J Huazhong Univ Sci Technol [Med Sci]
DOI 10.1007/s11596-015-1455-9 35(4):469-476,2015

Binding of Human SWI1 ARID Domain to DNA without Sequence Specificity: A Molecular Dynamics Study*

Qian SUN (孙 茜), Tao ZHU (朱 涛)#, Chang-yu WANG (王常玉), Ding MA (马 丁)
Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

Summary: SWI1 is a member of a new class of tumor DNA-binding proteins named as the AT-rich interaction domain family (ARID), and considered to bind with AT base pairs specifically. Genomic and functional data support ARID1A as a tumor suppressor because ARID1A/BAF250a (SWI1) subunit of the SWI/SNF chromatin-remodeling complex has emerged as recurrently mutated in a broad array of tumor types. But the crystal structure of SWI1 has not been solved as yet. Using docking and molecular dynamics, we predicted the DNA interaction pattern of human SWI1 ARID and made comparisons with the other two representative ARID family members, human Mrf-2 ARID and Drosophila Dri ARID. Dynamic results revealed that the N-terminal and loop L1 of SWI1 ARID bound with the DNA major groove, while the loop L2 and helix H6 bound with the minor groove. Moreover, it was found that SWI1 ARID bound with DNA apparently in a sequence-nonspecific manner. It was concluded that SWI1 ARID can form stable complex with sequence-nonspecific DNA segment comparing to Mrf-2 ARID/DNA and Dri ARID/DNA sequence-specific complexes.
Key words: AT-rich interaction domain family 1A (ARID); BAF250a; Mrf-2 ARID; Dri ARID; protein-DNA interaction; ovarian cancer

Recently, high-frequency mutations in epigenetic regulators have been a novel and active subject of studies involving cancer genome sequencing. Both genomic and functional studies have revealed that AT-rich interaction domain 1A (ARID1A) is a tumor suppressor, as many mutations in ARID1A/BAF250a (also known as human SWI1), a subunit of the SWI/SNF, have been identified in a wide array of tumor types[1]. For example, recurrent inactivating mutations in ARID1A have been identified in ovarian[2, 3] and uterine tumors[4], and next-generation sequencing of endoscopic biopsies identified ARID1A as a tumor-suppressor gene in Barrett's esophagus[5].

ARID1A encodes BAF250a, a key component of the SWI/SNF chromatin-remodeling complex. Currently, every known type of SWI/SNF chromatin-remodeling complex incorporates a subunit containing an ARID DNA binding domain. BAF250a-DNA interactions are physiologically relevant in higher eukaryotes, and they also facilitate SWI/SNF binding to target sites *in vivo*. Cooperative interactions among intrinsic subunit chromatin interaction domains and sequence-specific transcription factors drive SWI/SNF recruitment[6]. Loss of the BAF250a protein correlates strongly with the presence of ARID1A mutations in ovarian clear cell carcinoma[7] and high-grade endometrial carcinoma, although not in other types of endometrial carcinoma. Indeed, the loss of BAF250a expression does not have prognostic value for endometrial carcinoma[8]. By contrast, the expression of ARID1A/BAF250a could potentially be used by clinicians to assess the prognosis of cervical cancer[9] and gastric carcinoma[10]. As a ubiquitous component of the SWI/SNF complex, BAF250a plays an absolutely necessary physiological role in modulating the timing of replication at select replication domains in pluripotent cells[11], and it also acts as a key component of the gene regulatory machinery in embryonic stem cells, controlling self-renewal, differentiation, and cell lineage decisions[12].

BAF250a is a member of a new class of tumor DNA-binding proteins belonging to ARID family. Decades ago, it was thought that the ARID family bound to AT-rich DNA specifically, although it is now apparent that the DNA-binding activity of ARID-containing proteins is not necessarily sequence-specific. For example, Dri binds to a core ATTA motif[13], and Mrf-2 tends to bind the sequence AATAC/T in binding site selection assays[14]. However, the protein-DNA interaction binding sites and DNA-binding structures of SWI1 ARID have not been identified. Lacking a crystal structure for the SWI1/DNA complex, researchers nevertheless believe that the human SWI1 ARID is involved in DNA binding, as it can cross-link to DNA using UV-activated cross-linkers when the SWI/SNF complex is bound to DNA[15]. Another group used a PCR-assisted DNA binding selection method with a pool of random oligonucleotides and concluded that human SWI1 ARID binds to double-stranded DNA in a non-sequence-specific manner[16]. Moreover, the solution structure of yeast SWI1 ARID is similar to the human homolog, and it was

Qian SUN, E-mail: endeavor_forever@126.com
#Corresponding author, E-mail: tjogzt@163. com
*This project was supported by the College Scientific and Technological Innovation Project of Huazhong University of Science and Technology (No. 15A263).

shown to bind DNA non-specifically[17].

The basic structure of ARID family proteins is a series of six alpha-helices separated by beta-strands, loops, or turns. The most significant differences among the various ARID family members involve the presence of an additional helix at either or both ends of the basic six conserved regions[18]. Using solution nuclear magnetic resonance (NMR) methods, the ARID structures of human SWI1[19] and two other representative ARID proteins, human Mrf-2[20] and Drosophila Dri[21], have been solved. The ARIDs in SWI1, Mrf-2 and Dri share many similarities in their central regions, although significant differences exist at both the N- and C-termini.

Based on the above studies, we aimed to use the NMR structures of the Dri and Mrf-2 ARID proteins to simulate the structure of the SWI1 ARID/DNA complex, which should improve our understanding of the binding profile between the SWI1 ARID and DNA molecules in greater detail. More specifically, we isolated the SWI1 ARID structure and docked it with DNA sequences from the Dri ARID/DNA complex (sequence 1: 5'-CCTGTATTGATGTGG-3') or the Mrf-2 ARID/DNA complex (sequence 2: 5'-TACAATATAACGTCG-3'). The DNA bases of the sense sequences (5' to 3') are referred to as C1/T1 to G15, whereas antisense bases are marked with single quotations and are referred to as C'16 to G'30/A'30.

