UNIVERSITY OF CALIFORNIA

Los Angeles

Elucidating the mechanism of the *microrchidia* family of ATPases

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy in
Molecular, Cell and Developmental Biology

by

Linda Yen

2019

Elucidating the mechanism of the *microrchidia* family of ATPases

by

Linda Yen

Doctor of Philosophy in Molecular, Cell and Developmental Biology

University of California, Los Angeles,

2019

Professor Steve E. Jacobsen, Chair

The *microrchidia* (MORC) family of proteins are widely conserved eukaryotic and prokaryotic organisms. MORCs are GHKL (Gyrase, HSP90, Histidine Kinase, MutL) ATPases. Despite evidence that they are involved in gene silencing and genome compaction in multiple eukaryotic organisms, it is unknown what is their direct contribution to gene regulation *in vivo* or how they act on a molecular level. To elucidate how MORCs act to maintain gene silencing, I determined that MORCs are GHKL ATPases that form multimers and regulate a unique subset of genes. I show that MORCs with functional CW histone mark reader domains can be targeted to chromatin by recognition of histone tails, and validate MORC3 localization *in vivo.* From a structural analysis perspective, we also show that MORCs dimerize at the N terminus upon ATP

binding, and that they are DNA binding proteins which show length-dependent binding preferences. MORCs prefer to bind to longer DNA templates over short DNA templates, and they also have little sequence specificity with regards to their binding specificity. Using single molecule studies, we show that *C. elegans* MORC-1 compacts DNA by loop trapping it; MORC-1 also compacts nucleosome templates. These single molecule studies also clarify the role of ATP metabolism in MORC action. We show that ATP stimulates the rate of compaction on DNA, and is further stimulated by an ATP non-hydrolysable analog, demonstrating that nucleotide binding impacts MORC-1 stability on DNA. We also show that MORC-1 phase separates *in vitro* and during DNA compaction, MORC-1 forms large foci that are topologically entrapped on DNA. The foci mirror the 1,6 hexanediol resistant puncta observed *in vivo.* Their resistance to 1,6 hexanediol treatment is likely because MORC-1 puncta are topologically entrapped and use DNA as a scaffold.

MORCs thus depend on alternative mechanisms in order to bind to specific regions of chromatin for the same mechanistic process, which culminates in genome compaction. Altogether, my thesis reveals novel mechanistic insights into the MORC family. I propose a model whereby MORCs are targeted *in vivo* by either their own CW domains or other protein factors, trap loops of DNA at that locus to compact and silence the gene, and multimerize on themselves to create a phase separated environment that enforces genome compaction by inducing DNA compaction.

The dissertation of Linda Yen is approved.

Michael F. Carey
Amander T. Clark
Jeffrey A. Long
S. Lawrence Zipursky
Steve E. Jacobsen, Committee Chair

University of California, Los Angeles

2019

For my family …

Table of Contents

List of Figures

List of Tables

List of Abbreviations

C. elegans…… *Caenorhabditis elegans*
MORC……. *Microrchidia*
CTCF ……CCCTC-binding factor
SMC ……..Structural Maintenance of Chromosome
HP1 …… heterochromatic protein 1
PRC ……...polycomb repressive complex
TADs ……Topologically Associated Domains
H3K4me3 …….Histone 3 lysine 4 trimethylation
H3K9me3…….…….Histone 3 lysine 9 trimethylation
H3K27me3…….…….Histone 3 lysine 27 trimethylation
ChIPseq…….Chromatin immunoprecipitation and sequencing
DNA……...Deoxyribosenucleic acid
RNA……... Ribonucleic acid
PMSF……Phenylmethylsulfonyl fluoride
IPTG……...….Isopropyl β-D-1-thiogalactopyranoside
STORM …….…….Stochastic Optical Reconstruction Microscopy
PALM……Photoactivated Localization Microscopy
RNAi…… RNA interference
DIC ……..Differential Interference Contrast
LLPS……..... Liquid-liquid phase transition
HA……..Human influenza hemagglutinin
HUSH complex …….Human Silencing Hub
MPP8 ……..… M-phase phosphoprotein 8
GFP……...…Green fluorescent protein
HNs……... …histone-like nucleoid structuring protein
ALT…… alternative lengthening of telomere pathway
PiCH ……Proteomics of isolated chromosome segments
IAP elements …….Mouse Intra-cisternal A-type particle elements

Acknowledgements

I have been blessed to be surrounded by many amazing and inspiring individuals throughout my life. Graduate school was an especially difficult period where I also faced many changes and challenges in my life. It also led immense personal growth, none of which would have been possible for the people who supported me through these past eight years at UCLA.

