A Novel ARID/Bright-like Protein Involved in Transcriptional Activation of Cyst Wall Protein 1 Gene in *Giardia lamblia*

Chih-Hung Wang, Li-Hsin Su, and Chin-Hung Sun

From the Department of Parasitology, College of Medicine, National Taiwan University, Taipei 100, Taiwan

The capability of protozoan parasite *Giardia lamblia* to encyst is critical for survival outside the host and its transmission. AT-rich interaction domain (ARID) or Bright homologs constitute a large family of transcription factors in higher eukaryotes that regulate cell proliferation, development, and differentiation. We asked whether *Giardia* has ARID-like genes and whether they influence gene expression during Giardia encystation. Blast searches of the *Giardia* genome database identified two genes with putative ARID/Bright domains (gARID1 and 2). Epitope-tagged gARID1 was found to localize to nuclei. Recombinant gARID1 specifically bound to the encystation-induced cyst wall protein (*cwp*) gene promoters. Mutation analysis revealed that AT-rich initiators were required for binding of gARID1 to the *cwp* promoters. gARID1 contains several key residues for DNA binding, and its binding sequences are similar to those of the known ARID family proteins. The gARID1 binding sequences were positive cis-acting elements of the *cwp1* promoter during both vegetative growth and encystation. We also found that gARID1 transactivated the *cwp1* promoter through its binding sequences in vivo. Our results suggest that the ARID family has been conserved during evolution and that gARID1 is an important transactivator in regulation of the *Giardia cwp1* gene, which is key to *Giardia* differentiation into cysts.

*Giardia lamblia* is an important human intestinal parasite that causes outbreaks of waterborne diarrheal disease (1, 2). It has two life cycle stages that are well adapted to survival in different inhospitable environments (3–6). The motile flagellated trophozoites attach to and colonize the human small intestine to cause the symptoms of giardiasis. The dormant differentiated into cysts.

G. *lamblia* is of significant evolutionary interest because it has been proposed as one of the most early diverging eukaryotes (13, 14). The giardial transcription mechanism may be unusual because relatively short 5′-flanking regions (≤65 bp) are sufficient for the expression of many constitutive and encystation-induced giardial genes (11, 15–17). No classical TATA or CCAAT boxes or other cis-acting elements have been found in the promoters of many giardial protein-coding genes (16, 18). AT-rich sequences have been found spanning the transcription start sites of many genes, indicating that they are functionally similar to the initiator (Inr)² element in higher eukaryotes (8, 9, 11, 15–17, 19). These sequences are essential for promoter activity and play a key role in determining the transcription start sites (16, 17, 20). Specific regulatory regions have been identified in the encystation-induced *cwp2* promoter, including a positive cis-acting element (−23 to −10 relative to the translation start site) required for encystation-specific promoter activity and a negative cis-acting element (−64 to −23 region) required for promoter activity in vegetative cells (20). The latter region also contains a region (−61 to −52) that controls sterol-mediated decrease in the *cwp2* transcription (21).

Relatively little is known about the transcription machinery in *G. lamblia*. Only four of the twelve general transcription initiation factors have giardial homologs, and they appear to have diverged at a higher rate than those of crown group eukaryotes (22). Giardial TATA-binding protein is highly divergent with respect to archaeal and higher eukaryotic TATA-binding proteins (22). *Giardia* does not have a true TFIIB homolog but has a homolog to TFIIB-related factor, which is involved in RNA polymerase III transcription in higher eukaryotes (22). Moreover, giardial RNA polymerase II transcription is highly resistant to α-amanitin (23). Few transcription factors have been characterized to date (24, 25). *Giardia* has several Myb proteins, one of which is encystation-induced and is involved in coordinating up-regulation of four key encystation-induced genes, *cwp1, cwp2, cwp3*, and *g6pi-b* and *gmyb2*.

---

* This work was supported by grants from the National Science Council (NSC 94-2320-B-002-093) and the National Health Research Institutes (NHRI-EX95-9510NC) in Taiwan and in part by the Dept. of Medical Research in National Taiwan University Hospital. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the DDBJ/GenBank™/EBI Data Bank with accession number(s) DQ88039 and DQ88040.

1 To whom correspondence should be addressed. Tel.: 886-2-23123456 (ext. 8262); Fax: 886-2-23915294; E-mail: chsun@ha.mc.ntu.edu.tw.

2 The abbreviations used are: Inr, initiator; ARID, AT-rich interaction domain; RT, reverse transcription; RACE, rapid amplification of cDNA ends.
ARID Family Transcription Factors in Giardia

itself (24). The GARP family transcription factors may be involved in transcriptional regulation of many different genes, including the encystation-induced cwp1 gene and constitutive ran gene (25).