By comparing our predicted SWI1 ARID/DNA binding model with the Dri ARID/DNA and Mrf-2 ARID/DNA NMR structures, we were able to draw conclusions about various similarities and differences in the structure and DNA interaction profiles of these three ARID family members.

1 MATERIALS AND METHODS

1.1 Structure Modeling

We downloaded the solved structure of SWI1 ARID (PDB. 1RYU) from the protein data bank (PDB), which was determined using NMR in 2004. The molecule contains 120 amino acid residues representing region 1000–1119 aa of the ARID1A gene isoform A (NP_006006.3) reference sequence. The complete ARID1A reference sequence contains 2285 amino acids, and region 1017–1108 aa encodes the ARID. Therefore, the known structure 1RYU represents the SWI1 ARID domain as well as 17 N-terminal amino acids and 11 C-terminal amino acids.

In addition, we downloaded the ARID structures of the other two representative ARID family members, Dri ARID (PDB. 1KQQ) and Mrf-2 ARID (PDB. 2OEH), and aligned their sequences with SWI1 ARID separately. We also isolated DNA fragment structures from the Dri ARID/DNA complex (sequence 1: 5'-CCTGTATTGATGTGG-3') and the Mrf-2 ARID/DNA complex (sequence 2: 5'-TACAATATAACGTCG-3') and used them as DNA templates to predict the potential binding profiles of the SWI1 ARID/DNA complex. We docked SWI1 ARID with the above two DNA segments separately and then made comparison between the different binding complexes. The DNA sequences from the above two ARID complexes are referred to as Dri_DNA and Mrf-2_DNA.

1.2 Protein-DNA Docking

We used the software program Hex 8.0.0 to dock the DNA fragments described above with the SWI1 ARID. We used spherical polar Fourier (SPF) correlations instead of the Cartesian grid representations of protein shape and other properties as an improvement over conventional fast Fourier transform (FFT) docking methods[22]. Additionally, as an interactive molecular graphics program, Hex was also used to calculate and display feasible docking modes for the protein-DNA complexes.

The various orientations were sorted by docking energy and the 10 000 highest scoring (i.e., lowest energy) orientations were recorded. The optimal 2000 orientations that were spatially similar were then ordered by docking energy and retained for clustering. The conformation with minimal docking energy within the largest cluster was then selected as our initial conformation to run the molecular dynamics simulations.

1.3 Molecular Dynamics Simulations and Free-energy Decomposition Analysis

Molecular dynamics simulations were performed using the NAMD 2.9[23] and CHARMM27 force-field software programs. The proteins, both alone and in complex with DNA, were solvated in rectangular boxes of water. Sodium and chloride ions were added at the minimum amounts necessary to neutralize the overall charge of the simulation system.

In the preparatory steps of the simulations, the solvent molecules around the fixed protein and DNA complex were equilibrated with an initial energy minimization using the Conjugate Gradient (CG) algorithm for 10 000 steps. The constraints on the complex were then removed, and the systems were energy-minimized for another 10 000 steps. The systems were equilibrated for 1 ns at 310 K before the beginning of the production phase. The equilibration and production simulations were carried out under NPT conditions (constant number of particles, pressure and temperature) with periodic boundary conditions and Particle Mesh Ewald (PME) summation[24]. The production phases were carried out for another 4 ns.

We sampled information such as kinetic energy, van-der-Waals (VDW) energy, electric energy and various structural properties over all conformations from the production phases every 1 ps. Root-mean-square deviation (RMSD) to any frame in the trajectory and root mean square fluctuations (RMSF) per residues were calculated to determine structural stability and identify high fluctuant regions. We also used these trajectory files to analyze the binding sites of the complexes at 40 ps intervals.

2 RESULTS

2.1 Comparisons of DNA Binding Profiles among Structures of SWI1 ARID, Dri ARID and Mrf-2 ARID

The complete ARID1A protein reference sequence (NP_006006.3) contains 2285 amino acids, and region 1017–1108 aa encodes the ARID. The reference protein PDB 1RYU contains 120 amino acid residues, corresponding to the sequence region 1000–1119 aa. The secondary structure of SWI1 ARID contains six helices,

referred to as H1 (1010–1012 aa), H2 (1018–1033 aa), H3 (1051–1061 aa), H4 (1066–1068 aa), H5 (1072–1079 aa) and H6 (1087–1099 aa). There are also two beta strands (1014–1016 aa and 1045–1048 aa), two turns (1100–1107 aa and 1109–1111 aa), and two highly flexible regions that form loops, referred to as L1 (located between H2 and H3) and L2 (located between H5 and H6)[19].

NMR structures of the other two representative ARID family members, Mrf-2 ARID (PDB. 2OEH) and Dri ARID (PDB. 1KQQ) had been determined roughly a decade before. Multiple sequence alignment and secondary structure prediction of these three ARIDs revealed high sequence similarity and structural conservation (fig. 1A and 1B). Alignment results revealed sequence identities of 31.7% (SWI1 ARID and Dri ARID) or 17.2% (SWI1 ARID and Mrf-2 ARID), and sequence similarities of 59.3% (SWI1 ARID and Dri ARID) or 45.1% (SWI1 ARID and Mrf-2 ARID). Human SWI1 ARID contained all six well superimposed helices and two loops that were conserved in the structures of Mrf-2 ARID or Dri ARID. Helices H1–H6 of the SWI1 ARID can be superimposed on those of the Mrf-2 and Dri ARIDs with an average RMSD of 1.8 Å and 1.9 Å between the backbone atoms, respectively. However, the SWI1 ARID helices H1 (located at the N-terminus) and H4 (located between H3 and H5) contained only three residues and were thus much shorter than the helices in the same locations in Dri or Mrf-2 ARID. Residues at the C-termini of the three sequences were all highly flexible and not well superimposed. The RMSD of the matched sequence of the SWI1 and Mrf-2 ARID was 10.496, whereas the RMSD of SWI1 and Dri ARID was 9.327.