First, I'd like to give thanks to my advisor, Steve, for his many years of guidance and support. He has been a wonderful role model in living without fear and has taught me to be a more direct person. I also thank him for trusting me and allowing me to take my project and run with it in any direction I saw fit. Without his enthusiasm and limitless amount of understanding, this degree and what I have accomplished would not have been possible.

I also am indebted to:

Mike Carey for adopting me as a lab member, and letting me have bench space in your lab for so many years. You have watched after me as if I were your own student and for that I am forever grateful.

Joe Loparo, whose objectivity and wisdom has guided me and given me clarity through very challenging and chaotic times.

John Gross and the Gross lab, who took a chance on an average undergraduate. It is not an exaggeration to say that without the guidance and mentorship and training (and patience!) that everyone showed me, I would not be a scientist. You all allowed me to make (very silly and really expensive) mistakes, but gave me room to grow and better myself from those experiences.

I am grateful for all of the collaborators I have had along the way, and in particular HyeongJun, Sisi and Jiamu. The discoveries reported here were the result of our teamwork.

To the Jacobsen lab (and auxiliary members) for being my teachers and advisors, scientific and life related discussions, a wonderful work environment, offering unlimited moral and parenting support, being my second set of hands, and being my confidants: Mahnaz and Ming C., Guillaume, Will, Pop, Dylan, Javier, Marian, Peggy, Wanlu, Ashot, Lianna, Jixian, Jason, Magda, Colette, Qikun, Jake, Yan, Sylvain, Ming W., Suhua, Basu, Michael, Zheng, Marco, Lucia, Zhenhui, Israel, Julie, Xuehua, Chris, Martin, Hume, Josh, Haifeng, and Truman.

To BSRB 3rd floor and the broader UCLA community: for helping me with my very frequent reagent requests and constant questions, emotional and scientific guidance, and assistance when I could not physically be in two places at once: Suman, Fei, Mike, Chenyang, Ben, Jason, Maria, Chen, Narsis, Oscar, Trent, Yong, Siavash, Reid, Wenyang, Sridhar, Sophie, Andrew, Will, Ajay, Val, Jason, Isabel, Sam, Bryan, Biviana and Tiffany.

Thank you to my committee (Amander, Larry, Mike, Jeff) for their invaluable advice, enthusiasm and believing in me, especially when I tended to overthink things. And a special thank you because along with the administrators that have handled my student affairs over the past eight years: Pamela Hurley, Ashley TerHost, Chris Briganti and Jon Lowenson. They have all had to put up with my last-minute chaos and always being the last to do everything. (Last person to qualify, last person of my incoming year to graduate… and the last ever MCDB doctoral student).

To my counselors, who helped me to trust my intuition again.

To my friends: You have stood by me during tough times, kept me company during long lab hours and brought me food when I was too busy to even eat. We've survived so many adventures (and misadventures) together. You all have been unconditionally supportive of me but completely unafraid to force me to have difficult conversations when I needed them the most.

Finally, a thank you to my grandparents, who sacrificed everything to give our family the opportunities that we are fortunate to have now. And last but not least, I am forever grateful for my parents, my aunts and uncles, my sister, my son, and all of our adopted fur babies.

Thank you to the University of California education system for an unbeatable training experience and a wonderful postsecondary education. Go Bears and Go Bruins. I have been supported by: a Ruth L. Kirschstein National Research Service Award GM007185 and a University of California, Los Angeles Dissertation Year Fellowship, and additional financial support from UCLA Molecular, Cell and Developmental Biology and the Howard Hughes Medical Institute.

Chapter 3 is a version of an article published in Proceedings of National Academy of Sciences: Moissiard, G., Bischof, S., Husmann, D., Pastor, W.A., Hale, C.J., Yen, L., Stroud, H., Papikian, A., Vashisht, A.A., Wohlschlegel, J.A., Jacobsen S.E. (2014). Transcriptional gene silencing by Arabidopsis microrchidia homologues involves the formation of heteromers. Proceedings of the National Academy of Sciences of the United States of America 111, 7474-7479. Co-authors Guillaume Moissiard, Sylvain Bischof Dylan Husmann, William Pastor, Chris Hale, Hume Stroud, Ashot Papikian designed and performed experiments, generated reagents and analyzed data. Ajay Vashisht performed the mass spectrometry experiments. I also thank Drs. Steve Jacobsen and James Wohlschlegel for supervision of the project.

