Mouse MORC3 is a GHKL ATPase that localizes to H3K4me3 marked chromatin

Sisi Li[a,1,2], Linda Yen[b,1], William A. Pastor[b], Jonathan B. Johnston[c], Jiamu Du[d], Colin J. Shew[b], Wanlu Liu[e], Jamie Ho[b], Bryan Stender[b], Amander T. Clark[b,e,f,g], Alma L. Burlingame[c], Lucia Daxinger[h,3], Dinshaw J. Patel[a,3], and Steven E. Jacobsen[b,f,i,3]

[a]Structural Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065; [b]Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, CA 90095; [c]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94158; [d]Shanghai Center for Plant Stress Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 201602, China; [e]Molecular Biology Institute, University of California, Los Angeles, CA 90095; [f]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, CA 90095; [g]Jonsson Comprehensive Cancer Center, University of California, Los Angeles, CA 90095; [h]Department of Human and Clinical Genetics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands; and [i]Howard Hughes Medical Institute, University of California, Los Angeles, CA 90095

Microchidia (MORC) proteins are GHKL (gyrase, heat-shock protein 90, histidine kinase, MutL) ATPases that function in gene regulation in multiple organisms. Animal MORCs also contain CW-type zinc finger domains, which are known to bind to modified histones. We solved the crystal structure of the murine MORC3 ATPase-CW domain bound to the nucleotide analog AMPPNP (phosphoaminophosphonic acid-adenylate ester) and in complex with a trimethylated histone H3 lysine 4 (H3K4) peptide (H3K4me3). We observed that the MORC3 N-terminal ATPase domain forms a dimer when bound to AMPPNP. We used native mass spectrometry to show that dimerization is ATP-dependent, and that dimer formation is enhanced in the presence of nonhydrolyzable ATP analogs. The CW domain uses an aromatic cage to bind trimethylated Lys4 and forms extensive hydrogen bonds with the H3 tail. We found that MORC3 localizes to promoters marked by H3K4me3 throughout the genome, consistent with its binding to H3K4me3 in vitro. Our work sheds light on aspects of the molecular dynamics and function of MORC3.

X-ray crystallography | histone mark reader | ATPase | native mass spectrometry

The Microchidia (MORC) family of ATPase proteins has been shown to be an important regulator of gene silencing in multiple organisms. This family was first described in mice, when it was discovered that *morc1* null males showed arrested spermatogenesis (1). This arrest was later shown to be associated with transposon derepression, implicating murine MORC1 as a crucial mediator of transposon silencing (2). *Arabidopsis thaliana* MORC1 and MORC6 were shown to mediate silencing of transposons in a manner largely independent of changes in DNA methylation (3–5). Studies in *Caenorhabditis elegans*, which lack DNA methylation, also concluded that the single MORC gene in this organism plays a role in transgene silencing (4). Although the biological importance of MORC ATPases in enforcing gene silencing across multiple organisms is clear, how they are targeted and how they function are poorly understood.

The MORC ATPases share a similar domain arrangement. The N terminus contains a GHKL (gyrase, heat-shock protein 90, histidine kinase, MutL) type ATPase domain, and at the C terminus is a coiled-coil segment. MORCs have been reported to form functional homomultimers or heteromultimers, where multimerization is likely mediated by the N- and/or the C-terminal domains (4, 6, 7). The coiled-coil region has been proposed to promote constitutive dimerization, whereas N-terminal ATPase head dimerization occurs only on ATP binding (6). This is consistent with other GHKL ATPases described in the literature, many of which have been reported to undergo ATP-dependent dimerization (8–10). Both plant and animal MORCs are capable of forming nuclear bodies, and mutations that impair ATP binding and/or hydrolysis disrupt nuclear body formation of human MORC3 (6).

Animal MORCs also carry a CW-type zinc finger domain, which has been proposed to read histone H3 lysine 4 (H3K4)

dimethylation and trimethylation marks (H3K4me2 and H3K4me3) (11–13). Consistent with this hypothesis, human and murine MORC3 have been identified in mass spectrometry screens as H3K4me3 readers (12–14). Mutation of a critical tryptophan residue in the human MORC3 CW domain disrupts localization in the nucleus, suggesting that CW domain-mediated recognition of H3K4me3 is also critical for MORC3 targeting to chromatin (6).

Here we report a 2.6-Å crystal structure of the MORC3 (ATPase-CW cassette)–AMPPNP (phosphoaminophosphonic acid-adenylate ester)–H3K4me3 complex, which is a symmetrical dimer. We show that MORC3 is an active ATPase that forms ATP-dependent dimers and uses the CW domain to make contacts with the H3K4me3 peptide. Arg8 from the histone peptide is anchored by hydrogen bonding in a pocket on the CW domain, and the trimethylated Lys4 side chain is engaged via an aromatic cage. In vivo, MORC3 localizes to H3K4me3-marked chromatin. These observations reveal molecular details of MORC action that likely

Significance

The Microchidia (MORC) family of ATPases are important regulators of gene silencing in multiple organisms but little is known about their molecular behavior. In this study, we used crystallography and native mass spectrometry to show that MORC3 forms dimers when it binds to nonhydrolyzable ATP analogues. We also determined that the CW zinc finger-like domain of MORC3 can bind euchromatic histone H3 lysine 4 (H3K4) methylation and that MORC3 localizes to H3K4me3-marked chromatin. The MORC3 crystal structure provides details as to the intermolecular interactions that allow dimerization and the binding to ATP and histones. This work reveals key molecular activities of MORC3 that might apply to other MORC family members in eukaryotic organisms.

are conserved in other eukaryotic MORC proteins, and suggest an epigenetic mark associated with MORC3 localization in vivo.

Results

Overall Structure of the MORC3–AMPPNP–H3(1–15)K4me3 Complex.