NMR studies of Mrf-2 ARID[20] and Dri ARID[21] in complex with DNA have revealed that two regions of the ARID are involved in minor groove and phosphodiester backbone interactions, namely the loop L1 and the C-terminus. An interaction point with the major groove was mapped to helix H5 and the preceding loop L2. In particular, for the Dri/DNA complex, the amino acids Gln79, Thr92, Ser93, Arg99, Arg130 and Arg131, which were found in loop L2, helix H5, H6 and the C-terminus, formed intermolecular hydrogen bonds with the phosphate backbone or with A-T base pairs (fig. 1C). For the Mrf-2/DNA complex, the amino acids Arg24, Arg55, Lys58, Thr71, Ser72, His80, Arg106 and Lys107 located in loop L2, helix H5 and the C-terminus also formed intermolecular hydrogen bonds with the negatively charged phosphate backbone of the DNA molecule (fig. 1D). The roles of Arg45, Ala50 and Lys51 in the Dri ARID loop L1 were analogous to those of Arg24 and Pro26 in Mrf-2 ARID loop L1, both of which interacted with the deoxyribose phosphate backbone in the minor groove of the DNA. Residues Thr92–Ser93 of Dri ARID, corresponding to Thr71–Ser72 of Mrf-2 ARID, made base-specific contacts with A-T base pairs in the major groove of the DNA. In the Mrf-2 ARID/DNA complex, Lys58 bound to T'22, Thr71 bound to G'20, T'21 and T'22, and Ser72 bound to A7, T8, A9, T'21, T'22 and A'23. Similarly, in the Dri ARID/DNA complex, Thr92 bound to A10 and T11, Ser93 bound to A10 and T'21 (hydrophobic bond), and Arg131 bound to T13 and C'19.

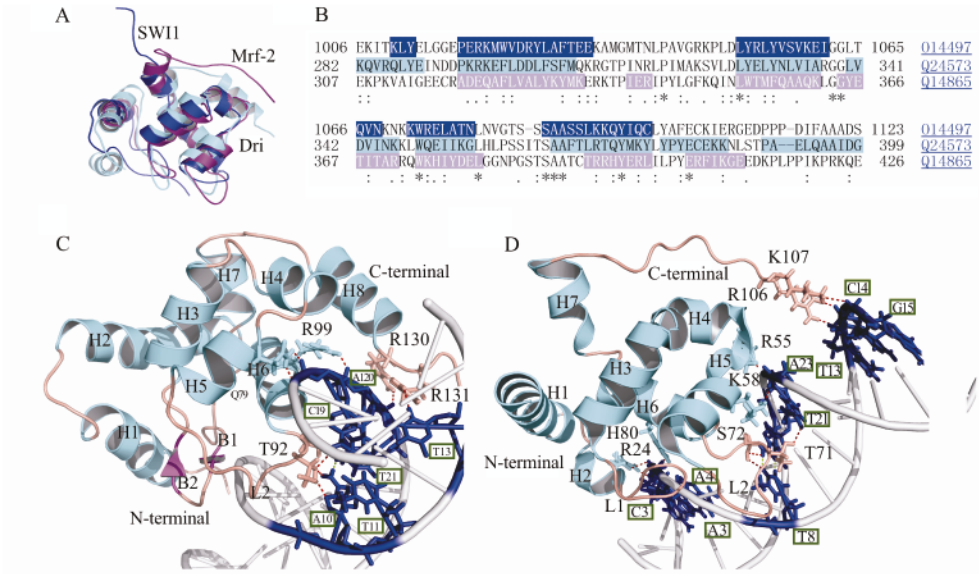

Fig. 1 Sequence alignment and structures among SWI1 ARID, Mrf-2 ARID and Dri ARID domains

A: superimposion of NMR structures among SWI1 ARID (blue), Mrf-2 ARID (purple) and Dri ARID (cyan) domains. ARID family members contain 6 conserved helices which are superimposed very well. SWI1 ARID contains an extra N-terminal that does not superimpose with others. Residues at the C termini of those three structures are highly flexible and not structured. B: comparison of sequence and predicted second structures between SWI1 ARID (blue), Mrf-2 ARID (purple) and Dri ARID (cyan). Sequence alignment results show that the sequence identity is equal to 31.7% (SWI1 ARID & Dri ARID) or 17.2% (SWI1 ARID and Mrf-2 ARID), and sequence similarity is 59.3% (SWI1 ARID and Dri ARID) or 45.1% (SWI1 ARID and Mrf-2 ARID). C: NMR structure of Dri ARID/Dri_DNA complex; D: NMR structure of Mrf-2 ARID/Mrf-2_DNA complex

2.2 Comparison of DNA Binding Profiles among Human SWI1 ARID, Mrf-2_DNA and Dri_DNA

In general, regular secondary structure motifs are important for maintaining the overall architecture of a protein, whereas flexible regions are more important for interactions with other molecules and functions. As mentioned previously, the H5 and L2 regions of both Mrf-2 ARID and Dri ARID are highly flexible and can bind to

the major groove of DNA, whereas L1 and the C-terminus are involved in minor groove and phosphodiester backbone interactions.