Chapter 4 is a version of an article published in Proceedings of National Academy of Sciences: Li, S*., Yen, L*., Pastor W.A., Johnston J.B., Du, J., Shew, C., Liu, W., Ho, J., Stender, B., Clark, A.T., Burlingame, A.L., Daxinger, L., Patel, D.J., Jacobsen S.E. (2016) Mouse MORC3 is a GHKL ATPase that localizes to H3K4me3 marked chromatin. Proc. Natl. Acad. Sci. U. S. A. 113: E5108-16. *Denotes equal contribution. Sisi Li and Jiamu Du solved the

crystal structure; Will Pastor generated reagents and along with Wanlu Liu, performed the genome wide studies and data analysis; Colin Shew, Bryan Stender and Jamie Ho performed experiments. Jonathan Johnston performed the native mass spectrometry experiments. I also thank Drs. Steve Jacobsen, Dinshaw Patel, Lucia Daxinger, Al Burlingame, and Amander Clark, for supervision of the project.

Chapter 5 is a version of an article that has been submitted and is currently under revision at *Molecular Cell*: Kim, H.J.*, **Yen, L.***, Wongpalee, S., Kirschner, J.A., Mehta, N., Xue, Y., Johnston, J.B., Burlingame, A.L., Kim, J.K., Loparo, J.J., Jacobsen, S.E. The gene silencing protein MORC-1 topologically entraps DNA and forms multimeric assemblies to cause DNA compaction *Denotes equal contribution. *Molecular Cell* (Under revision). I'd like to thank HyeongJun Kim for the single molecule experiments and data analysis, Jessica Kirschner, Nicita Mehta and Yan Xue for performing the imaging, and Somsakul Wongpalee for generating reagents. Jonathan Johnston performed the native mass spectrometry experiments. I also thank Drs. Steve Jacobsen, Joe Loparo, John Kim, and Al Burlingame for supervision of the project. In addition, we thank Michael Carey, Reid Johnson, Geeta Narlikar, Sridhar Mandali, and Elizabeth Thrall for discussion and advice, Sean Carney for assistance with reagent preparation, and Alison Frand and Sophie Katz for guidance and assistance with microscopy. The supercoiled pBR322 derivative was a gift from Sridhar Mandali and Reid Johnson. The MBP vector was a gift from Andrew Ah Young. This work was supported by National Institutes of Health (NIH) Grant 1R35GM130272 and a grant from the W. M. Keck Foundation (S.E.J); NIH R01 GM114054 (J.J.L) and a John and Virginia Kaneb Fellowship (J.J.L); the University of Texas Rio Grande Valley Start-up Grant and the University of Texas System's Rising STARs Award (H.J.K.); NIH grant P41GM103481 and a University of California, San Francisco Research Resource Program

(RRP) Shared Equipment Award (A.L.B.) for the purchase of the Exactive EMR mass spectrometer; a Ruth L. Kirschstein National Research Service Award GM007185 and a University of California, Los Angeles Dissertation Year Fellowship (L.Y); a NIH R01 GM129301 (J.K.K). S.E.J is an Investigator of the Howard Hughes Medical Institute.

Vita/Biographical sketch

Education

University of California, Berkeley (January 2006-December 2009)
B.A. in Molecular and Cell Biology
Minor in Public Policy

Publications

1. Wongpalee, S.P., Liu, S., Gallego-Bartolomé, J., Leitner, A., Aebersold, R., Liu, W., **Yen, L.**, Nohales, M.A., Kuo, P.H., Vashisht, A.A., Wohlschlegel, J.A., Feng, S., Kay, S.A., Zhou, Z.H., Jacobsen, S.E. CryoEM Structures of Arabidopsis DDR Complexes Involved in RNA-directed DNA Methylation. *Nature Communications* (Accepted)

2. Kim, H.J.*, **Yen, L.***, Wongpalee, S., Kirschner, J.A., Mehta, N., Xue, Y., Johnston, J.B., Burlingame, A.L., Kim, J.K., Loparo, J.J., Jacobsen, S.E. The gene silencing protein MORC-1 topologically entraps DNA and forms multimeric assemblies to cause DNA compaction *Denotes equal contribution. *Molecular Cell* (Under revision)

3. Attar, N.A.*, Campos, O.A.*, Vogelaur, M.*, Xue, Y., Schmollinger, S., Mallipeddi, N.V., Cheng, C., **Yen, L.**, Yang, S., Zikovich, S., Dardine, J., Carey, M.F., Merchant, S.S., Kurdistani, S.K. The Histone H3-H4 tetramer is a copper reductase enzyme. *Denotes equal contribution (In preparation)

4. Harris, C.J., Scheibe, M., Wongpalee, S., Liu, W., Cornett, E., Vaughan, R., Li, X., Chen, W., Xue, Y., Zhong, Z., **Yen, L.**, Barshop, W.D., Rayatpisheh, S., Gallego-Bartolome, J., Groth, M., Wohlschlegel, J.A., Du, J., Rothbart, S., Butter, F., Jacobsen, S.E. (2018) A DNA methylation-directed transcriptional activator complex in *Arabidopsis*. *Science* 362(6419):1182-1186