MORC3 is composed of an N-terminal ATPase domain followed by a CW-type zinc finger, a flexible linker segment without predictable secondary structure, and a C-terminal coiled-coil domain (Fig. 1A). The CW-type zinc finger has been identified as an H3K4me reader module as assayed by structural and biochemical methods (11–13, 15). To gain insight into the molecular mechanism of MORC3, we solved the crystal structure of the MORC3 ATPase-CW domain cassette in complex with the nonhydrolyzable ATP analog AMPPNP and H3(1–15)K4me3 peptide by single-wavelength anomalous diffraction. The structure of the complex (Fig. 1A) was refined to 2.6-Å resolution, yielding an R factor of
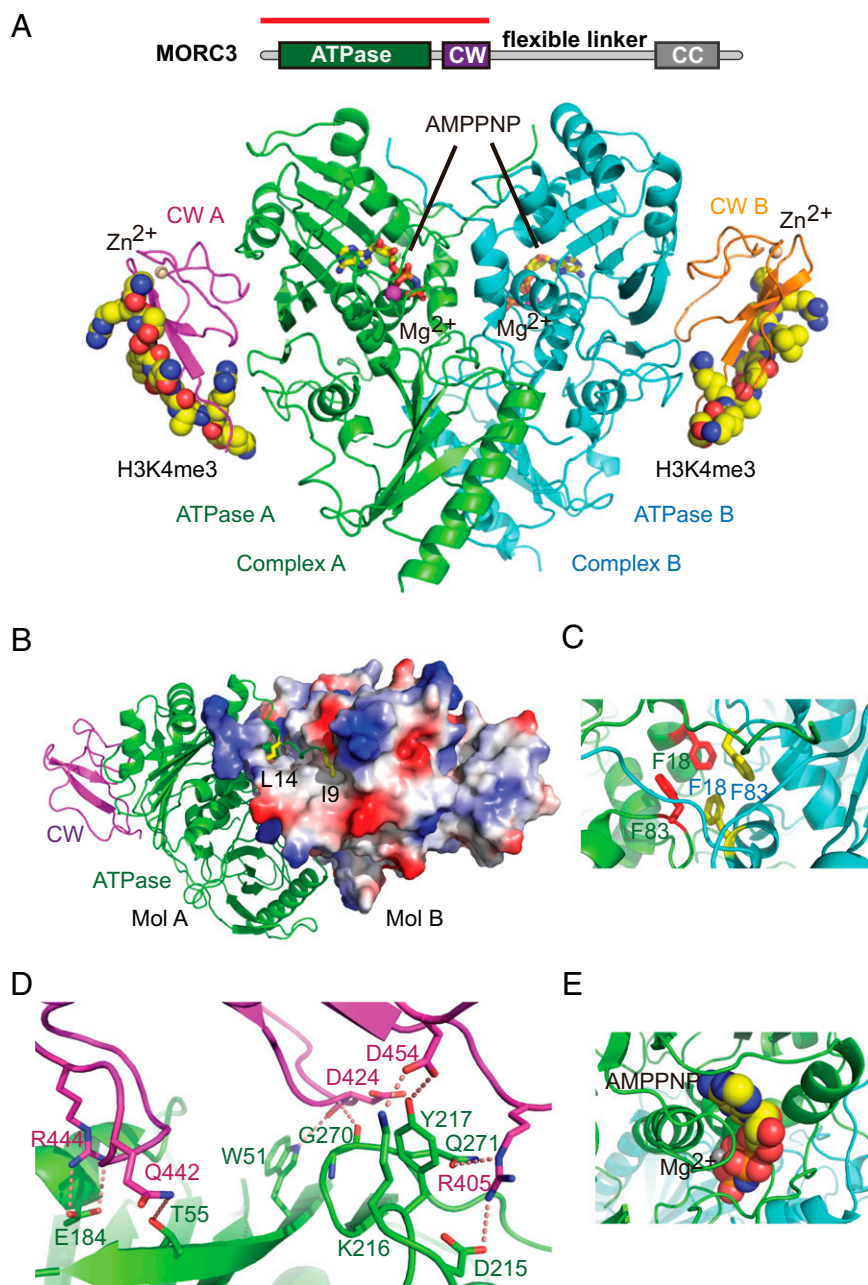
Fig. 1. Structure of MORC3 ATPase-CW cassette in complex with AMPPNP and H3K4me3 peptide. (A) Overall crystal structure of MORC3–AMPPNP–H3K4me3 complex. The crystallized region is delineated by the red line. The ATPase and CW domains are shown in ribbon representation and colored green and magenta in monomer Mol A and cyan and orange in monomer Mol B, respectively. The bound peptides are shown in space filling representation, whereas the AMPPNP molecules are in stick representation. (B) The MORC3 dimer shown with one monomer in ribbon representation and the other monomer in electrostatic surface representation, demonstrating the extensive interactions between the two monomers. The N terminus of one monomer interacts with the other monomer, with Ile9 and Leu14 residues positioned within two small hydrophobic pockets. (C) Two aromatic residues, Phe18 and Phe83, of one monomer pack against the same residues in the other monomer, forming a hydrophobic core of the dimer interface. (D) The interactions between the ATPase domain and the CW domain. The hydrogen bonds are highlighted with dashed red lines. (E) The AMPPNP fits into a narrow binding pocket inside the ATPase domain. AMPPNP is shown in space-filling representation.

Li et al.

21.8% and a free R factor of 23.0% (Table S1). A symmetric dimer of the MORC3 ATPase-CW domain cassettes (designated Mol A and Mol B) was observed in the asymmetric unit in the structure of the complex, with each monomer in the complex bound to an AMPPNP and an H3K4me3 peptide (Fig. 1A). Overall, we observed well-defined electron density in the structure of the complex, except for some loop segments within the ATPase domain (residues 225–233, 239–242, and 334–337 in Mol A and residues 225–233 and 240–241 in Mol B) and the linker between ATPase and CW domains (residues 387–403 in both Mol A and Mol B).

The MORC3 CW domain structure in the complex resembled the fold reported for human CW-zinc finger and PWWP domain-containing protein 1 and *Arabidopsis* ASH1 HOMOLOG2 (11, 15). The structure of the CW domain as part of the MORC3 ATPase-CW domain cassette also agreed with the recently described structure of an isolated MORC3 CW domain in complex with H3K4me3 peptide (rmsd ~0.7 Å) (PDB ID code 4QQ4) (16). Overall, the CW fold featured an N-terminal β-hairpin, with two cysteine residues projecting from the β-hairpin combining with two additional nearby cysteine residues to form a zinc finger.

The MORC3 ATPase Domain Adopts a GHKL ATPase Fold and Forms a Dimer in the Complex. The MORC3 ATPase domain adopted an ATPase fold, similar to those observed in other GHKL domain structures such as the HSP90–ATP–Sba1 complex (PDB ID code 2CG9) (17). Alignment of the monomeric MORC3 ATPase domain fold with the HSP90 ATPase domain fold showed an rmsd of 2.8 Å for 241 aligned Cα atoms despite low sequence conservation (~16%). The MORC3 ATPase domains formed a conserved twofold symmetric dimer in the complex, with a large interface that buried ~2,334 Å^2 of surface area for each monomer, a value similar to that observed for other GHKL dimeric ATPase domain structures. The dimer interface showed extensive and symmetric interactions between the two monomers, such as the N-terminal end of one ATPase domain (residues 8–15) bound within a surface cleft of the other ATPase domain (Fig. 1B).

Notably, Ile9 and Leu14 of one monomer extended their side chains into two small hydrophobic pockets located on the other monomer (Fig. 1B). The aromatic side chains of Phe18 and Phe83 from one monomer formed hydrophobic contacts with the same residues from the other monomer, resulting in an interdigitated hydrophobic core that contributes to the stability of the dimer (Fig. 1C). The stability of the dimeric interface was further augmented by additional hydrophobic and hydrophilic interactions. The CW domain was positioned along one side of the ATPase domain, forming an interface with an average of 704 Å^2 buried surface area by each component. Extensive salt bridges and hydrogen-bonding interactions were involved in the interaction between ATPase and CW domains along this interface (Fig. 1D). The ATP analog AMPPNP fit into the active site of each monomer and formed extensive interactions with the highly conserved residues involved in ATP binding and hydrolysis (Fig. 1E and Fig. S1).