For SWI ARID, we docked the ARID with Mrf-2_DNA and Dri_DNA separately, and then selected the best pose as the initial conformation for molecular dynamics. The total energy of the best docking pose was –437.3 kCal/mol for the SWI1 ARID/Mrf-2_DNA complex and –424.3kCal/mol for the SWI1 ARID/Dri_DNA complex. Analysis of the hydrogen bonds between ARID and the DNA chains for both best conformations before molecular dynamics showed different binding profiles. For the SWI1 ARID/Mrf-2_DNA complex, Ser1-Thr3 in the N terminus and Ser92, Lys95, and Gln96 in helix H6 interacted with the deoxyribose or phosphate backbone of T8, A9, A10 and C11 in the minor groove. Ser1 and Thr40, which are located in loop L1, formed base-specific contacts with A9 and A18 in the major groove. By contrast, although the total energy for the SWI1 ARID/Dri_DNA complex was similar to the above complex, hydrogen bonds inside the complex were very weak. Only Arg54 and Tyr56 in helix H3 formed hydrogen bonds with the deoxyribose of A10 and C17 in the minor groove. Lys60 in helix H3 formed a base-specific contact with C17 in the major groove.

Molecular dynamics simulations were performed on the SWI1 ARID and the specific DNA sequences bound by the Dri and Mrf-2 monomers to gain insights into the structural stability and conformational dynamics of these ARID/DNA complexes. The RMSD of the backbone Cα atoms of the simulated protein over time can be used to analyze the structural stability of the system. To characterize the stability of the SWI1 ARID/DNA complex and to compare it with the Mrf-2 ARID /DNA and Dri ARID /DNA complexes, we performed molecular dynamics on the SWI1 monomer as well as the SWI1 ARID/Mrf-2_DNA, SWI1 ARID/Dri_DNA, Mrf-2 ARID/Mrf-2_DNA and Dri ARID/Dri_DNA complexes. During the first 1 ns of the simulation, all systems underwent conformational readjustments based on the bound DNA segment and the initial ARID structure, and the systems monotonically tended towards an equilibrium state (fig. 2A).

Unsurprisingly, results showed that the RMSD values of the SWI ARID monomer were larger than the RMSD values of the other systems, with the exception of the last 0.5 ns of the SWI1 ARID/Dri_DNA system. RMSD values for the SWI1 ARID/Mrf-2_DNA system were similar to those for the Mrf-2 ARID/Mrf-2_DNA NMR structure and were significantly lower than for the SWI1 ARID monomer, indicating that SWI1 ARID reached a more stable equilibrium state when binding with the Mrf-2_DNA and that conformational dynamics were similar to those for the Mrf-2 ARID/Mrf-2_DNA complex NMR structure. By contrast, the SWI1 ARID/Dri_DNA system showed higher fluctuations over its assumed equilibrium state than did the Dri-2 ARID/Dri_DNA complex NMR structure, indicating that the SWI1 ARID might not form stable complex with Dri_DNA. These results were consistent with the structural analysis of the SWI1 ARID/Dri_DNA complex showing that only relatively weak hydrogen bonds were formed between these two molecules.

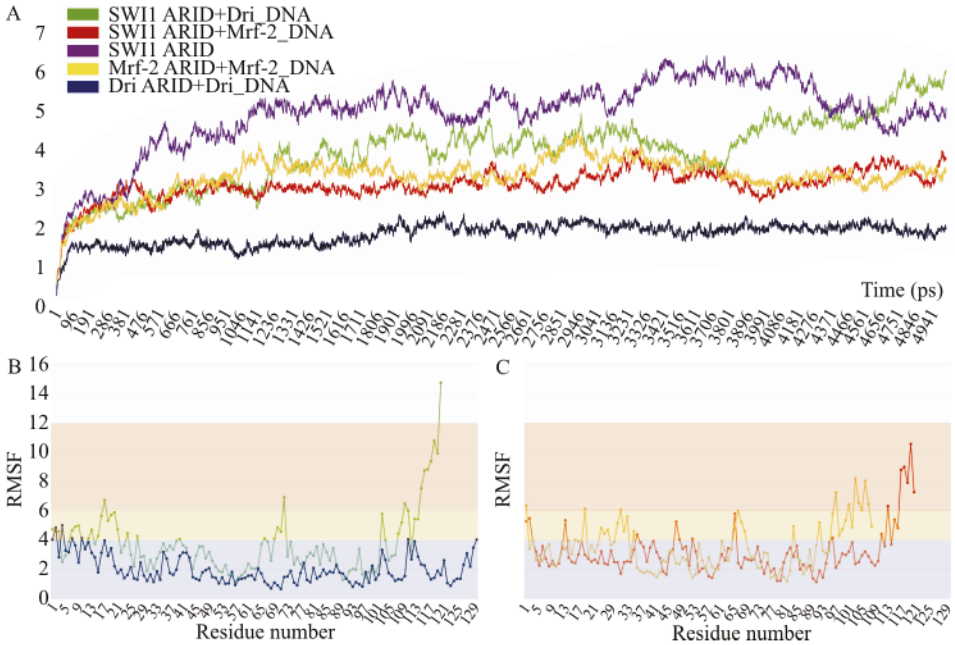

Fig. 2 RMSD of the backbone Cα atoms and RMSF per residues of the simulated protein over simulation time

A: RMSD-time curve of SWI1 monomer, SWI1 bound with Dri-DNA and Mrf-2-DNA separately, Dri/DNA and Mrf-2/DNA complexes. During the first 1 ns of the simulation, all the systems reach an equilibrium state. B: RMSF-time curve of SWI1 bound with Dri-DNA and Dri/DNA complexes. C: RMSF-time curve of SWI1 bound with Mrf-2-DNA and Mrf-2/DNA complexes

To understand the structural basis for the observed differences in RMSD fluctuations between DNA bound to SWI1 ARID, Mrf-2 ARID and Dri ARID during the simulations, we also analyzed the RMSF values for each residue to identify regions with high fluctuations (fig. 2B and 2C). It was necessary to set levels for RMSF values in advance, and we divided the range of RMSF values into three arbitrary intervals to facilitate this property. In

particular, we assumed that RMSF value greater than or equal to 6 represented the high fluctuation group, values between 4 and 6 represented the moderate fluctuation group, and residues less than or equal to 4 represented the low fluctuation group. In addition, we summarized residues of high or moderate fluctuation and associated their secondary structural properties with their RMSF levels to explore the relationship between secondary structure and function. In general, binding to DNA or other ligands makes protein structures more stable. Therefore, residues with decreased RMSF values after binding with DNA could serve as binding sites, as was observed in the N terminus (Ser1–Ile9), L1 (Ala36–Leu50) and L2 (Asn82–Ser86) of SWI1 ARID (table 1).