AT-rich interaction domain (ARID) proteins are known to be involved in chromatin remodeling and regulation of development, cellular differentiation, and cell cycle control (16, 26). The ARID protein family has been found in yeast, Caenorhabditis elegans, plants, Drosophila, and mammals (26, 27). They have been shown to have AT-rich sequence-specific DNA binding activity and have been reported to function as transcriptional activators or repressors (26, 27). Mouse Bright (B cell regulator of IgH transcription) is a B cell-specific transactivator (28). Drosophila dead ringer (dri) is expressed in a developmentally regulated set of tissues, and its product may be a transcriptional activator or repressor depending on the cis-regulatory context (29). Human or mouse ARID family proteins have been grouped into seven subfamilies based on their sequence homology (Fig. 1) (30). The binding sites of the ARID3 subfamily Bright and DRI are (A/G)AT(T/A)AA, and (A/G)ATTAA or TATGTAT, respectively, all of which contain the sequence ATT embedded within a larger AT-rich sequence (27, 28, 31). The Bright core binding sites also flanked with AT or ATC run containing AT dimers, which are characteristic of matrix attachment region recognition sites (28). The binding site of the ARID5 subfamily MRF2 is AATA(C/T) (32). Some ARID family proteins are not sequence-specific DNA-binding proteins or they do not bind AT-rich sequences (26, 27). For example, the ARID1 subfamily p270 and Osa, which are members of SWI/SNF complexes, have DNA binding activity but no preference for specific DNA sequences (33–35). The JARID2 subfamily JUMONJI seems to bind both AT-rich and non-AT-rich sequences (36).

Because most giardial promoters contain AT-rich sequences spanning the transcription start sites, we asked whether Giardia has ARID family proteins and whether they influence gene expression. We searched the Giardia genome database for genes encoding ARID-like domains. We identified two giardial ARID homologs and determined the function of gARID1 in Giardia. We found that gARID1 bound to specific AT-rich Inr sequences and functioned as a transcriptional activator in G. lamblia.

EXPERIMENTAL PROCEDURES

Giardia Culture—Trophozoites of G. lamblia WB (ATCC 30957), clone C6, were cultured in modified TYI-S33 medium (37) and encysted as described previously (10).

Isolation and Analysis of the ARID Genes—The G. lamblia genome data base (www.mbl.edu/Giardia) (38) was searched with the amino acid sequence of the mouse MRF2 (GenBankTM accession number AF280065) using the BLAST program (39). The gARID1 coding region with 255 nt of 5′- and 190 nt of 3′-flanking regions was cloned and the nucleotide sequence was determined. The garriv1 gene sequence in the data base was essentially correct. The sequence of the garid2 gene was also confirmed. To isolate the cDNA of the garid1 gene, we performed RT–PCR with garid1-specific primers using total RNA from Giardia. For RT–PCR, 5 μg of DNase-treated total RNA from vegetative and 24-h encysting cells was mixed with oligo(dT)12–18 and Superscript II RNase H reverse transcriptase (Invitrogen). Synthesized cDNA was used as a template in subsequent PCR with primers A1F (ATGATGTCGCTCCTCCT) and A1R (ATACCCAAAGCTGTTAAT). Genomic and RT–PCR products were cloned into pGEM-T easy vector (Promega) and sequenced (ABI; Applied Biosystems). RT–PCR analysis of cwp1 (U09330) and ran (U02589) gene expression was performed using primers cwp1F (ATGATGTCGCTCCTCCTT) and cwp1R (TCAGGCGGAGGTGAGGCA) and ranF (ATGTCTGACCACCAATCAGC) and ranR (TCAATCATCGTCCGGGAA), respectively.

RNA and DNA Extraction, Northern Blot, and 5′-RACE Analysis—Total RNA was extracted from Giardia at the indicated differentiation stages using TRIzol reagent (Invitrogen). For Northern blot analysis, 10 μg of total RNA was fractionated and transferred to charged nylon membranes (Biodyne B membrane; Pall Corp.). Full-length coding region probes of luciferase, cwp1 (U09330), and ran (U02589) genes were prepared by PCR amplification of genomic DNA using primers lucF (ATGGAGAGCCTAAAAAC) and lucR (TTACACGGCGATCT- TTCC), cwp1F and cwp1R, and ranF and ranR, respectively. Radiolabeled probes were prepared using the Rediprime II kit (Amersham Biosciences). The membranes were hybridized and washed as described previously (11). Hybridization signals were imaged and quantified using a Storm system (Molecular Dynamics). 5′-RACE analysis was performed using the 5′-RACE system (Invitrogen). Oligonucleotides lucRA1 (CAACACTTTAATCGCAGTAT) and lucRA2 (CGGGAATGATTTGATTGCA) were used as first-strand primer and nested primer, respectively.

Plasmid Construction—All garid1 fragments were amplified from genomic DNA by PCR. All constructs were verified by DNA sequencing with an BigDye Terminator 3.1 DNA sequencing kit and an ABI 3100 DNA Analyser (Applied Biosystems). Plasmid pPW1 has been described previously (40). The garid1 gene and its 255-bp 5′-flanking region was amplified with oligonucleotides A5XF (GGCGTCTAGATCAGCTGACGGACGCTTATATTGATG) and A5AUR (GGCGGAATTCCTAGATGATCGACGACGCTATATTG) and digested with XbaI/EcoRI, and cloned into pPW1m3, pPW1m4, and pPW1m5, respectively.