5. Li, S*., **Yen, L***., Pastor W.A., Johnston J.B., Du, J., Shew, C., Liu, W., Ho, J., Stender, B., Clark, A.T., Burlingame, A.L., Daxinger, L., Patel, D.J., Jacobsen S.E. (2016) Mouse MORC3 is a GHKL ATPase that localizes to H3K4me3 marked chromatin. *Proc. Natl. Acad. Sci. U. S. A.* 113: E5108-16. *Denotes equal contribution

6. Pradhan, S.K., Su, T., **Yen, L.**, Jacquet, K., Huang, C., Côté, J. Kurdistani, S.K., Carey, M.F. (2016) EP400 deposits H3.3 into promoters and enhancers during gene activation. *Molecular Cell* 61(1):27-38

7. Moissiard, G., Bischof, S., Husmann, D., Pastor, W.A., Hale, C.J., **Yen L.**, Stroud, H., Papikian, A., Vashisht, A.A., Wohlschlegel, J.A., Jacobsen, S.E. (2014) Transcriptional gene silencing by Arabidopsis microrchidia homologues

involves the formation of heteromers. *Proc. Nat. Acad. Sci. U. S.A.* 111:7474-7479

8. Stanley, D.S., Bartholomeeusen, K., Crosby, D.C., Kim, D.Y., Kwon, E., Cartozo, N.C., **Yen, L.**, Li, M., Jaeger, S., Mason-Herr, J.M., Hayashi, F., Yokoyama, S., Krogan, N., Peterlin, B.M. and Gross, J.D. (2012) Inhibition of a NEDD8 cascade restores restriction of HIV-1 by APOBEC3G. *PLoS Pathogens* 8(12): e1003085

9. Jaeger, S., Kim, D.Y., Hultquist, J., Shindo, K., LaRue, R.S., Kwon, E., Li, M., Anderson, B., **Yen, L.**, Stanley, D.J., Mahon, C., Kane, J., Franks-Skiba, K., Cimermancic, P., Burlingame, A., Sali, A., Craik, C., Harris, R., Gross J.D., Krogan, N.J. (2011) "Vif Hijacks CBFβ to degrade APOBEC3G and promote HIV-1 infection. *Nature* 481(7381): 371-375

Poster presentations

1. **Gordon Research Conference: Chromatin Structure and Function.** (May 22-26, 2016) Les Diablerets, Switzerland
   Poster title: Investigating the molecular mechanisms of MORC ATPases

2. **UCLA Molecular Biology Institute Annual Retreat**. (April 22-24, 2016) Lake Arrowhead, CA
   Poster title: Investigating how the Microrchidia family of ATPases bind chromatin to enforce gene silencing

3. **European Molecular Biology Organization: Chromatin and Epigenetics.** (May 6-10, 2015) Heidelberg, Germany
   Poster title: *Arabidopsis thaliana* homologues regulate gene silencing.

4. **UCLA Molecular Biology Institute Annual Retreat.** (April 24-25, 2015) Los Angeles, CA
   Poster title: Investigating how the Microrchidia family of ATPases bind chromatin to enforce gene *silencing*

5. **Wellcome Trust: From Nucleosomes to Chromosomes.** (April 30-May 2, 2014) Hinxton, Cambridge, United Kingdom (Poster)
   Poster title: *Arabidopsis thaliana* homologues regulate gene silencing.

6. **UCLA Molecular, Cell and Developmental Biology Research Conference.** (November 30-December 2, 2012) Los Angeles, CA
   Poster title: Functional analysis of AtMORC ATPases to elucidate their role in gene silencing in *Arabidopsis thaliana*

Chapter 1: General mechanisms of gene regulation and gene silencing

1.1 Genome architecture and gene regulation are intertwined.

A genome comprises essential DNA sequences that contain the genetic information, or coding sequences, to make proteins (Annunziato, 2008; Ou et al., 2017). The full length of a stretched-out genome is up to 2 meters in some eukaryotic organisms, and yet somehow it is tightly packaged into a nucleus a few micrometers in diameter (Ou et al., 2017) . In any given interphase eukaryotic cell, DNA is packaged by histones and other histone like proteins to form chromatin fibers (Annunziato, 2008; Ou et al., 2017). Histones have tails largely composed of basic residues that can be chemically modified (Annunziato, 2008). Depending on the modification, this signature can promote a permissive environment for transcription (such as histone 3 lysine 4 tri-methylation), or a repressive environment that prevents transcription (such as histone 3 lysine 9 tri-methylation). These markers can also recruit other protein complexes to enforce either chromatin state.