Table 1 RMSF per residue of SWI1 ARID monomer and four ARID/DNA complexes during molecular dynamics simulation

Monomer/complex	4<RMSF<6 (moderate fluctuation)	RMSF≥6 (high fluctuation)
SWI1 ARID	T3..T4 E7 I9 L12..Y13 L15..G16 E18..E20 F31 E34..G38 T40..L42 Q67 K70..N71 N82..S86 S88 C100 E105 K107 E109 E112 P114	K8 G17 M39 K72 Y102 R110 P115..A120
SWI1 ARID/Mrf-2_DNA	S1..S2 Y13 K35 R47..K48 L52 L65 K95 R110 P114..P115	E112 P116..A120
SWI1 ARID/Dri_DNA	S1..T3 E7..T10 L12..Y13 L15..G16 E18..R21 W24 R27 T40 T66 N69..N71 Y102..A103 K107..I108 E112..D113	G17 K72 E109..R110 P114..A120
Mrf-2 ARID/Mrf-2_DNA	K16 R24..P26 G29 K31..Q32 N34 G65..P68 R83 F91 E95 D97..P99 P101 K107	R1 K19 F30 E96 L100 P102..R106
Dri ARID/Dri_DNA	G1..W2 F4 Q7 Q10 I17 E110 G131	–

2.3 Formation of a Stable Associative Ring on Binding Surface of SWI1 ARID/Mrf-2_DNA Complexes

The conformations of the SWI1 ARID/Mrf-2_DNA complexes with maximal (RMSD=3.98) and minimal (RMSD=2.60) RMSD values during the molecular dynamics simulation are discussed in detail. Using Pymol[25], the complexes are shown as "Cartoon" representation, and the binding residues are shown in "Sticks" mode (fig. 3A and 3B). Potential binding sites between these two conformations were somewhat similar, involving Ser1, Thr4, Asn6, Asn41, Ser86, Ser88, Ser92, and Lys95 in SWI1 ARID and A9, A10, C11, G12, T13, A'18, and C'19 in the DNA segment. In addition, Ser2, Lys8, Thr40, Ser87, and Lys94 in SWI1 ARID and T6, C'16, G'20, T'24, and T'26 in the DNA segment were found to be important in one of the conformations. In particular, Ser1–Thr2 in the N terminus, Ser86 in loop L2 and Ser88, Ser92 and Lys95 in helix H6 interacted with the deoxyribose or phosphate backbone of A9, A10 and C11 in the minor groove. Thr4 and Asn6 in the N-terminus, together with Thr40 and Asn41, which are located in loop L1, formed base contacts with C'19, A'18, G12 and T13 in the major groove of the DNA molecule. Interestingly, G12-C'19 and T13-A'18 formed hydrogen bonds inside two complementary DNA strands. Therefore, we observed a stable hydrogen interactive ring on the binding surface of the SWI1 ARID/Mrf-2_DNA complex. These hydrogen-bonding interactions between or within SWI1 ARID and the Mrf-2_DNA segment were quite clear, suggesting that the binding surface of the SWI1 ARID/Mrf-2_DNA complex forms a stable consecutive ring.

We also sampled conformations from the SWI1 ARID/Mrf-2_DNA and SWI1 ARID/Dri_DNA complexes every 40 ps during the last 4 ns of the simulations, and we summarize the total potential binding sites of these 100 samples in descending order. The results showed that the frequency distributional percentages of LYS95, THR40, ASN41, THR4, LYS8, SER92, LYS72, THR3, ARG75 and SER88 of SWI ARID ranged from 55% to 199% of the total samples. Likewise, for the Mrf-2_DNA segment (sequence 5'-TACAATATAAC GTCG-3') binding to SWI1 ARID, bases C11, A9, A10, G12 and T13 were potential binding sites, and bases C'16, G'17, A'18, G'20, C'19, T'24, A'25 and T'26 were potential binding sites on the complementary DNA strand (fig. 3C and 3D). These potential binding sites were similar to those predicted using maximal and minimal RMSD values. In particular, residues Thr3 and Thr4 of SWI1 ARID both bound to C11, C'19 and G'20 of the Mrf-2_DNA; Thr4 and Asn41 both bound to G12; Thr40 and Asn41 both bound to T13, G'17 and A'18; and Ser88 bound to T8 and T'24 (fig. 4A). For the SWI1 ARID/Dri_DNA complex, Val57 bound to T11 (hydrophobic bond), Lys60 bound to G12 and C'19, and Glu61 bound to A10, C'19, A'20 (fig. 4B).

In summary, we discovered that the SWI1 ARID can bind with not only AT-base pairs but also CG-base pairs. In the SWI1 ARID/Mrf-2_DNA and SWI1 ARID/Dri_DNA complexes, A10–T13 and A'18–T'21 of the DNA segments could form a stable connection within the SWI1 ARID backbone, whereas they bound to each other by hydrogen bonds inside the double-stranded DNA. Our findings provide new evidence that the SWI1 ARID forms stable, non-specific complexes with DNA segments based on comparisons with Mrf-2 ARID/DNA and Dri ARID/DNA complexes.