The gARID1 coding region with 255 nt of 5′- and 190 nt of 3′-flanking regions was cloned and the nucleotide sequence was determined. The garid1 gene sequence in the data base was essentially correct. The sequence of the garid2 gene was also confirmed. To isolate the cDNA of the garid1 gene, we performed RT–PCR with garid1-specific primers using total RNA from Giardia. For RT–PCR, 5 μg of DNase-treated total RNA from vegetative and 24-h encysting cells was mixed with oligo(dT)12–18 and Superscript II RNase H reverse transcriptase (Invitrogen). Synthesized cDNA was used as a template in subsequent PCR with primers A1F (ATGATGTCGCTCCTCCTACT) and A1R (ATACCCAAAGCTGTTAATATG). Genomic and RT–PCR products were cloned into pGEM-T easy vector (Promega) and sequenced (ABI; Applied Biosystems). RT–PCR analysis of cwp1 (U09330) and ran (U02589) gene expression was performed using primers cwp1F (ATGATGTCGCTCCTCCTT) and cwp1R (TCAGGCGGAGGTGAGGCA) and ranF (ATGTCTGACCACCAATCAGC) and ranR (TCAATCATCGTCCGGGAA), respectively.

DNA and RNA Extraction, Northern Blot, and 5′-RACE Analysis—Total RNA was extracted from Giardia at the indicated differentiation stages using TRIzol reagent (Invitrogen). For Northern blot analysis, 10 μg of total RNA was fractionated and transferred to charged nylon membranes (Biodyne B membrane; Pall Corp.). Full-length coding region probes of luciferase, cwp1 (U09330), and ran (U02589) genes were prepared by PCR amplification of genomic DNA using primers lucF (ATGGAGAGCCTAAAAAC) and lucR (TTACACGGCGATCTTTCC), cwp1F and cwp1R, and ranF and ranR, respectively. Radiolabeled probes were prepared using the Rediprime II kit (Amersham Biosciences). The membranes were hybridized and washed as described previously (11). Hybridization signals were imaged and quantified using a Storm system (Molecular Dynamics). 5′-RACE analysis was performed using the 5′-RACE system (Invitrogen). Oligonucleotides lucRA1 (CAACACTTTAATCGCAGTAT) and lucRA2 (CGGGAATGATTTGATTGCA) were used as first-strand primer and nested primer, respectively.

Plasmid Construction—All garid1 fragments were amplified from genomic DNA by PCR. All constructs were verified by DNA sequencing with an BigDye Terminator 3.1 DNA sequencing kit and an ABI 3100 DNA Analyser (Applied Biosystems). Plasmid pPW1 has been described previously (40). The garid1 gene and its 255-bp 5′-flanking region was amplified with oligonucleotides A5XF (GGCGTCTAGATCAGCTGACGGACGCTTATATTGATG) and A5AUR (GGCGGAATTCCTAGATGATCGACGACGCTATATTG) and digested with XbaI/EcoRI, and cloned into pPW1m3, pPW1m4, and pPW1m5, respectively.

The gARID1 coding region with 255 nt of 5′- and 190 nt of 3′-flanking regions was cloned and the nucleotide sequence was determined. The garid1 gene sequence in the data base was essentially correct. The sequence of the garid2 gene was also confirmed. To isolate the cDNA of the garid1 gene, we performed RT–PCR with garid1-specific primers using total RNA from Giardia. For RT–PCR, 5 μg of DNase-treated total RNA from vegetative and 24-h encysting cells was mixed with oligo(dT)12–18 and Superscript II RNase H reverse transcriptase (Invitro
AU1 monoclonal antibody (1/5000 in blocking buffer; BAbCO) and detected with peroxidase-conjugated goat anti-mouse IgG (1/5000; Pierce) and enhanced chemiluminescence (Amer sham Biosciences).

**Immunofluorescence Assay**—The pNA1 stable transfectants were cultured in growth medium under G418 selection. Cells were harvested after 0 or 24 h of encystation, washed in phosphate-buffered saline, attached to glass coverslips (2×10⁶ cells/cover-slip), and then fixed and stained (11). Cells were reacted with anti-AU1 monoclonal antibody (1/300 in blocking buffer; BAbCO), and anti-mouse ALEXA 568 (1/500 in blocking buffer; Molecular Probes) was used as the detector. gARID1 was visualized using a Leica TCS SP2 Spectral Confocal System.

**Expression and Purification of Recombinant gARID1 Protein**—The genomic garid1 gene was amplified using oligonucleotides A1F and A1R. The product was cloned into the expression vector pCRT7/CT-TOPO (Invitrogen) in-frame with the C-terminal His and V5 tags to generate plasmid pA1. To make the gARID1m expression vector, the garid1 gene was amplified using two primer pairs A1F and A1mR (CTGGG-TACGAGAGTGGCCCCAGCATGAC) and A1mF (GTCAACATGGCGGACCATTTG) and A1mF (GTCAACATGGCGGACCATTTG) (the underlined GCC encode Ala) and A1R. The two PCR products were purified and used as templates for a second PCR. The second PCR reaction also included primers A1F and A1R, and the product was cloned into the expression vector to generate plasmid pA1m. The pA1 or pA1m plasmid was freshly transformed into E. coli BL21(DE3)pLysS (QIAexpressionist; Qiagen). An overnight pre-culture was used to start a 250 ml culture. E. coli cells were grown to an A600 of 0.5 and then induced with 1 mM isopropyl-D-thiogalactopyranoside (Sham Biosciences). Some bacteria were centrifuged, and the supernatant was mixed with 1 ml of a 50% slurry of nickel-nitrilotriacetic acid superflow resin (Promega) for 2 h. Bacteria were harvested by centrifugation and sonicated in 10 ml of buffer A (50 mM sodium phosphate, pH 8.0, 300 mM NaCl) containing 10 mM imidazole and complete protease inhibitor mixture (Roche Applied Science). The samples were centrifuged, and the supernatant was mixed with 1 ml of a 50% slurry of nickel-nitrilotriacetic acid superflow (Qiagen). The resin was washed with buffer A containing 20 mM imidazole and then eluted with buffer A containing 250 mM imidazole. Fractions containing gARID1 were pooled, dialyzed in 25 mM HEPES, pH 7.9, 20 mM KCl, and 15% glycerol and stored at 70 °C. Protein purity and concentration were estimated by Coomassie Blue and silver staining compared with bovine serum albumin. gARID1 was purified to apparent homogeneity as a predicted molecular mass of 52.5 kDa and a pI of ~6.4. The minimal structure of the ARID domain includes six α-helices (H1–H6) separated by β-strands or loops, and it may extend to additional helix at the N terminus or additional helices at both ends (H0 and H7) (26, 27). The H4-loop2-H5 region is similar to a helix-turn-helix structure (27). Structural studies of Drosophila DRI show that loop2 and helix 5 contact the minor groove and a β-sheet or loop between H1 and H2 and sequences downstream of H6 contact the minor groove (Fig. 2) (27).