The MORC3 ATPase Domain Forms a Dimer on ATP Binding. In the cocrystal structure, we observed that the ATPase domain adopted a canonical GHKL ATPase fold, with each ATPase domain bound by a molecule of AMPPNP-Mg^{2+}. The ATPase domains formed a symmetrical dimer in the crystal structure, consistent with the observation that GHKL ATPases dimerize on ATP binding (10, 18) Given that the MORC ATPase domain harbored residues that were highly conserved across other GHKL ATPase domains, we hypothesized that this dimer was physiologically relevant and likely induced by the presence of AMPPNP.

To confirm that the AMPPNP-bound dimer could be observed, we analyzed the ratio of MORC3 (7-456) monomer and dimer in the presence of nucleotide by native mass spectrometry. We also validated these results by protein cross-linking followed by SDS/PAGE. MORC3 (7-456) analyzed without added ligand ionized in a manner most consistent with the protein bound to one zinc atom per monomer (Fig. 2 and Table S2). The observed zinc binding was consistent with both our crystal structure and previous reports that CW domains use four cysteine residues to coordinate a single zinc atom (Fig. 1A) (15). With each protein and under all conditions, we observed two distinct charge envelopes in the native mass spectrum within the region of 3,000–5,500 m/z. The first charge envelope (~3,000–4,200 m/z) corresponded to monomer bearing between +13 to +16 charges. The second charge envelope (~4,200 to ~5,500 m/z) corresponded to dimers, representing the +20 to +24 charge states of the dimer (Fig. 2A). The MORC3 (7-456) holoenzyme showed a ratio of roughly 70% monomer/30% dimer without added ligand (Fig. 2 and Table S2). Incubation of substrate (ATP-Mg^{2+}) resulted in a small shift in the equilibrium ratio to 59% monomer/41% dimer, similar to that of the ADP-Mg^{2+} condition, where we observed a ratio of 64% monomer/36% dimer. Presumably, the rapid hydrolysis of ATP resulted in ADP-Mg^{2+} being bound in the active site (product-like), explaining why we observed both ADP and ATP conditions producing similar monomer/dimer ratios. This interpretation is further supported by the observation that MORC3 (7-456) hydrolyzed ATP very efficiently under these conditions (Fig. S2A). Furthermore, protein cross-linking experiments showed that the population of dimers increased on incubation with ATP or AMPPNP compared with the protein without nucleotide (Fig. S2B). Our native mass spectrometry analysis also revealed that when the nonhydrolyzable ATP analogs AMPPNP and ATP-γ-S were incubated with MORC3 (7-456), the ratio shifted strongly, such that there was now ~90% dimer (Fig. 2 and Table S2). These results strongly support the view that ATP binding induces MORC3 dimer formation.

To further investigate the ligand-dependent dimerization, we examined the native mass spectra for evidence of ligand-enzyme complexes in the gas phase. We confirmed that one molar equivalent of ATP-Mg^{2+} could be bound to the monomer and that both one and two molar equivalents were bound to the dimer (Table S3). With the nonhydrolyzable analogs, we observed two molar equivalents of each analog binding to the dimer form (Fig. S3). With AMPPNP, we observed the expected one molar equivalent of AMPPNP-Mg^{2+} bound per monomer, but for ATP-γ-S-Mg^{2+}, we did not detect a monomer-ligand complex. The observed dimer bound to two molar equivalents of AMPPNP-Mg^{2+} was consistent with the crystal structure, and demonstrates that these interactions could be captured in the gas phase.

From the structure, we identified Ile9 as a key residue that could provide hydrophobic contact between monomer units and help anchor the dimer interface (Fig. 1B). A similar isoleucine residue was previously found to significantly contribute to dimer stabilization in other GHKLs (9). We mutated Ile9 to an alanine (I9A) and analyzed dimer formation in the presence of nucleotide by native mass spectrometry. Consistent with a functional role in mediating monomer–monomer contacts and stabilizing the dimer, abolishing this contact severely attenuated the extent of dimer formation despite being able to bind nucleotide (Fig. 2B, Figs. S2B and S4, and Tables S4 and S5). The I9A mutant incubated with Mg^{2+} showed only monomeric protein without any detectable dimer; however, when substrate (ATP-Mg^{2+}) was incubated with the I9A mutant, the dimer was now present (16%), albeit at a lower level than with the wild-type protein (31%). Presumably, the energetic contributions stabilizing the dimer interface are weak and numerous and include additional contributions from Phe18 and Phe83 (Fig. 1 B and C). Nonetheless, I9A dimer formation was not as productive as with the wild-type. The nonhydrolyzable analog ATP-γ-S-Mg^{2+} stabilized the dimer to the greatest degree (35%) in our native mass spectrometry assay,
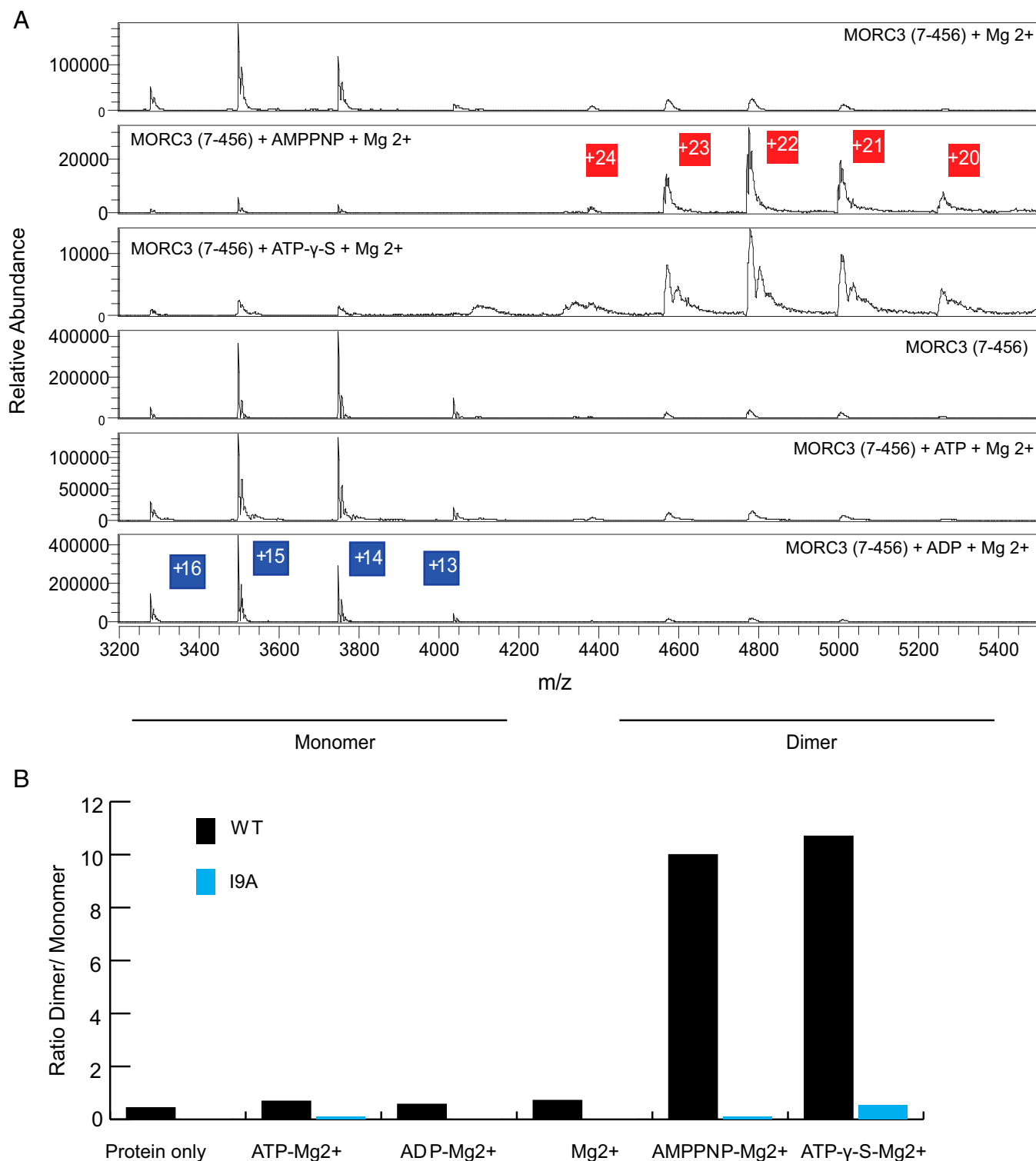
Fig. 2. The N terminus of MORC3 forms ATP-dependent dimers. (*A*) Native mass spectra (*m/z* vs. relative abundance) of MORC3 (7-456) under various conditions. The charge envelope on the left corresponds to the MORC3 monomer at +16, +15, +14, and +13 ions and to the MORC3 (7-456) dimer at +24, +23, +22, +21, and +20 ions. (*B*) Ratio of MORC3 (7-456) distribution of monomers and dimers in the presence of nucleotide and nonhydrolyzable nucleotide analogs as determined by native mass spectrometry.