Finally, we discussed the total energy for each binding system. Energy values were sampled during the last 4 ns of the production simulations, and average values were calculated for each system. The interval of the average total energy values was –71355.5 kCal/mol to –34194.1 kCal/mol. All energy values for the SWI1 ARID monomer were lower than in the other systems, as the system size of the SWI1 ARID monomer was the smallest among the five systems. The system size of the SWI1 AR-ID/Mrf-2_DNA complex was comparable to the SWI1 ARID/Dri_DNA complex, whereas the total energy of the Dri_DNA-bound complex was higher than that of the Mrf-2_DNA-bound complex. These findings are consistent with the conclusion that binding to the Mrf-2_DNA segment makes these systems much more stable (table 2).

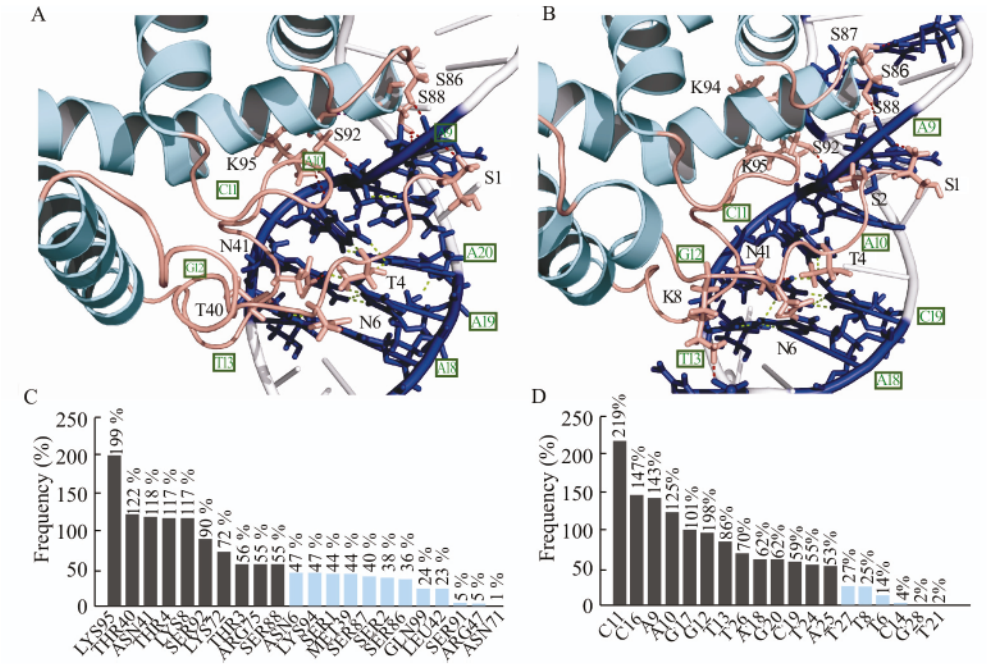

Fig. 3 Structures of SWI1 ARID/DNA complexes obtained from MD simulation

A: the structure of SWI1 ARID/DNA (pose 3271) complex with maximal RMSD value (RMSD=3.98) during the dynamics. Using Pymol, the complex is shown in Cartoon representation, and binding residues are shown in Sticks mode. B: the structure of SWI1 ARID/DNA (pose 3966) complex with minimal RMSD value (RMSD=2.60) during the dynamics. Details of the hydrogen bonding interactions between SWI1 ARID and DNA are shown (red dotted line). The hydrogen-bonding contacts within SWI1 ARID (magenta dotted line) or DNA (green dotted line) are shown. C: total potential binding residues of SWI1 ARID predicted by 100 SWI1_ARID/Mrf-2_DNA complex conformations sampled during the last 4 ns of molecular dynamics simulation in descending order. D: potential binding sites of Mrf-2_DNA segment

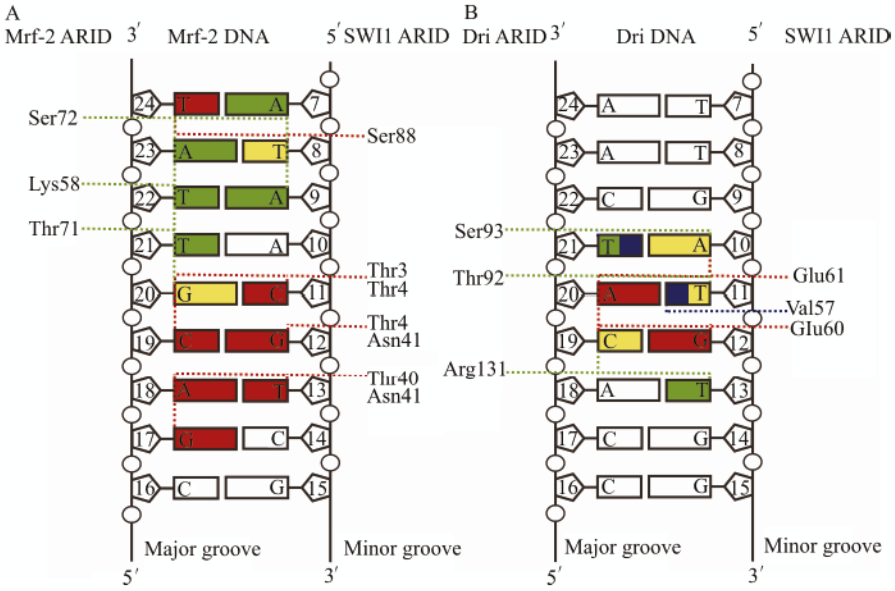

Fig. 4 Schematic diagram of specific ARID/DNA interactions

A: SWI1 ARID/Mrf-2_DNA and Mrf-2 ARID/DNA complexes; B: SWI1 ARID/Dri_DNA and Dri ARID/DNA complexes. Red regions indicate that those bases bind with SWI1 ARID residues by hydrogen bond, green regions indicate that those bases bind with Mrf-2 ARID or Dri ARID residues by hydrogen bond, yellow regions indicate that those bases bind with both of them by hydrogen bond, and blue regions indicate hydrophobic interactions between ARID and DNA.