**RESULTS**

**Identification and Expression of garid1 Gene**—To identify genes encoding novel ARID proteins from *Giardia*, we searched the *G. lamblia* genome data base (www.mbl.edu/Giardia) (38) using the amino acid sequence of a mouse ARID transcription factor, MRF2, that regulates smooth muscle cell differentiation, as a query sequence (GenBank™ accession number AF280065) (42). Amino acid sequences with similarity to the ARID domain were found in two proteins that we named gARID1 and 2 (GenBank™ accession number DQ88039 and DQ88040, respectively) (Fig. 1). We first focused on understanding the role of gARID1 in *Giardia*. Comparison of genomic and cDNA sequences showed that the garid1 gene contained no introns.

The deduced gARID1 protein contains 469 amino acids with a predicted molecular mass of ~52.5 kDa and a pI of ~6.4. The minimal structure of the ARID domain includes six α-helices (H1–H6) separated by β-strands or loops, and it may extend to an additional helix at the N terminus or additional helices at both ends (H0 and H7) (26, 27). The H4-loop2-H5 region is similar to a helix-turn-helix structure (26, 27). Structural studies of Drosophila DRI show that loop2 and H5 contact the major groove and a β-sheet or loop between H1 and H2 and sequences downstream of H6 contact the minor groove (Fig. 2) (27). Sequence alignment of the loop2 and helix 5 of the ARID domain shows that the ARID domain of gARID1 is highly similar to those of the human ARID3 subfamily (Fig. 1) (30). A neighbor-joining (43) phylogenetic tree obtained from the alignment of the loop2 and helix 5 also revealed similarity between gARID1 and the human ARID3 subfamily and between gARID2 and the human JARID2 or AIRD4 subfamily (data not shown). We further aligned the ARID domain of gARID1 with that of the human ARID3 subfamily and Drosophila DRI (Fig. 2). The ARID domain of gARID1 has 35% sequence identity and 51% sequence similarity to Drosophila DRI. Some of the key contact residues identified by structural studies of...
ARID Family Transcription Factors in Giardia

Drosophila DRI are conserved in gARID1 (Fig. 2) (27). Four residues in mouse Bright are important for DNA binding, Pro-268, Trp-299, Phe-317, and Tyr-330 (Fig. 2) (44). With the exception of Phe-317, all are conserved in Giardia gARID1 (Fig. 2). In contrast to the human ARID3 subfamily and Drosophila DRI, the ARID domains of which are central, the giardial ARID domains are near the N terminus (residues 1–98) (Fig. 2). Unlike the human ARID3 subfamily and Drosophila DRI, which contain H0 and H7 motifs at both ends (27), gARID1 may have an incomplete H0 or H7 (Fig. 2). The similarity between gARID1 and the human ARID3 subfamily or Drosophila DRI is limited to their ARID domains (data not shown). The C-terminal half of gARID1 was not highly conserved and had no apparent functional motif.

Expression of the garid1 Gene—RT-PCR analysis of total RNA showed a single ~1.4-kb transcript, in agreement with the size of the full-length garid1 gene (Fig. 3A). The garid1 transcript was present in vegetative cells and decreased significantly in 24-h encysting cells (Fig. 3A). The transcript levels of the garid1 gene also decreased slightly during early encystation (data not shown).

To determine the expression of gARID1 protein, we prepared construct pNA1 in which the garid1 gene is controlled by its own promoter and contains an AU1 epitope tag at its C terminus (Fig. 3B) and stably transfected it into Giardia. A ~56-kDa protein was detected (Fig. 3C), which is slightly larger than the predicted ~52.5-kDa molecular mass of gARID1 with the AU1 tag (~0.8 kDa). The levels of the gARID1 protein increased significantly in encysting cells (Fig. 3C). However, the levels of the garid1 message decreased significantly in encysting pNA1 cells, similar to the result in wild-type cells (data not shown, see Fig. 3A for wild-type cells). The lack of correlation between the steady state mRNA and protein levels could be due to an increase in translation rate or protein half-life.

Localization of the gARID1 Protein—The AU1-tagged gARID1 was detected exclusively in the nuclei during vegetative growth and encystation (Fig. 4, A and B), indicating that gARID1 is a nuclear protein in Giardia. Expression was ~25 and ~50% positive in vegetative and encysting cells, respectively. Some cytosolic staining was also observed in ~5 and ~10% of stained vegetative and encysting trophozoites, respectively (Fig. 4B; data not shown). The AU1-tagged gARID1 was also detected in four nuclei and the cytoplasm of some cysts (Fig. 4C).