Organisms also depend on tightly coordinated mechanisms to access the information that is packaged in the genome. Temporally precise gene expression (and silencing other genes at the appropriate time) is critical for regulating cellular processes, and must distinguish which regions need to be accessible and which need to be silenced in any given cell. There are two general states of chromatin – euchromatin, characterized by loosely packed nucleosomes, accessible and transcriptionally active regions, and heterochromatin, composed of tightly packed, inaccessible and silenced genes.

There are multiple mechanisms of gene regulation, and each of these can occur at various levels. At the primary level, gene expression itself is under the control of other DNA elements such as promoters and enhancers. Promoter sequences are regions of DNA which recruit RNA polymerase II and other protein factors that comprise of the pre-initiation complex that can

regulate the rate of transcription. Transcriptionally inactive promotors are bound by silencing factors that occlude or hinder transcriptional machinery accessibility to the promoter. Enhancers are DNA sequences that are exist further away in the genome from their target promoter, but are often brought in proximity by mediating proteins (Boija et al., 2018; Kagey et al., 2010; Raab and Kamakaka, 2010; Sabari et al., 2018; Sun et al., 2019). This enhancer-promoter looping enables gene expression by enhancing the local concentration of requisite proteins (Boija et al., 2018; Sabari et al., 2018).

Little of the genome exists as free DNA within the nucleus. Approximately 147 base pairs of DNA are packaged by histone protein into the thousands of nucleosomes constituting chromatin (Ou et al., 2017). Chromatin itself exists in various states, and the modifications that are present on the histone (and DNA, i.e., methylation) recruit the protein activators or repressors that dictate whether chromatin possesses an euchromatic signature or a heterochromatic signature (Kouzarides, 2007). In sum, these combinations, depending on the modification and residue, can dictate whether a gene is expressed or silenced (Kouzarides, 2007).

DNA itself can be covalently modified with methylation on a cytosine. Methylation, depending on the context and whether it is in the promoter or gene body, is typically used to signal silencing (Law and Jacobsen, 2010). DNA methylation is also tightly coupled with the recruitment of silencing factors that serve to create a strong silencing environment.

1.2 Genome architecture can be manipulated by protein factors

To access specific parts of the genome at opportune times, many chromatin associated proteins contain histone binding domains that allow them to recognize certain chromatin states that are characterized by various histone modifications. This acts as a way to target protein action

to a specific region in the genome. These proteins can affect the local environment by inducing changes such as local chromatin compaction or remodeling the chromatin. All of these therefore impact accessibility to the underlying DNA (Kagey et al., 2010).

Another layer of epigenetic regulation is through chromosome folding and higher order structure. Recent work has demonstrated that the genome is a mixture of fibers of varying diameters and thicknesses (Ou et al., 2017). These differences reflect the impact of locally bound proteins because the chromatin landscape is determined by proteins that regulate the local packing. These proteins include CTCF and readers of specific histone modifications such as HP1 (which deposits H3K9me3) and the polycomb repressive complex (PRC; which deposits H3K27me3) family (Larson et al., 2017; Ong and Corces, 2014; Strom et al., 2017; Sun et al., 2019).

In some organisms, long range chromatin interactions known as topological associating domains (TADs) organize the chromatin further. These domains are bounded by CTCF and other architectural proteins, including cohesion. Cohesin is a member of the SMC family of proteins, SMC family members serve to create structure out of an otherwise orderless, very long strand of chromatin during the cell cycle and have a well characterized role in regulating chromatin architecture *in vivo (*Hirano, 2016). SMC mediated chromatin organization is crucial for multiple processes throughout mitosis and meiosis (Kagey et al., 2010; Murayama and Uhlmann, 2013). Failure to maintain proper chromosome packaging during these processes would lead to pre-mature termination and exit from recombination and cell division (Lee et al., 2013).

1.3 Mechanisms of chromosome compaction and genome folding

Chromosome organizing proteins fall into two categories: one brings distal pieces of chromatin/DNA together or the other compacts nearby chromatin/DNA. Distal examples include bacterial ParB (Spo0J), which compacts via an ATP-independent loop trapping method, or condensin which compacts via ATP-dependent loop extrusion. Although all these proteins are involved in chromatin structuring, they have vastly distinct mechanisms and measurable behaviors in a flow stretched DNA experiment.

The bacterial ParB binds to ParS sites located throughout the bacterial genome and oligomerizes, thereby bridging distal DNA filaments (Graham et al., 2014). Chromatin loops are caused by ATP-independent protein oligomerization (Graham et al., 2014). Condensin, on the other hand, translocates passively, then pauses at specific locations where it uses ATP dependent translocation to extrude a loop (Ganji et al., 2018).

