* promoter activity was reduced in the DNA contact mutants. The mutant proteins were present at higher levels than wild-type RBP2, possibly as a result of decreased degradation of the DNA binding mutants by the proteasome, as previously observed[21]. Furthermore, all of the mutants retained intrinsic histone demethylase activity (**Supplementary Fig. 5** online), excluding the possibility that the reduced activity of the DNA contact mutants was caused by reduced protein expression or defective demethylase activity. Possible reasons for the loss of histone demethylase activity in the ARID-deletion mutant but not site-specific DNA binding mutants include alteration of spatial arrangements or global folding in the deletion mutant. Mutation of one of the CCGCCC motifs in the *BRD2* promoter led to a decrease in the expression of a reporter (**Supplementary Fig. 6** online). These results show that ARID residues essential for recognizing CCGCCC are required for RBP2 to regulate specific gene transcription. Thus, DNA binding by RBP2 might have a recruiting role and regulate the demethylase activity on specific genes. This finding is interesting considering that few histone modifiers contain sequence-specific DNA binding domains.

Accession codes. Protein Data Bank: Coordinates for the ARID domain of RBP2 have been deposited with accession code 2JXJ.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank B. Lamarche and H.-M. Shih for useful discussion, W. Hong at the Institute of Molecular and Cell Biology, Singapore, for providing RBP2 plasmids, and W.G. Kaelin, Jr. at Harvard Medical School, Boston, for *BRD2* promoter construct and the RBP2 antibody. NMR was performed at Campus Chemical Instrument Center of Ohio State University and the National Research Program for Genomic Medicine (Taiwan) NMR Core. This work was supported by funding from the Genomics Research Center (to L.-J.J. and M.-D.T.), and research grant DOH96-TD-G-111-003 from National Research Program for Genomic Medicine of Taiwan (to L.-J.J.) and CA69472 from US National Institutes of Health (to M.-D.T.).

AUTHOR CONTRIBUTIONS

S.T., Y.-T.W., L.-J.J. and M.-D.T. designed the research; S.T., Y.-C.T., C.Y., M.-Y.C., A.-N.C. and P.-H.L. performed the research; S.T., L.-J.J. and M.-D.T. analyzed the data and wrote the paper.

Published online at http://www.nature.com/nsmb/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Klose, R.J. *et al. Cell* **128**, 889–900 (2007).
2. Christensen, J. *et al. Cell* **128**, 1063–1076 (2007).
3. Benevolenskaya, E.V., Murray, H.L., Branton, P., Young, R.A. & Kaelin, W.G., Jr. *Mol. Cell* **18**, 623–635 (2005).
4. Chan, S.W. & Hong, W. *J. Biol. Chem.* **276**, 28402–28412 (2001).
5. Hayakawa, T. *et al. Genes Cells* **12**, 811–826 (2007).
6. Kortschak, R.D., Tucker, P.W. & Saint, R. *Trends Biochem. Sci.* **25**, 294–299 (2000).
7. Herrscher, R.F. *et al. Genes Dev.* **9**, 3067–3082 (1995).
8. Gregory, S.L., Kortschak, R.D., Kalionis, B. & Saint, R. *Mol. Cell. Biol.* **16**, 792–799 (1996).
9. Whitson, R.H., Huang, T. & Itakura, K. *Biochem. Biophys. Res. Commun.* **258**, 326–331 (1999).
10. Collins, R.T., Furukawa, T., Tanese, N. & Treisman, J.E. *EMBO J.* **18**, 7029–7040 (1999).
11. Dallas, P.B. *et al. Mol. Cell. Biol.* **20**, 3137–3146 (2000).
12. Yuan, Y.C., Whitson, R.H., Liu, Q., Itakura, K. & Chen, Y. *Nat. Struct. Biol.* **5**, 959–964 (1998).
13. Iwahara, J., Iwahara, M., Daughdrill, G.W., Ford, J. & Clubb, R.T. *EMBO J.* **21**, 1197–1209 (2002).
14. Kim, S., Zhang, Z., Upchurch, S., Isern, N. & Chen, Y. *J. Biol. Chem.* **279**, 16670–16676 (2004).
15. Patsialou, A., Wilsker, D. & Moran, E. *Nucleic Acids Res.* **33**, 66–80 (2005).
16. Scibetta, A.G. *et al. Mol. Cell. Biol.* **27**, 7220–7235 (2007).
17. Lee, M.G., Norman, J., Shilatifard, A. & Shiekhattar, R. *Cell* **128**, 877–887 (2007).
18. Yamane, K. *et al. Mol. Cell* **25**, 801–812 (2007).
19. Iwahara, J. & Clubb, R.T. *EMBO J.* **18**, 6084–6094 (1999).
20. Luscombe, N.M., Laskowski, R.A. & Thornton, J.M. *Nucleic Acids Res.* **29**, 2860–2874 (2001).
21. Molinari, E., Gilman, M. & Natesan, S. *EMBO J.* **18**, 6439–6447 (1999).