Identification of the gARID1 DNA Binding Sites—The nuclear localization of gARID1 suggested that it might function as a transcription factor. Because its protein level increases during encystation, we tested the hypothesis that it may bind DNA and regulate the transcription of genes required for encystation. To test its DNA binding activity, we expressed gARID1 with a C-terminal V5 tag in E. coli and purified it to >95% homogeneity as assessed by silver-stained gels (data not shown). An anti-V5-horseradish peroxidase antibody specifically recognized the recombinant gARID1 (Fig. 5A).

To determine whether purified gARID1 binds DNA, we performed electrophoretic mobility shift assays with double-stranded DNA sequences from the 5′-flanking region of an
encystation-induced gene, cwp1, and a constitutive gene, ran. Incubation of a labeled double-stranded DNA probe cwp1 −45/−1 (Table 1) with gARID1 resulted in the formation of shifted bands (Fig. 5B, lane 2). cwp1 −45/−1 is the region from −45 to −1 bp relative to the translation start site of the cwp1 gene. The sequence of this promoter is shown in Table 1. gARID1 did not bind to either single strand of the cwp1 −45/−1 probe (data not shown). gARID1 was shown to bind to cwp1 −90/−46, and within this region it bound to the 3’-region (cwp1 −68/−46), but not to the region cwp1 −90/−69 (Fig. 5B, lanes 6 and 7; Table 1). gARID1 also bound to a well characterized core promoter, ran −51/−20 (16), but not to ran −30/−1 or ran −81/−52 (Fig. 5B, lane 8; Table 1). The mobility of the gARID1-DNA complex did not change too much with the size of the probe, possibly because the mobility depends much on the size and charge of gARID1 and on the conformation of the gARID1-DNA complex (45).

The binding specificity was confirmed by competition and supershift assays (Fig. 5B, lanes 3–5; additional data not shown). gARID1 bound to cwp1 −45/−1 or −68/−46 could be supershifted by an anti-V5-horseradish peroxidase antibody (lane 3; data not shown). The formation of the shifted cwp1 −45/−1 bands was almost totally competed by a 200-fold molar excess of unlabeled cwp1 −45/−1, but not by the same excess of a nontarget competitor, ran −30/−1 (lanes 4 and 5).

Scanning mutagenesis of the cwp1 −45/−1 probe showed that substitutions within the AGATC or AATAAAATA sequence significantly decreased the DNA-protein interaction (Fig. 6, lanes 4, 8, and 9) but mutations of the other regions caused a minor decrease in binding (lanes 2, 3, 5–7, and 10). Mutation of both regions almost abolished binding (lane 11). Substitutions of the three Ts within the AATAAAATA region decreased binding by ~50% (lane 12, Table 1). The gARID1 binding sequences or their antisense sequences in the cwp1 −45/−1 probe are similar to those of the known ARID family proteins that contain ATT, AT, or ATC sequences (27, 28, 31, 36). Likewise, two gARID1 binding regions, cwp1 −68/−46 and ran −51/−20, also contain the sequence ATC (Fig. 5B, lanes 7 and 8; Table 1). Mutation analysis of the cwp1 −68/−46 or ran −51/−20 probe revealed that the short AATCT or AATCG sequence was required for binding, respectively (Fig. 5B, lanes 9 and 10; data not shown).

We also tested whether gARID1 binds to the 5’-flanking region of two other encystation-induced genes, cwp2 and cwp3. We found that gARID1 bound to the cwp3 −30/−1 probe (weakly) and the extended probes cwp2 −30/+8 and cwp3 −30/+10 (Fig. 5B, lanes 10 and 11; Table 1) but it did not bind to the cwp2 −30/−1 probe (Table 1 and data not shown), indicating that the extended probes cwp2 −30/+8 and cwp3 −30/+10 may include the whole gARID1 binding sites. The results suggest that gARID1 can bind to the cwp2 and cwp3 promoter regions. We also tested whether gARID1 binds to specific AT-rich sequences. Interestingly, gARID1 did not bind to a poly(A) sequence (Table 1 and data not shown) but bound to a poly(A) sequence with a T insertion or a TC insertion (Fig. 5B, lanes 12 and 13), indicating that the gARID1 binding sequence contains AT or ATC sequences.
gARID1 Binds to the Minor Groove and Contains Key Residues for DNA Binding—Studies suggest that Drosophila DRI and mouse Bright can bind to DNA minor grooves (27, 28). To investigate whether gARID1 can also bind to the minor groove, we used distamycin A, which binds to the minor groove of AT-rich DNA sequences, as a competitive inhibitor of gARID1 binding (28, 46). As shown in Fig. 6B, the binding of gARID1 to DNA decreased with increasing concentrations of distamycin A. The binding was completely inhibited at concentrations of 0.56 mM.

Four residues of mouse Bright are known to be important for DNA binding, Pro-268, Trp-299, Phe-317, and Tyr-330 (Fig. 2) (44). These residues, with the exception of Phe-317, are conserved in Giardia gARID1 (Fig. 2). We tried to understand whether Tyr-82 of gARID1, which corresponds to Phe-317 of the mouse Bright, is also important for DNA binding (Fig. 2).