Local compaction is exemplified by bacterial HNs (histone-like nucleoid structuring protein). HNs also oligomerizes, but as it binds DNA it induces a bend in the strand and therefore compacts by modifying the local segment (Graham et al., 2014). HNs finds its nearest neighbor HNs and induces compaction by all binding and bending in the same region (Graham et al., 2014).

Another mechanism of protein-mediated manipulation of genome architecture is through intracellular condensates (Boija et al., 2018; Sabari et al., 2018; Shin and Brangwynne, 2017; Shin et al., 2018). The nucleus is an organelle densely filled with proteins at concentrations estimated to be from 100-400 mg/ml (Larson and Narlikar, 2018). Molecular crowding and unique protein domains cause protein oligomerization that forms "condensate" structures in the nucleus, representing a membraneless sub-nuclear domain that clusters specific factors together (Larson and Narlikar, 2018). One mechanism that has regained attention recently is the idea that

such structure are created by phase separation and a critical questions is how this phenomenon impacts genome compartmentalization (Boija et al., 2018; Larson et al., 2017; Sabari et al., 2018; Strom et al., 2017). Proteins form these condensate structures frequently undergo liquid-liquid phase separation (LLPS) *in vitro,* as assayed by differential inference contrast (DIC) and confocal microscopy. Phase separated proteins *in vitro* undergo fusion and fission, and when photobleached recover fluorescence because the proteins are constantly mixing within and do not form a static aggregate. The ability to "phase separate" is typically mediated by intrinsically disordered regions (IDR). IDRs can be of different amino acid compositions, such as hydrophobic, hydrophobic and uncharged, or even polar and charged. These IDRs self-interact to create an environment that typically organizes water (Lin et al., 2017).

In vivo, phase transitioned bodies form intracellular condensates which are considered functionally membraneless compartments. Such compartments have properties that are either more fluid or gel like, or more solid in feature. Amyloid fibers, RNA binding proteins and stress granules are examples of complex protein-nucleic acid interactions that are known to phase transition, but their condensates are more fibrous in nature. These bodies are also insensitive to 1,6 hexanediol treatment (a chemical known to disrupt the punctate localization of intracellular condensates by interrupting weak, hydrophobic interactions mediated by some IDRs) (Strom et al., 2018). It is possible that the nature of the phase transitioned body *in vivo* is sensitive to additional interactions that scaffold the body. For example, Heterochromatic protein 1 (HP1) bodies are heavily scaffolded by chromatin, whereas RNA binding protein and stress granules include RNAs (Larsen and Narlikar, 2018).

There are two proposed mechanisms for how phase separation shapes genome organization. Optogenetic studies have demonstrated that droplets can form in euchromatic

environments. Proteins such as BRD4 and complexes such as Mediator have been shown to form puncta (intracellular condensates) in cells, and these proteins associate with various DNA sequences to pull them together into a local environment that creates a membraneless compartment. This sequestration increases the local concentration of the protein, creating multivalent contacts to further anchor these contacts. This clustering is also important for the fidelity of gene activation by enriching enhancers and cognate promoters in close proximity.

Alternatively, optogenetic studies have demonstrated that heterochromatic environments may have different behaviors. Activators that normally phase transition *in vivo,* when targeted to heterochromatic environments, are excluded from dense heterochromatin (Shin et al., 2018). This suggests that the density of these opposing environments may play a larger role in whether there is further mixing (Shin et al., 2018).

Dense heterochromatin is often marked by the presence of histone 3 lysine 9 trimethylation (H3K9me3). HP1 uses its chromodomains to bind to H3K9me3, and forms heterodimers that bridge nucleosomes to bring them together. In this way, HP1 creates a compact environment in regions where genes are meant to be silenced (Larson et al., 2017; Strom et al., 2017). HP1 has been observed to form puncta *in vivo*, though they are partially resistant to 1,6 hexanediol treatment and phase transitions *in vitro* as assayed by DIC microscopy upon phosphorylation of an unstructured region (Larson et al., 2017; Strom et al., 2017). Phosphorylation reveals a DNA binding domain, which then allows HP1 to further engage chromatin by binding to both H3K9me3 and linker DNA (Larson et al., 2017). This ability creates multivalent interactions between nucleosomes and HP1 that generates an environment inaccessible to other proteins (Larson et al., 2017; Shin et al., 2018; Strom et al., 2017).

Annunziato, A.T. (2008). DNA Packaging: Nucleosomes and Chromatin. Nature Education *1*.