whereas AMPPNP-Mg^{2+} and ATP-Mg^{2+} resulted in dimer detection of only 10%. Although dimer formation was still observed, the proportion was significantly lower than that formed by the wild-type protein (~90%). This attenuation in productive dimer formation was confirmed by protein cross-linking (Fig. S2*B*).

The I9A monomer and dimer were found in complex with one and two molar equivalents of nucleotide, respectively (Fig. S4 and Table S5), suggesting that the I9A mutation has minimal effect on nucleotide binding. The various charge states observed and their relative ratios across the wild-type and mutant enzymes

Li et al.

also suggest that the wild-type and I9A proteins have similar overall structures (Fig. 2A and Fig. S4). We observed stoichiometric binding of ligand independent of productive dimer formation, consistent with the model in which the I9 residue is largely responsible for stabilizing the dimer interface rather than contributing to nucleotide binding.

Recognition of the H3K4me3 Peptide by the MORC3 CW Domain. The H3(1–15)K4me3 peptide adopted a β-strand–like conformation that aligned along one side of the β-hairpin of the CW domain, forming a continuous three-stranded β-sheet (Fig. 3A). The ATPase domain formed minimal contacts (as small as 102 Å² buried on each side) with the bound peptide, indicative of a weak or packing-induced but not biologically relevant interaction, as predicted by the PISA server (19). The peptide was bound within a negatively charged surface of the CW domain (Fig. 3 B and C).

Along with main chain hydrogen-bonding interactions, which mediated the three-stranded β-sheet formation between the peptide and the β-hairpin of the CW domain (Fig. S5A), there were additional side chain interactions involved in the specific recognition of the bound H3K4me3 peptide (Fig. 3 B and D). The amino group of Ala1 from the peptide was anchored within a small surface pocket, forming two hydrogen bonds with the

main chain carbonyls of Pro430 and Glu431 of the CW domain. This interaction allowed the CW domain to specifically recognize the H3 peptide N terminus (Fig. 3D). The trimethyl-lysine side chain of Lys4 was positioned within an aromatic-lined surface groove pocket formed by the Trp410 and Trp419 of the CW domain and stabilized by the cation–π interactions (Fig. 3E), similar to other classical methyl-lysine recognition modules (20). In addition, Thr6 of the peptide formed one hydrogen bond with Trp410 of the CW domain. Most strikingly, the guanadinium side chain of Arg8, which was not identified as important in the recognition of other reported CW domains (15, 21), formed extensive hydrogen bonds and salt bridge interactions with the CW domain (Fig. 3F). The side chain of H3R8 inserted into a deep pocket of the CW domain and formed a salt bridge interaction with Asp424 and hydrogen bonds with Pro406 and Gln408 (Fig. 3F).

Structure of the MORC3–AMPPNP–H3 Complex. To evaluate the importance of trimethylated Lys4 for recognition, we also determined the crystal structure of MORC3 in complex with AMPPNP and an unmodified H3(1–32) peptide. The structure was solved using the molecular replacement method using the MORC3–AMPPNP–H3K4me3 complex as a model and refined to 2.9-Å resolution, yielding an R factor of 22.0% and a free R factor of 23.6%
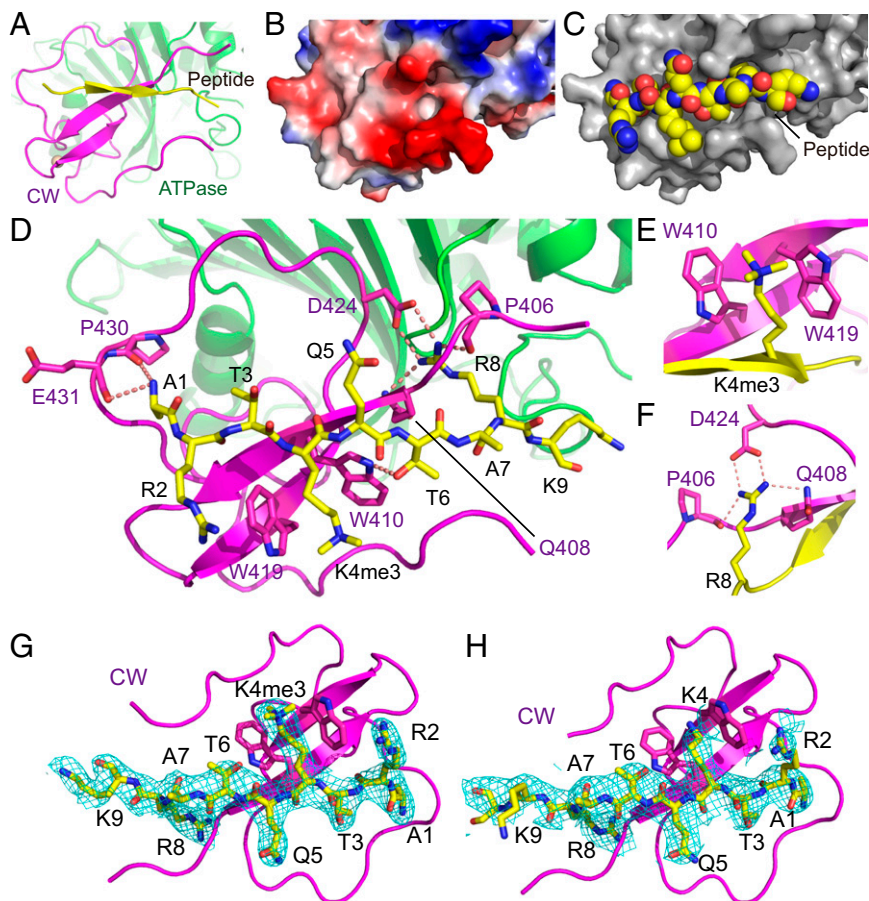