The Tyr-82 was mutated to Ala, and the resulting gARID1 mutant (gARID1m) was expressed in E. coli and purified. We found that the purified gARID1m did not bind to cwp1-45/1, cwp1-90/46, ran-51/20, or the other probes listed in Table 1 (data not shown), indicating that the Tyr-82 is important for DNA binding.

The gARID1 Binding Sequences Are Sufficient for Transcriptional Activation—We further investigated the ability of the gARID1 binding sites to regulate the cwp1 promoter

### TABLE 1

| Name      | Sequence (5' to 3') | Binding Activity* |
|-----------|--------------------|-------------------|
| cwp1 -45/-1 | GTTTACACTCTAGATCAAGAAGAACACAGAAATAATTCAGGG | 100±4.1 |
| cwp1 -45/-1m1 | GCGGGGACTCTAGATCAAGAAGAACACAGAAATAATTCAGGG | 70±3.1 |
| cwp1 -45/-1m2 | GTTTACACTCTAGATCAAGAAGAACACAGAAATAATTCAGGG | 80±3.3 |
| cwp1 -45/-1m3 | GTTACACTCTCTGGATCAAGAAGAACACAGAAATAATTCAGGG | 30±2.2 |
| cwp1 -45/-1m4 | GTTACACTCTCTGGATCAAGAAGAACACAGAAATAATTCAGGG | 70±4.1 |
| cwp1 -45/-1m5 | GTTACACTCTCTGGATCAAGAAGAACACAGAAATAATTCAGGG | 50±2.7 |
| cwp1 -45/-1m6 | GTTACACTCTCTGGATCAAGAAGAACACAGAAATAATTCAGGG | 60±3.2 |
| cwp1 -45/-1m7 | GTTACACTCTCTGGATCAAGAAGAACACAGAAATAATTCAGGG | 30±2.1 |
| cwp1 -45/-1m8 | GTTACACTCTCTGGATCAAGAAGAACACAGAAATAATTCAGGG | 20±1.5 |
| cwp1 -45/-1m9 | GTTACACTCTCTGGATCAAGAAGAACACAGAAATAATTCAGGG | 50±3.1 |
| cwp1 -45/-1m10 | GTTACACTCTCTGGATCAAGAAGAACACAGAAATAATTCAGGG | 40±2.9 |
| cwp1 -45/-1m11 | GTTACACTCTCTGGATCAAGAAGAACACAGAAATAATTCAGGG | 50±2.9 |
| cwp1 -90/46 | CTGGCACTACGTCGCTGCCTCAATTTACGATATACATCA | + |
| cwp1 -90/69 | CTGGCACTACGTCGCTGCCTCAATTTACGATATACATCA | - |
| cwp1 -68/46 | CAAATTACGATATACATCA | + |
| cwp1 -68/46m | CAAATTACGATATACATCA | - |
| cwp2 -30/1 | CAGGACCTACGTCGCTGGAGAAAATAAA | - |
| cwp2 -30/+8 | CAGGACCTACGTCGCTGGAGAAAATAAA | - |
| cwp2 -60/31 | TAGACTGCTGTCGCTGAGGGGGACAGATTA | - |
| cwp3 -30/1 | GTCCTTTCCCCAAAAATAAGTGTTACTCTGT | +/- |
| cwp3 -30/+10 | GTCCTTTCCCCAAAAATAAGTGTTACTCTGT | + |
| cwp3 -60/31 | TAGACTGCTGTCGCTGAGGGGGACAGATTA | - |
| g6pi-b -30/1 | TATATATGATTACGCTGCGCAGAAAATAAAA | - |
| g6pi-b -60/31 | TCTGATGAGGGTTTACTACGAGTGACAG | - |
| ran -30/-1 | TAAACCTTTAACAGGAGACGAGAGTAGC | - |
| ran -51/-20 | TAAATGTTATAATTCGGAATTAAACTTAA | + |
| ran -51/-20m | TAAATGTTATAATTCGGAATTAAACTTAA | +/- |
| ran -81/-52 | CAGAAAAAGGGTCTCGAGAGGTTAGAGGCG | - |
| poly A | AAAAAAAAAAAAAAAAAAAAAAAAAA | - |
| poly A+1T | AAAAAAAAAAAAAAAAAAAAAAAAAA | + |
| poly A+2T | AAAAAAAAAAAAAAAAAAAAAAAAAA | + |

*The ability of oligonucleotides to bind to the gARID1 protein. The numbers show the binding activity of mutants relative to that of the wild-type cwp1 -45/1 probe (mean ± S.E. of three independent experiments). "+", "++", and "--" represent DNA binding, weak binding, and no binding, respectively. The translation start sites are underlined. Base changes in the mutants tested are framed.
by mutation analysis. The 5'-flanking region −409/−1 of the cwp1 gene was sufficient for up-regulation of the luciferase reporter gene during encystation (construct pPW1, induction ratio ∼47.6, Fig. 7). Mutation of the −34 to −30 and −17 to −12 region, which spans the first gARID1 binding site and the AT-rich Inr element, resulted in significant decreases of luciferase activity to ∼1.8 and ∼3.7% of the pPW1 activity in vegetative and encysting cells, respectively (Fig. 7, pPW1m3). Mutation of the −18 to −8 region, which spans the whole AT-rich Inr element, also resulted in significant decreases in luciferase activity to ∼0.5 and ∼1.5% of the pPW1 activity in vegetative and encysting cells, respectively (Fig. 7, pPW1m4).