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M.*, et al.,* (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842-1855.e1816.

Ganji, M., Shaltiel, I.A., Bisht, S., Kim, E., Kalichava, A., Haering, C.H., and Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. Science *360*, 102-105.

Graham, T.G., Wang, X., Song, D., Etson, C.M., van Oijen, A.M., Rudner, D.Z., and Loparo, J.J. (2014). ParB spreading requires DNA bridging. Genes & development *28*, 1228-1238.
Hirano, T. (2016). Condensin-Based Chromosome Organization from Bacteria to Vertebrates. Cell *164*, 847-857.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.,* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Kouzarides, T. (2007). Chromatin Modifications and Their Function. Cell *128*, 693-705.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. Nature *547*, 236-240.

Larson, A.G., and Narlikar, G.J. (2018). The Role of Phase Separation in Heterochromatin Formation, Function, and Regulation. Biochemistry *57*, 2540-2548.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics *11*, 204-220.

Lee, J. (2013). Roles of Cohesin and Condensin in Chromosome Dynamics During Mammalian Meiosis. Journal of Reproduction and Development *59*, 431-436.

Lin, Y., Currie, S.L., and Rosen, M.K. (2017). Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. Journal of Biological Chemistry *292*, 19110-19120.

Murayama, Y., and Uhlmann, F. (2013). Biochemical reconstitution of topological DNA binding by the cohesin ring. Nature *505*, 367-371.

Ong, C.-T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nature Reviews Genetics *15*, 234-246.

Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H., and O'Shea, C.C. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science *357*.

Raab, J.R., and Kamakaka, R.T. (2010). Insulators and promoters: closer than we think. Nature Reviews Genetics *11*, 439-446.
Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C*., et al.,* (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science *361*.

Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. Science *357*.

Shin, Y., Chang, Y.-C., Lee, D.S.W., Berry, J., Sanders, D.W., Ronceray, P., Wingreen, N.S., Haataja, M., and Brangwynne, C.P. (2018). Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. Cell *175*, 1481-1491.e1413.

Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. Nature *547*, 241-245.

Sun, F., Chronis, C., Kronenberg, M., Chen, X.-F., Su, T., Lay, F.D., Plath, K., Kurdistani, S.K., and Carey, M.F. (2019). Promoter-Enhancer Communication Occurs Primarily within Insulated Neighborhoods. Molecular Cell *73*, 250-263.e255.

Chapter 2: A primer on the *microrchidia* family of ATPases

2.1 MORCs are evolved from prokaryotic origins and are found in several eukaryotic systems

The *microrchidia* (MORC) family of ATPases was first discovered by the phenotype generated when a transgene interrupted the *MORC1* gene in mice (Watson et al., 1998). Homozygous mutant mice display the phenotype of small testis. This phenotype was later identified to be the result of a failure of sperm to undergo proper meiosis due to a failure of homologous chromosome pairing during pachytene (Watson et al., 1998). MORCs are found in multiple prokaryotic, archaeal and eukaryotic systems, and in eukaryotes are nuclear localized (Dong et al., 2018; Harris et al., 2016; Iyer et al., 2008; Mimura et al., 2010; Moissiard et al., 2012; Pastor et al., 2014). MORC mutations result in a variety of phenotypes, depending on the expression pattern. Some MORCs have tissue-specific expression, whereas others are expressed in multiple tissue types (Li et al., 2016 [See Chapter 4]; Pastor et al., 2014). There is evidence of redundancy of some MORCs in various organisms as well (Moissiard et al., 2014 [See Chapter 3]).

There are seven *Arabidopsis thaliana* MORCs (AtMORC1, AtMORC2, AtMORC3, AtMORC4, AtMORC5, AtMORC6, AtMORC7) among which AtMORC3 creates a protein product that prematurely terminates after the ATPase domain, and likely is not considered functional. Although there is redundant expression of some MORCs in multiple tissues, others display tissue-specific expression. For example, AtMORC5 expression is pollen specific. This observation argues that although there is some redundancy for these family members, they have also evolved to perform specific roles in certain contexts. For example, MORC1 has been shown to be involved in pathogen response (Kang et al., 2012), but this section will exclusively address the MORC family's role in gene regulation.

All AtMORCs have the GHL ATPase domain (Moissiard et al., 2012) followed by a flexible linker region and a coiled coil domain, but are missing CW histone mark reader domains (Figure 2-1).