Fig. 3. Intermolecular interactions between MORC3 and the H3K4me3 peptide. (A) Overall intermolecular interactions between H3K4me3 peptide (yellow) and the MORC3 CW domain. The peptide forms a β-strand–like conformation, resulting in generation of a three-stranded β-sheet with the β-hairpin of the CW domain. (B) Electrostatic surface view of the MORC3 CW domain where it binds the H3K4me3 peptide. The domain contains a pocket to accommodate Arg8 of the peptide. (C) Surface view of the H3K4me3 peptide (space-filling representation) in complex with MORC3 (surface representation in silver). Ala1 is anchored in a small pocket, K4me3 binds within a tryptophan-lined surface cleft, and Arg8 inserts into a deep pocket of the CW domain as noted in B. (D) Outline of intermolecular interactions between the bound H3(1–15)K4me3 peptide and the CW domain in the complex. (E) K4me3 inserts into a partial aromatic surface groove lined by Trp410 and Trp419 in the complex. (F) Anchoring of the Arg8 guanidinium group in a deep pocket through a network of hydrogen-bonding interactions with side chains from the CW domain. (G and H) 2Fo–Fc electron density maps at 1σ for the bound peptide in the complex containing H3(1–15)K4me3 peptide (G) and the complex containing H3(1–32)K4 peptide (H).
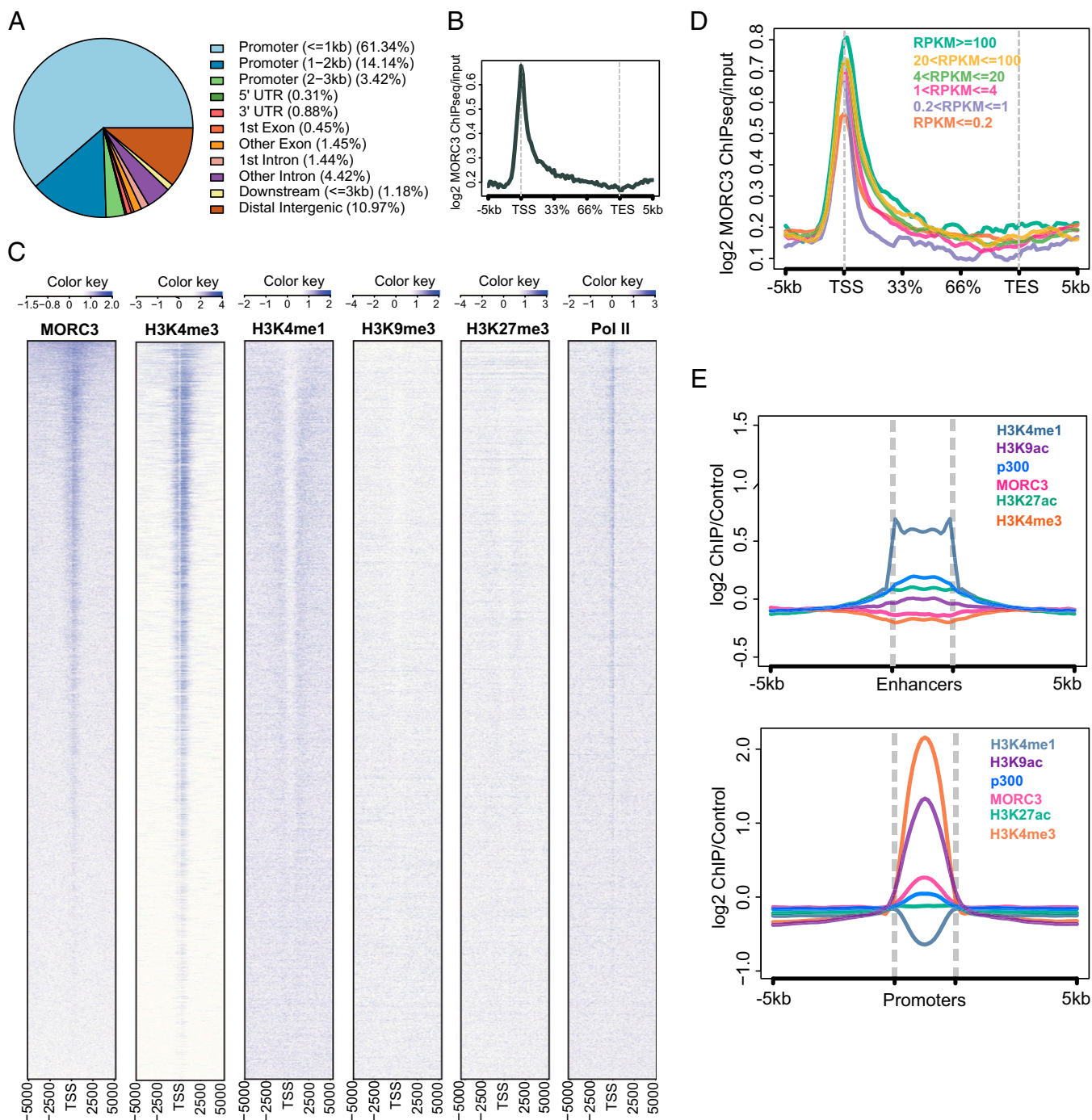
www.manaraa.com

Fig. 4. MORC3 is localized to active gene promoters in vivo. (*A*) Pie chart showing the distribution of MORC3 peaks at promoters, exons, introns, and intergenic regions. (*B*) Metaplot showing MORC3 distribution mapped over all protein-coding genes. (*C*) Heat map of histone modifications over all protein-coding gene promoters ranked by MORC3 abundance. Distribution is mapped over −5,000 to 5,000 base pairs surrounding the TSS. Log$_2$ fold change is indicated in the color key. (*D*) MORC3 enrichment at a gene is correlated to gene expression level. Shown is MORC3 distribution analyzed over a gene unit as binned by gene expression level. (*E*) MORC3 is enriched at active promoters but not enhancers. Shown are metaplots of MORC3, selected histone marks, and p300 over enhancers (H3K4me1$^+$ H3K4me3$^-$) or promoters (H3K4me3$^+$) and surrounding regions.

(Table S1). Overall, the structure of MORC3–AMPPNP–H3 complex was nearly identical to the structure of MORC3–AMPPNP–H3K4me3 complex with an rmsd of only 0.27 Å for 414 aligned Cα atoms. The bound unmodified H3(1–32) showed a similar β-strand–like conformation as was observed for the bound H3K4me3 peptide. However, the electron density of the unmodified K4 side chain was poorer than that observed for

the K4me3 side chain, indicating that the position of K4me3 can be stabilized through the cation–π interaction with the CW domain (Fig. 3 *G* and *H*).