We further asked whether the decrease in luciferase activity was due to decreased mRNA levels. Two different transcripts, a full-length (1.7-kb) transcript and a 1.2-kb transcript, were detected in the pPW1 transfectants during both vegetative growth and encystation (Fig. 8). Two major transcription start sites, 13 bp upstream (−13) and 462 bp downstream (+462) of the luciferase translation start codon, were identified by 5'-RACE analysis of total RNA isolated from the pPW1 transfectants during encystation (Fig. 7, and data not shown). The 1.7- and 1.2-kb transcripts could be the result of transcription initiation upstream and downstream of the luciferase translation start codon, respectively. During vegetative growth, the levels of the 1.7-kb transcript decreased more than the levels of the 1.2-kb transcript in the pPW1m3 and pPW1m4 transfectants compared with the pPW1 transfectants (Fig. 8), indicating that mutation of the gARID1 binding sites spanning the AT-rich Inr resulted in a downstream shift in transcription start site selection. Therefore, the gARID1 binding sites are important for transcription start site selection during vegetative growth. During encystation, an upstream shift in transcription start site selection was found in all cell lines, irrespective of the presence of the mutated gARID1 binding sites (Fig. 8). The upstream transcription start sites in the pPW1m3 or pPW1m4 transfectants during encystation were shifted to −11, −19, and −21 upon mutating the gARID1 binding sites/Inr to GC-rich sequences (Fig. 7). The levels of both 1.7- and 1.2-kb transcripts decreased significantly in the pPW1m3 and pPW1m4 transfectants compared with the pPW1 transfectants during encystation (Fig. 8). These results indicate that the gARID1 binding sequences in the cwp1 gene promoter function as posi-
ARID Family Transcription Factors in Giardia

tive cis-acting elements during both vegetative growth and encystation.

Transactivation of cwp1 Promoter Activity by gARID1—Identification of the gARID1 binding sequences in the cwp1 promoter allowed us to examine whether the gARID1 protein actually activates transcription of these promoters. Transactivation was examined by stably transfecting a reporter plasmid containing a luciferase gene into Giardia together with an effector plasmid pNA1 (Fig. 3B). Co-transfection of the pPW1 reporter construct that expressed the luciferase gene under the control of the cwp1 promoter (Fig. 7) with pNA1 resulted in a 1.6 to 1.8-fold increase in luciferase activity during normal growth compared with the activity in the pPW1/pRANneo co-transfectants (data not shown) (41). In addition, when the luciferase gene was under the control of the ran32-bp promoter in the ran32 construct (24), co-transfection with pNA1 did not increase luciferase activity compared with the activity in the ran32/pRANneo co-transfectants (data not shown). The results indicate that the gARID1 can transactivate the cwp1 promoter but not the ran promoter.

To understand whether the increase in luciferase activity was due to increased mRNA levels, we measured the luciferase mRNA levels from different transfectants. Co-transfection of pNA1 into the pPW1 transfectants resulted in a 1.6 to 1.8-fold increase in the levels of full-length 1.7-kb luciferase mRNA and endogenous cwp1 mRNA compared with the levels in the pPW1/pRANneo co-transfectants (Fig. 9), indicating that gARID1 can transactivate the cwp1 promoter.

We further asked whether gARID1 transactivates the cwp1 promoter through its binding sites. We found that the levels of the full-length 1.7-kb luciferase transcript decreased significantly in the pPW1m3/pNA1 and pPW1m4/pNA1 co-transfectants compared with their control cell lines (Fig. 9). However, the levels of the 1.2-kb transcript increased by 1.8-fold in the pPW1m3/pNA1 and pPW1m4/pNA1 co-transfectants. Therefore, mutation of the gARID1 target sequences in the cwp1 promoter canceled the transactivation through the upstream transcription start sites by gARID1 but increased the transactivation through the downstream transcription start sites. The results suggest that gARID1 can transactivate the cwp1 promoter in vivo through its DNA binding sequences.

DISCUSSION

ARID domains have been found in a wide variety of species, including yeast, nematodes, plants, insects, and, mammals (26, 27). Although divergent in sequence, giardial ARID1 and ARID2 are clearly recognizable as members of the ARID family. This suggests that the ARID family may have evolved before divergence of Giardia from the main eukaryotic line of descent. To date, gARID1 is the first ARID transcription factor identified in early diverging protozoan parasites. Further genomic analyses will reveal whether ARID family proteins are shared with other eukaryotic lineages such as Trichomonas vaginalis, Entamoeba histolytica, Plasmodium falciparum, and Trypanosoma brucei.
Although the ARID domain of gARID1 is similar to that of the human ARID3 subfamily, it may have incomplete H0 and H7 motifs. Our results indicate that, like members of the human ARID3 subfamily, gARID1 can bind to similar AT-rich specific sequences. gARID1 has three of the four residues in mouse Bright that are known to be important for DNA binding (Fig. 2) (44). We found that Tyr-82 of gARID1, which corresponds to Phe-317 of mouse Bright, was also required for gARID1 binding, indicating the importance of this residue. The ARID domain of the gARID2 is similar to that of the human ARID4 or JARID2 subfamily. However, these two subfamilies have no sequence-specific DNA binding activity or they do not bind AT-rich sequences. Further studies are needed to characterize the DNA binding activity of gARID2.