Table 2 Average energy of SWI1 monomer and four ARID/DNA complexes

Monomer/complex	BOND	ELECT	VDW	KINETIC	TOTAL	Residue	Total number
SWI1 ARID/Mrf-2_DNA	743.0119	−101071	7790.408	17548.93	−71355.5	149aa	27758
SWI1 ARID/Dri_DNA	756.6095	−93927.6	7050.919	16254.41	−66163.6	150aa	25642
Mrf-2 ARID/Mrf-2_DNA	729.7896	−78936.6	5701.775	13648.56	−55308.7	107aa	21443
Dri ARID/Dri_DNA	804.3102	−71123.2	4667.11	12326.4	−49422	131aa	19179
SWI1 ARID	389.5936	−48888.5	3580.951	9046.53	−34194.1	120aa	14240

BOND: bond energy; ELECT: electrostatic energy; VDW: Van der Waals force energy; KINETIC: kinetic energy; TOTAL: total energy. All the above data are presented in the unit of Kcal/mol. Residue represents protein residue number among each complex, and total number represents molecular number of the whole ARID/DNA complex together with water and ions.

3 DISCUSSION

Protein-DNA interactions are integral to many biological processes, including transcriptional regulation, DNA replication and DNA repair[26]. In essence, proteins can "read" DNA sequences by directly probing the floors of the DNA grooves. Different DNA-binding protein families can vary greatly in the types of major groove motifs they bind to, whereas different proteins within the same family use sequence-specific information to achieve higher resolution specificity[27]. Although DNA deformations can be readily observed using structural methods such as X-ray diffraction and NMR spectroscopy, the forces imparted by a protein on DNA that cause deformation are often subtle and can be difficult to parse or rank in importance[28]. In many systems involving protein-DNA interactions and other molecular complexes, the binding interfaces often involve flexible loops and termini. These flexible regions undergo local folding that is coupled to complex formation. In addition, changes in flexibility can influence the affinity of these interactions through changes in conformational entropy and free energy changes. DNA molecules with attached cationic charges or neutralized phosphate backbones demonstrate the importance of electrostatic charges in DNA deformation, which play crucial roles in the intrinsic, sequence-dependent shape of DNA as well as shape changes induced by protein binding[29].

The ARID is a helix-turn-helix motif-based DNA-binding domain that is conserved in all eukaryotes and is characteristic of a family that includes 15 distinct human proteins with important roles in development, tissue-specific gene expression and cell proliferation control[18]. The 15 human ARID family proteins can be divided into seven subfamilies based on the degree of sequence identity between individual members. Most ARID family members have not been characterized with respect to their DNA-binding behavior, although it is apparent that not all ARIDs conform to the pattern of binding AT-rich sequences. DNA-binding affinity also varies somewhat between the ARID subfamilies. Site-specific mutagenesis experiments do not support the suggestions made based on structural analyses that specific amino acids in loop 2 and helix 5 are the main determinants of sequence specificity. Most probably, specificity is determined through multiple interactions across the entire ARID structure[30].

Helices H1–H6 of the SWI1 ARID can be superimposed on those of the Mrf-2 ARID with an average RMSD of 1.8 Å among the backbone atoms, and it can be superimposed on those of the Dri ARID with an average RMSD of 1.9 Å. The primary unique feature of the SWI1 ARID is that the N-terminus forms an extended structure followed by a single turn helix (H1), which interacts extensively with loop L1. This interaction is analogous to that observed between the α-helix H1 and L1 in Dri ARID. However, the orientation of loop L1 in the SWI1 ARID is more similar to that found in the Mrf-2 AR-ID: L1 is positioned close to loop L2 in both the Mrf-2 ARID and SWI1 ARID, but it is positioned away from L2 in Dri ARID[19]. Indeed, such structural differences between these ARIDs may affect their interactions with DNA.

RMSD calculations of the SWI1 ARID/Dri_DNA system showed higher fluctuations over its assumed equilibrium state than did the Dri ARID/Dri_DNA complex NMR structure, indicating that the SWI1 ARID may not form a stable complex with Dri_DNA. These findings are consistent with the structural analysis of the SWI1 ARID/Dri_DNA complex, which showed that only weak hydrogen bond interactions were formed. Additionally, the trend of decreasing RMSF values for residues in the Dri_DNA-bound complex was significantly different from the Mrf-2_DNA-bound systems. After binding to the Mrf-2_DNA segment, the largest decline in RMSF values was observed for the SWI1 ARID residues located at the N terminus (S1-I9), L1 (A36-L50) and L2 (N82-S86), which could represent DNA-interactive binding sites. Molecular dynamics simulations predicted multiple potential ARID/DNA binding sites, among which SWI1 ARID residues Thr3 and Thr4 in the N-terminus, Thr40 and Asn41 in loop L1, Val57, Glu60 and Glu61 in helix H3, and Ser88 in helix H6 bound specifically to the DNA segment. Moreover, SWI1 ARID bound to the ATG motif of the Dri_DNA segment and to the GACG motif of the Mrf-2_DNA segment, suggesting that SWI1 ARID may bind a GAC/TG motif. In other words, the SWI1 ARID could bind not only AT-base pairs but also CG-base pairs. DNA binding sites A10–T13 and A'18–T'21 could form a stable connection within the SWI1 ARID backbone, whereas A10–T13 and A'18–T'21 also bound to each other by hydrogen bonds inside the double-stranded DNA. Our findings provide evidence for a new model whereby SWI1 ARID family members form stable nonspecific complexes with DNA segments in a manner that is distinct from Mrf-2 ARID/DNA and Dri ARID/DNA complexes.