The CW Domain of MORC3 Recognizes the H3K4me3 Mark. We performed in vitro peptide pull-down experiments to confirm that the MORC3 CW domain could bind preferentially H3K4me3

over unmodified H3 peptide tail (Fig. S5B). The MORC3 CW domain exhibited a significant preference for H3K4me2 and H3K4me3 over H3K4me0, consistent with previous reports of human MORC3 binding preferences (12, 13).

We used isothermal titration calorimetry (ITC) to calculate the binding affinity of the MORC3 CW domain for both unmodified H3K4me0 and H3K4me3 H3 peptides (Fig. S5C). We used H3K9me3 peptide as a negative control, because it appeared to be anticorrelated with MORC3 in the genome (14). The MORC3 CW domain bound to the H3K4me3 peptide with a high affinity of 0.49 μM (Fig. S5C). In contrast, it bound to the unmethylated H3K4 peptide with sixfold lower affinity (K_d = 2.8 μM) compared with the H3K4me3 peptide, revealing that the methylation modification of H3K4 is essential for achieving high binding affinity. Notably, the MORC3 CW domain bound very weakly, if at all, to the H3K9me3 mark (K_d >200 μM) (Fig. S5C), suggesting that binding was methylated K4 site-specific.

MORC3 Is Enriched Over H3K4 Trimethylation Sites Genome-Wide. To validate the significance of MORC3 H3K4me3 binding in vitro, we profiled MORC3 localization in mouse embryonic stem cells (mESCs) by chromatin immunoprecipitation (ChIP) sequencing (ChIP-seq) using an endogenous MORC3 antibody (Fig. S6A). We normalized ChIP signals to their respective inputs to define sites of enrichment as MORC3 peaks, and found that these peaks frequently overlapped with H3K4me3 (Fig. S6B). We called 14,032 MORC3 peaks, of which 11,071 were located in promoters (Fig. 4 A and B). We compared MORC3 peaks with ENCODE (Encyclopedia of DNA Elements) datasets for various histone modifications or genomic features, and plotted heat maps of ChIP-seq signals over transcription start sites (TSSs). When we ranked MORC3 sites from highest to lowest enrichment and compared them with corresponding H3K4me3 intensity, the heat maps exhibited a similar pattern, suggesting that most, if not all, MORC3 peaks were colocalized with H3K4me3 (Fig. 4C). In contrast, we observed a strong anticorrelation between MORC3 signals and silencing features, such as H3K9 trimethylation (H3K9me3) (Fig. 4C). Further analysis indicated that ∼94% of the defined MORC3 peaks exhibited overlap with H3K4me3 ChIP-seq signals, indicating a strong correlation of MORC3 peaks with H3K4me3 sites in vivo. Examination of Pol II at these loci revealed a slightly increased intensity for genes with the highest MORC3 signals, confirming that the top MORC3 sites consist of more active genes (Fig. 4C).

We also analyzed the relationship between levels of gene expression and MORC3 enrichment. Using existing RNAseq datasets, we binned genes according to their expression levels, then mapped and analyzed their levels of MORC3 enrichment over genes (Fig. 4D). Consistent with the heat maps (Fig. 4C), MORC3 was somewhat more abundant at genes with higher expression and somewhat lower at genes with lower expression; however, MORC3 enrichment was consistently found at the TSS for all expression categories.

In addition to H3K4me3, such features as p300, H3K27 acetylation (H3K27ac), and H3K9 acetylation (H3K9ac), which are found in the promoters of active genes (22), were found together with MORC3 at promoters (Fig. 4E). It is unlikely that MORC3 is attracted by these marks, however, given that H3K4me1$^+$ H3K4me3$^-$ enhancers, which also feature H3K9 and H3K27 acetylation, are devoid of MORC3 (Fig. 4E) (23). Taken together, these findings suggest that MORC3 is attracted by H3K4me3 sites to the promoters of active genes.

Discussion

MORC proteins are involved in gene regulation in a number of eukaryotic species. However, the mechanism of MORC action is very poorly understood. All MORC proteins contain a GHKL ATPase domain with high conservation of residues known to be involved in ATP binding and hydrolysis in other well-studied GHKL

ATPases (18). Our structural data confirm that the MORC3 ATPase domain adopts a GHKL fold with the necessary residues for it to function as an active ATPase, and our in vitro assays show that it can hydrolyze ATP. The crystal structure shows that the MORC3 N-terminal ATPase domain is a dimer in complex with the nonhydrolyzable ATP analog AMPPNP. Furthermore, using native mass spectrometry, we observed that stable dimer formation is nucleotide-dependent.

The population distributions of monomers and dimers obtained by incubating MORC3 with no additional ligand, ADP, or ATP were similar, consistent with the observation that MORC3 can hydrolyze ATP very efficiently under native mass spectrometry conditions. However, the addition of a nonhydrolyzable ATP analog was sufficient to shift the equilibrium toward a predominantly dimeric population. These results suggest that MORC3 is likely to function similarly to other GHKL ATPase-containing proteins, using ATP for dimer formation and ATP hydrolysis for dimer dissolution. It is also likely that other MORCs similarly use dimer formation and dissolution in their modes of action. Our data are consistent with an earlier proposal that MORC3 may act as a molecular clamp for DNA, with constitutive dimerization through the C-terminal coiled-coil domain and ATP-dependent dimerization through the ATPase domains (6). Indeed, bacterial GHKL ATPases, such as topoisomerase VI and gyrase, can trap DNA through dimerization of their ATPase domains. In this way, MORCs may act in a manner analogous to structural maintenance of chromosomes (SMC) proteins, such as condensins and cohesins, which use topological trapping of DNA in their mechanisms of action (24, 25).

Whether there is a universal mechanism by which MORCs are recruited to chromatin is unclear. The mammalian and *C. elegans* MORC proteins contain a CW domain that is lacking in the plant MORC proteins. In addition, both the mammalian MORC3 and MORC4 CW domains have been reported to bind to H3K4 methylation in vitro (12, 14–16), and in the present study we confirm that MORC3 binds preferentially to H3K4 trimethylation. Our structure of the MORC3 N-terminal ATPase-CW cassette in complex with H3K4 trimethylated peptides shows that the methyl group is recognized by a typical methyl-binding aromatic cage. In addition, our ChIP analysis shows that MORC3 is localized to H3K4 trimethylation marked chromatin in vivo. These results suggest that MORC3 uses H3K4 methylation to guide its localization to specific regions of chromatin. How the other MORC proteins are directed to chromatin, especially those lacking CW domains, remains a mystery. It was previously shown that MORC3 forms punctate nuclear bodies that appear to be ATP-dependent, and that loss of function in either the CW domain or ATPase domain is sufficient to abrogate body formation (6). These results are consistent with the hypothesis that both CW domain-mediated interactions with H3K4me3 and ATPase domain-mediated interactions with DNA are required for stable localization of MORC3 to chromatin.

Although the function of MORC3 remains unknown, our present results shed light on the molecular aspects of MORC3 protein domains. Because all MORC proteins share an overall domain architecture of an N-terminal GHKL ATPase domain and a C-terminal coiled-coil domain, future work aimed at determining the precise molecular mechanism of MORC3 should provide insight into the action of MORC in other eukaryotes as well.