gARID1 has some but not all of the conserved residues in Drosophila DRI that contact the phosphodiester backbone, minor groove, or major groove (Fig. 2) (27). The binding sites of the ARID3 subfamily Bright and DRI and ARID5 subfamily MRF2 are (A/G)AT(T/A)AA, (A/G)ATTAA or TATTGAT, and AATA(C/T), respectively (28, 29, 31, 32). These sequences contain ATT or AT/ATC runs. Our results indicate that the AGATC or AATAAAATA sequence may be important for gARID1 binding. Further studies also indicate that gARID1 can bind to the poly(A) sequence with a T or TC insertion. These gARID1 binding sequences contain ATT or AT/ATC sequences on the coding or noncoding strand, indicating that the ARID family binding sites may have been conserved in evolution.

The giardial promoters, including the encystation-specific cwp promoters, defined to date are notably short and contain AT-rich Inr elements (8–10, 15–17, 20, 24). Their proximal upstream regions also contain gMyb2 or GARP-like protein binding sites and these sequences are positive cis-acting elements (24, 25). Our current mutation analysis provides further evidence of the function of the AT-rich Inr elements. Mutation of the gARID1 target sequences/Inr in the cwp1 promoter might impair the binding of gARID1 or other Inr-binding proteins to the Inr, leading to a downstream shift in transcription start site selection during vegetative growth. During encystation, transcription start site selection of the cwp1 promoter was shifted upstream, irrespective of mutations in the gARID1 binding sites/Inr. Therefore, the binding of some encystation-specific transcription factors to proximal upstream regions might be important for the selection of the upstream transcription start sites. Interestingly, in the cwp2 promoter, an encystation-specific regulatory element (−23 to −10 region) is just next to the gARID1 binding sequences/Inr (−8 to −1 region) (20), making it more likely that there is an interaction between regulatory and general transcription factors. It should be noted that, during both vegetative growth and encystation, the levels of both transcripts resulting from upstream or downstream transcription start sites decreased significantly when the gARID1 binding sequences/Inr were mutated, suggesting that the gARID1 binding sequences/Inr could be important for basal promoter activity. Deletion and mutation analysis of the ran, α2-tubulin, and cwp2 promoters has provided evidence for involvement of AT-rich Inr elements in basal promoter activity (15, 16, 20). In addition, there may be a synergistic effect between the upstream or downstream transcription start sites. The synergistic effect may come from the interaction of gARID1 and other transcription factors with both transcription start sites.

ARID proteins in higher eukaryotes are involved in differentiation and function as transcriptional activators or repressors (26, 27). Our results show that gARID1 localizes mainly to the nuclei in both vegetative and encysting Giardia trophozoites. The increased levels of the gARID1 protein during encystation indicate that it may also play a role in differentiation. The findings that gARID1 binding sites/Inr in the cwp1 promoter are positive cis-acting elements and that overexpressed gARID1 can induce the cwp1 promoter activity indicate that gARID1 is involved, at least in part, in promoting cwp1 gene expression during encystation. The ability of gARID1 to transactivate the encystation-induced cwp1 gene may require the binding of other transcription factors to promoter elements or interaction with gARID1.

gARID1 can bind AT-rich Inr elements of both the constitutive ran gene and the encystation-induced cwp1, cwp2, and cwp3 genes, suggesting that gARID1 may be involved in transcriptional regulation of many different genes. It has been shown that the AT-rich Inr in the ran promoter is an important positive cis-acting element (16). However, in this study, overexpressed gARID1 did not transactivate the ran promoter. Although not proven, it is still possible that gARID1 is involved in transcriptional regulation of many different genes.

As gARID1 may bind to more than one AT-rich sequence in the cwp1 promoter region, it is likely that additional transcription factors binding to the AT-rich Inr element or interacting with gARID1 increase the specificity for transcriptional regulation in vivo. When the gARID1 target sequences/Inr element were mutated, the overexpressed gARID1 could not transactivate the cwp1 promoter through the upstream transcription start sites within the AT-rich Inr element, but its transactivation through the downstream transcription start sites increased. This indicates that gARID1 may transactivate the cwp1 promoter in vivo through its DNA binding sequences/Inr element and that binding of gARID1 or other Inr-binding proteins to the Inr are required for accurate transcription start site selection.

In summary, we have characterized gARID1 as a transcriptional activator of the Giardia cwp1 gene. Our studies provide new insights into the evolution of eukaryotic pathways of DNA binding in the early diverging protozoan G. lamblia.

Acknowledgments—We thank the researchers and administrators of the G. lamblia Genome data base. We also thank Dr. Frances D. Gillin for valuable discussions and Dr. Janet Yee for helpful comments on the electrophoretic mobility shift assay experiments.

REFERENCES
1. Adam, R. D. (2001) Clin. Microbiol. Rev. 14, 447–475
2. Wolfe, M. S. (1992) Clin. Microbiol. Rev. 5, 93–100
3. Eichinger, D. (2001) Curr. Opin. Microbiol. 4, 421–426
4. Gillin, F. D., Reiner, D. S., Gault, M. J., Douglas, H., Das, S., Wunderlich, A., and Sauch, J. F. (1987) Science 235, 1040–1043
5. Gillin, F. D., and Reiner, D. S. (1996) Annu. Rev. Microbiol. 50, 679–705