This study was based on structural predictions and ARID-DNA binding pattern simulations, and further computational and experimental studies are required to

verify our conclusions. Nevertheless, we provide new evidence showing that not all ARID family members specifically bind to AT-rich DNA segment. As mentioned above, the N-terminal, loop L1 and loop L2 regions of SWI1 ARID likely play key roles in ARID-DNA interactions. To more closely characterize ARID-DNA interaction patterns, researchers should mutate these predicted binding sites and compare mutant and wild-type structures.

Conflict of Interest Statement

The authors declared no potential conflicts of interest in this work.

REFERENCES

1 Wu JN, Roberts CW. ARID1A mutations in cancer: another epigenetic tumor suppressor? Cancer Discov, 2013,3(1):35-43
2 Jones S, Wang TL, Shih Ie M, *et al*. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. Science, 2010,330(6001):228-231
3 Wiegand KC, Lee AF, Al-Agha OM, *et al*. Loss of BAF250a (ARID1A) is frequent in high-grade endometrial carcinomas. J Pathol, 2011,224(3):328-333
4 Guan B, Mao TL, Panuganti PK, *et al*. Mutation and loss of expression of ARID1A in uterine low-grade endometrioid carcinoma. Am J Surg Pathol, 2011,35(5):625-632
5 Streppel MM, Lata S, DelaBastide M, *et al*. Next-generation sequencing of endoscopic biopsies identifies ARID1A as a tumor-suppressor gene in Barrett's esophagus. Oncogene, 2014,33(3):347-357
6 Chandler RL, Brennan J, Schisler JC, *et al*. ARID1a-DNA interactions are required for promoter occupancy by SWI/SNF. Mol Cell Biol, 2013,33(2):265-280
7 Wiegand KC, Shah SP, Al-Agha OM, *et al*. ARID1A mutations in endometriosis-associated ovarian carcinomas. New Engl J Med, 2010,363(16):1532-1543
8 Zhang ZM, Xiao S, Sun GY, *et al*. The clinicopathologic significance of the loss of BAF250a (ARID1A) expression in endometrial carcinoma. Int J Gynecol Cancer, 2014,24(3):534-540
9 Cho H, Kim JS, Chung H, *et al*. Loss of ARID1A/BAF250a expression is linked to tumor progression and adverse prognosis in cervical cancer. Hum Pathol, 2013,44(7):1365-1374
10 Wiegand KC, Sy K, Kalloger SE, *et al*. ARID1A/BAF250a as a prognostic marker for gastric carcinoma: a study of 2 cohorts. Hum Pathol, 2014,45(6):1258-1268
11 Takebayashi S, Lei I, Ryba T, *et al*. Murine esBAF chromatin remodeling complex subunits BAF250a and Brg1 are necessary to maintain and reprogram pluripotency-specific replication timing of select replication domains. Epigenet Chromatin, 2013,6(1):42
12 Gao X, Tate P, Hu P, *et al*. ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. P Natl Acad Sci USA, 2008,105(18):6656-6661
13 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
14 Whitson RH, Huang T, Itakura K. The novel Mrf-2 DNA-binding domain recognizes a five-base core sequence through major and minor-groove contacts. Biochem Bioph Res Co, 1999,258(2):326-331
15 Wilsker D, Patsialou A, Zumbrun SD, *et al*. The DNA-binding properties of the ARID-containing subunits of yeast and mammalian SWI/SNF complexes. Nucleic Acids Res, 2004,32(4):1345-1353
16 Dallas PB, Pacchione S, Wilsker D, *et al*. The human SWI-SNF complex protein p270 is an ARID family member with non-sequence-specific DNA binding activity. Mol Cell Biol, 2000,20(9):3137-3146
17 Wang T, Zhang J, Zhang X, *et al*. Solution structure of SWI1 AT-rich interaction domain from Saccharomyces cerevisiae and its nonspecific binding to DNA. Proteins, 2012,80(7):1911-1917
18 Wilsker D, Patsialou A, Dallas PB, *et al*. ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development. Cell Growth Differ, 2002,13(3):95-106
19 Kim S, Zhang Z, Upchurch S, *et al*. Structure and DNA-binding sites of the SWI1 AT-rich interaction domain (ARID) suggest determinants for sequence-specific DNA recognition. J Biol Chem, 2004,279(16):16670-16676
20 Cai S, Zhu LY, Zhang ZM, *et al*. Determination of the three-dimensional structure of the Mrf2-DNA complex using paramagnetic spin labeling. Biochemistry, 2007,46(17):4943-4950
21 Iwahara J, Iwahara M, Daughdrill GW, *et al*. The structure of the Dead ringer-DNA complex reveals how AT-rich interaction domains (ARIDs) recognize DNA. Embo J, 2002,21(5):1197-1209
22 Ritchie DW, Venkatraman V. Ultra-fast FFT protein docking on graphics processors. Bioinformatics, 2010,26(19):2398-2405
23 Phillips JC, Braun R, Wang W, *et al*. Scalable molecular dynamics with NAMD. J Comput Chem, 2005,26(16):1781-1802
24 Ewald PP. Die Berechnung optischer und elektrostatischer Gitterpotentiale. Ann Phys (German), 1921,369(3):253-287
25 DeLano WL. The PyMOL molecular graphics system, www.pymol.org. CA
26 Von Hippel PH. From "simple" DNA-protein interactions to the macromolecular machines of gene expression. Annu Rev Bioph Biom, 2007,36:79-105
27 Rohs R, Jin X, West SM, *et al*. Origins of specificity in protein-DNA recognition. Annu Rev Biochem, 2010,79:233-269
28 Kalodimos CG, Biris N, Bonvin AM, *et al*. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science, 2004,305(5682):386-389
29 Williams LD, Maher LJ 3rd. Electrostatic mechanisms of DNA deformation. Annu Rev Bioph Biom, 2000,29:497-521
30 Patsialou A, Wilsker D, Moran E. DNA-binding properties of ARID family proteins. Nucleic Acids Res, 2005,33(1):66-80

(Received Sep. 29, 2014; revised Dec, 21, 2014)