Experimental Procedures

Protein Expression and Purification. MORC3 (7-456) was expressed in *Escherichia coli* strain BL21(DE3) recombinant inbred line (Stratagene) in a self-modified vector with an N-terminal His-yeast sumo tag. Protein expression was induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside, followed by incubation at 17 °C overnight. The recombinant expressed protein was purified using a nickel affinity column (GE Healthcare). After cleavage by Ulp1 protease, tags were removed in a second step using a nickel affinity column (GE Healthcare). The target protein was further purified using a heparin column and a Superdex G200 gel filtration column (GE Healthcare). MORC3 398–456 was cloned and expressed using the same strategy, except

that it was fused to a His-GST N-terminal tag. The CW domain protein was purified using a nickel affinity column (GE Healthcare) and a Superdex G75 gel-filtration column (GE Healthcare). The purified protein was concentrated to 20 mg/mL and stored at −80 °C. The peptides used in the experiment were ordered from Genscript (Nanjing) and GL Biochem (Shanghai).

Crystallization, Data Collection, and Structure Determination. Before crystallization screening, the purified MORC3 protein in the presence of 2 mM $MgCl_2$ was mixed with AMPPNP and H3(1–15)K4me3 or H3(1–32) peptides with a molar ratio of 1:4:4 and then incubated at 4 °C for 1 h. Crystallization was carried out by the sitting-drop vapor diffusion method at 20 °C. The MORC3 in complexes with AMPPNP and H3(1–15)K4me3 or H3(1–32) peptides were crystallized under a condition of 5% (vol/vol) ethanol, 5% (vol/vol) 2-methyl-2,4-pentanediol, and 0.1 M Hepes-Na, pH 7.5. To obtain the heavy atom derivative, the MORC3–AMPPNP–H3(1–15)K4me3 crystal was soaked in the reservoir solution supplemented with 10 mM ethylmercurithiosalicylic acid at 20 °C for 2 h. The crystals were cryoprotected with the reservoir solution supplemented with 10% glycerol, followed by flash-freezing in liquid nitrogen. The diffraction data were collected at Shanghai Synchrotron Radiation Facility beamlines BL17U1 and BL19U and were processed with HKL2000 (26). The mercury derivate single-wavelength anomalous diffraction (SAD) data were used for phasing using Phenix (21). The model was built using the program Coot (27) and refined using Phenix (21). All of the molecular graphics were generated with PyMol (DeLano Scientific) and LigPlus (28). Diffraction data and structure refinement statistics are summarized in Table S1.

Native Mass Spectrometry. Native mass spectrometry was carried out using an Exactive Plus EMR mass spectrometer (Thermo Fisher Scientific). Before analysis, samples were buffer-exchanged into 150 mM ammonium acetate (AmAc), pH 7.5 using MicroBioSpin6 columns (Bio-Rad), and then sprayed from borosilicate capillaries (NanoES spray capillaries, borosilicate; Thermo Fisher Scientific) at flow rate of 5–40 μL/min. The instrument was calibrated in the extended mass range using a 5 mg/mL solution of CsI prepared in water. Protein samples were analyzed in positive ion mode. The experimental parameters were optimized for each sample, but were generally as follows: spray voltages, 0.8–1.5 kV; injection flatapole, 5; interflatapole lens, 5; bent flatapole, 5; transfer multipole, +4 to −4; C-trap entrance lens, −10 to +10; source DC offset, 25 V; fragmentation collision energy, 20–150 and collision-induced dissociation, 5–150; injection times, 50–200 μs; trapping gas pressure, 7.5; resolution, 17,500 arbitrary units; mass range, 500–20,000 m/z; capillary temperature, 250 °C; S-lens RF value was set to 200 V; microscans, 10; and automatic gain control was set to 1e⁶. Nucleotide-Mg complexes were prepared fresh by incubating equimolar ratios of nucleotide and $MgCl_2$ for 30 min at 4 °C and then freshly diluting to the final working concentration. Protein (~0.5–1 μM) was incubated with excess nucleotide (20 μM) at room temperature for 30 min before analysis.

Native Mass Spectrometry Data Analysis. Mass spectra were analyzed using MagTran (29) and PeakSeeker (30) by converting the m/z values to charge state envelopes. The zero-charge state of each species was calculated from the charge state envelopes. Average protein masses were calculated from the primary sequence using the ExPASy ProtParam tool (web.expasy.org/protparam).

ChIP of MORC3. V6.5 embryonic stem cells were passaged on mouse embryonic fibroblasts in serum + leukemia inhibitor factor until a final split, when they were expanded on gelatin for ChIP. Passage 28 V6.5 cells were harvested by trypsinization. Trypsin was quenched with medium containing FBS, and the cells were washed with medium and 1× PBS. Cells were fixed by treatment with 1% formaldehyde and 1× PBS for 10 min at room temperature, then quenched by the addition of glycine to a final concentration of 0.14 M, followed by incubation for another 10 min. Cells were washed with 1× PBS, then aliquoted and flash-frozen.

Ten million cells per replicate were thawed on ice and resuspended in 1 mL 10 mM Tris pH 8.0, 0.25% Triton X-100, 10 mM EDTA, 0.5 mM EGTA, and 1 mM PMSF and then incubated with rotation for 15 min at room temperature. Nuclei were pelleted by centrifugation at 1,500 × g for 5 min at 4 °C. The nuclei were resuspended in 10 mM Tris pH 8.0, 200 mM NaCl, 10 mM EDTA, 0.5 mM EGTA, and 1 mM PMSF, incubated for 10 min at room temperature with rotation, and then centrifuged again. Nuclei were then resuspended in 10 mM Tris pH 8.0, 10 mM EDTA, 0.5 mM EGTA, 0.1% SDS, and 1 mM PMSF, and disrupted by sonication at high intensity using a Bioruptor (Diagenode).

Sonicated lysate was cleared by centrifugation at 16,000 × g for 10 min, and the supernatant was used for ChIP. Samples were diluted with an equal

volume of 16.7 mM Tris pH 8.0, 0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, and 167 mM NaCl. The samples were then precleared with 30 μL of protein A magnetic Dynabeads (Thermo Fisher Scientific), which had been washed with 16.7 mM Tris pH 8.0, 0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, and 167 mM NaCl before use, followed by incubation for 2 h at 4 °C. The beads were then collected on a magnet, and the supernatant was retained. Once 10% of the sample was saved for input, the sample was then split into two halves. One half was treated with 1 μL of rabbit IgG (Active Motif 27478), and the other half was incubated with 1 μL of anti-MORC3 antibody (100-401-N97; Rockland). Samples were incubated overnight at 4 °C with rotation.

The next day, 60 μL of protein A beads, which had been washed with 16.7 mM Tris pH 8.0, 0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, and 167 mM NaCl before use, were added to each sample, followed by incubation for another 2 h. The beads were washed twice for 4 min each time with rotation with 50 mM Hepes pH 7.9, 1% Triton X-100, 0.1% deoxycholate, 1 mM EDTA, and 140 mM NaCl; washed twice for 4 min each time under rotation with 50 mM Hepes pH 7.9, 0.1% SDS, 1% Triton X-100, 0.1% deoxycholate, 1 mM EDTA, and 500 mM NaCl; and then washed twice for 4 min each time under rotation with 500 μL of 10 mM Tris pH 8.0 and 1 mM EDTA. The purified DNA was eluted by incubation with elution buffer (100 μL of 50 mM Tris pH 8.0, 1 mM EDTA, and 1% SDS) at 65 °C for 10 min. Eluent was collected on a magnetic rack, and the beads were resuspended with 150 μL of elution buffer and then incubated at 65 °C for 10 min. The two eluents were pooled and de–cross-linked by incubation at 65 °C overnight.

The samples were brought to room temperature and then warmed to 37 °C, followed by incubation with 10 μg RNase A. The samples were then treated with 15 μg of proteinase K and incubated for 2 h at 56 °C. Finally, the samples were purified with Qiagen MinElute columns. Purified DNA was quantified with Qubit High-Sensitivity reagent (Thermo Fisher Scientific), and libraries were generated with the Ovation Ultralow Library System Kit (Nugen) using 10 ng of input DNA.

Analysis of ChIPseq Data. Raw sequencing reads were mapped to the mm9 genome with Bowtie (31) by allowing up to two mismatches and keeping only uniquely mapped reads. PCR duplicated reads were removed with SAMtools (32). Heatmaps and metaplots were generated in R with the ngs.plot package (33). MORC3 ChIPseq peaks were identified with MACS version 2.1.0 (34). Defined MORC3 peaks were then annotated to the mm9 genome with the ChIPseeker R package (35). For H3K4me1 (GSM769009), H3K9me3 (GSM1000147), H3K27me3 (GSM1000089), H3K27ac (GSM1000099), H3K4me3 (GSM769008), H3K9ac (GSM1000127), p300(GSM918750), and RNA polymerase II (GSM723019) ChIPseq, data raw sequencing data were downloaded from ENCODE and processed as described above.

Identification of promoters relied exclusively on the presence of H3K4me3 as described previously (23, 36, 37). Enhancer lists were identified as done previously (23, 37, 38) by considering the presence of H3K4me1 but the absence of H3K4me3.

Analysis of ENCODE RNAseq Data. For mESC RNAseq (GSM929718), aligned data were downloaded from ENCODE. The number of reads mapping to genes (mm9) were calculated by HTseq (39) with default parameters. Expression levels were determined by reads per kilobase of exons per million aligned reads.

Li et al.

BIOCHEMISTRY

1. Watson ML, et al. (1998) Identification of MORC (*Microrchidia*), a mutation that results in arrest of spermatogenesis at an early meiotic stage in the mouse. *Proc Natl Acad Sci USA* 95(24):14361–14366.
2. Pastor WA, et al. (2014) MORC1 represses transposable elements in the mouse male germline. *Nat Commun* 5:5795.
3. Lorković ZJ, Naumann U, Matzke AJ, Matzke M (2012) Involvement of a GHKL ATPase in RNA-directed DNA methylation in *Arabidopsis thaliana*. *Curr Biol* 22(10):933–938.
4. Moissiard G, et al. (2012) MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* 336(6087):1448–1451.
5. Brabbs TR, et al. (2013) The stochastic silencing phenotype of *Arabidopsis* morc6 mutants reveals a role in efficient RNA-directed DNA methylation. *Plant J* 75(5):836–846.
6. Mimura Y, Takahashi K, Kawata K, Akazawa T, Inoue N (2010) Two-step colocalization of MORC3 with PML nuclear bodies. *J Cell Sci* 123(Pt 12):2014–2024.
7. Moissiard G, et al. (2014) Transcriptional gene silencing by *Arabidopsis microrchidia* homologues involves the formation of heteromers. *Proc Natl Acad Sci USA* 111(20):7474–7479.
8. Dutta R, Inouye M (2000) GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem Sci* 25(1):24–28.
9. Corbett KD, Berger JM (2003) Structure of the topoisomerase VI-B subunit: Implications for type II topoisomerase mechanism and evolution. *EMBO J* 22(1):151–163.
10. Corbett KD, Berger JM (2005) Structural dissection of ATP turnover in the prototypical GHL ATPase TopoVI. *Structure* 13(6):873–882.
11. He F, et al. (2010) Structural insight into the zinc finger CW domain as a histone modification reader. *Structure* 18(9):1127–1139.
12. Li X, et al. (2012) Quantitative chemical proteomics approach to identify post-translational modification-mediated protein-protein interactions. *J Am Chem Soc* 134(4):1982–1985.
13. Eberl HC, Spruijt CG, Kelstrup CD, Vermeulen M, Mann M (2013) A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. *Mol Cell* 49(2):368–378.
14. Engelen E, et al. (2015) Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. *Nat Commun* 6:7155.
15. Hoppmann V, et al. (2011) The CW domain, a new histone recognition module in chromatin proteins. *EMBO J* 30(10):1939–1952.
16. Liu Y, et al. (2016) Family-wide characterization of histone binding abilities of human CW domain-containing proteins. *J Biol Chem* 291(17):9000–9013.
17. Ali MM, et al. (2006) Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* 440(7087):1013–1017.
18. Iyer LM, Abhiman S, Aravind L (2008) MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. *Biol Direct* 3:8.
19. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372(3):774–797.
20. Patel DJ, Wang Z (2013) Readout of epigenetic modifications. *Annu Rev Biochem* 82:81–118.
21. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
22. Calo E, Wysocka J (2013) Modification of enhancer chromatin: What, how, and why? *Mol Cell* 49(5):825–837.
23. Heintzman ND, et al. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39(3):311–318.
24. Nolivos S, Sherratt D (2014) The bacterial chromosome: Architecture and action of bacterial SMC and SMC-like complexes. *FEMS Microbiol Rev* 38(3):380–392.
25. Hirano T (2016) Condensin-based chromosome organization from bacteria to vertebrates. *Cell* 164(5):847–857.
26. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
27. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.
28. Laskowski RA, Swindells MB (2011) LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* 51(10):2778–2786.
29. Zhang Z, Marshall AG (1998) A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *J Am Soc Mass Spectrom* 9(3):225–233.
30. Lu J, et al. (2015) Improved peak detection and deconvolution of native electrospray mass spectra from large protein complexes. *J Am Soc Mass Spectrom* 26(12):2141–2151.
31. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
32. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
33. Shen L, Shao N, Liu X, Nestler E (2014) ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* 15(1):284.
34. Zhang Y, et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9):R137.
35. Yu G, Wang L-G, He Q-Y (2015) ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31(14):2382–2383.
36. Bernstein BE, et al. (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120(2):169–181.
37. Shen Y, et al. (2012) A map of the *cis*-regulatory sequences in the mouse genome. *Nature* 488(7409):116–120.
38. Heintzman ND, et al. (2009) Histone modifications at human enhancers reflect global cell type-specific gene expression. *Nature* 459(7243):108–112.
39. Anders S, Pyl PT, Huber W (2015) HTSeq: A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166–169.
40. Hung T, et al. (2009) ING4 mediates crosstalk between histone H3 K4 trimethylation and H3 acetylation to attenuate cellular transformation. *Mol Cell* 33(2):248–256.
41. Cong L, et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339(6121):819–823.

www.manaraa.com