There are five MORCs in mice (MORC1, MORC2a, MORC2b, MORC3 [also known as NXP2], and MORC4). MORC3 and MORC4 have functional CW domains with a triad of tryptophans that are used to specifically recognize methylated lysines on histone tails (Eberl et al., 2013; Engelen et al., 2015; Hoppmann et al., 2011; Li et al., 2012; Liu et al., 2016). In comparison, MORC1, MORC2a and MORC2b contain CW domains but are lacking the conserved tryptophans necessary to bind methylated lysines. MORC2a and MORC4's specific roles and molecular mechanisms have yet to be established. MORC2b is an intronless version of MORC2a, and is likely to be redundant as they are expressed in similar/same tissue types.

Similarly, there are four MORCs in humans (MORC1, MORC2, MORC3, MORC4). hMORC2 is the best characterized and mutations in this protein have been linked to Charcot-Tooth-Marie disease (CMT) in addition to other neurological disorders (Tchasovnikarova et al., 2017, Douse et al., 2018). MORC3 has also been linked to influenza infectivity.

In *C. elegans*, there is only one MORC (MORC-1). However, mutations in this gene lead to a progressive loss in fertility over generations.

This list is not exhaustive; MORCs can be found in multiple other organisms, such as mosses and zebrafish. However, the role of MORCs in these systems have not yet been determined.

2.2 MORCs are GHL ATPases

MORCs are GHL (Gyrase, HSP90, MutL) ATPases. All members of this superfamily contain a Bergerat fold ATPase domain, followed by the S5 domain containing conserved

catalytic residues required for ATP hydrolysis (stabilizing the ATP molecule during the transition state) (Corbett and Berger, 2005).

The MORC family, however, specifically also share a common domain architecture, where a GHL ATPase fold is followed by a flexible linker region and ends with a coiled coil domain (Dong et al., 2018; Inoue et al., 1999; Iyer et al., 2008; Li et al., 2016 [See Chapter 4]; Mimura et al., 2010). MORCs in multiple organisms form heteromers or homomers, and these multimers are presumably dictated by interactions among the coiled-coil domains (Mimura et al., 2010). MORC multimers have been observed as nuclear bodies *in vivo* (Harris et al., 2016; Mimura et al., 2010; Moissiard et al., 2012). Collectively, these observations suggest that multimerization is an important aspect of MORC function.

GHLs are GHKL (Gyrase, histidine kinase, MutL, HSP90) ATPases that do not contain the S5 domain (Corbett and Berger, 2005). The GHKL ATPases ubiquitiously engage ATP using their Bergerat fold at a 1:1 stoichiometry. ATP binding induces conformational changes that allow the ATPase domains to dimerize (Corbett and Berger, 2005). The dimerization interface in the ATPase domain is formed by a cage of aromatic residues (identical ones on either side) in addition to hydrophobic residues that interlock to create a solvent inaccessible core. This conservation is so strict that even alignments of MORCs ATPase domains in conjunction with other prokaryotic GHKLs show conservative substitutions in these regions (Figure 2-2).

The residues involved in ATP binding and hydrolysis are equally conserved. At the catalytic cleft of the ATPase domain, there is an Asparagine (N) required for magnesium binding (Corbett and Berger, 2005; Iyer et al., 2008); a catalytic glutamic acid coordinated by conserved basic amino acid; and a lysine required for stabilizing the transition state during ATP hydrolysis. For ATP binding, there is an aspartic acid, followed by a threonine and a flexible loop.

Mutations in any of these residues abolish ATP binding and hydrolysis (Figure 2-2) (Corbett and Berger, 2005; Iyer et al., 2008). Interestingly, for the MORC ATPases, it has been reported that ATP binding mutants are not stably expressed, suggesting that the ATP binding fold is important for protein stability (Douse et al., 2018). Similarly, during ATP wash out experiments, a mutation that theoretically abolished ATP hydrolysis exhibited defects that appeared to be similar to ATP binding mutants (Mimura et al., 2010). This observation suggests that at least for the MORC family, residues involved in ATP hydrolysis also play a role in stabilizing the binding pocket.

2.3 MORC ATPases regulate gene expression

MORCs were discovered to regulate gene silencing and genome condensation in plants as the result of a forward genetic screen in a sensitized background where DNA methylation was compromised. In this background, the SDC promoter was fused to green fluorescent protein (GFP) (Moissiard et al., 2012). SDC is a promoter that is normally methylated, and therefore expression is silenced in wild type plants. *morc6 or morc1* mutants exhibited transgene reactivation. *MORC6* and *MORC1* were later identified to regulate the same sets of transposable elements, and both did so in a manner independent of DNA methylation or H3K9me3, a histone modification associated with gene silencing in plants (Moissiard et al., 2012). Transposon reactivation was approximately two fold greater in *morc6* mutants compared to *morc1* and its close homologue *morc2* (Moissiard et al., 2014 [See Chapter 3]).

morc1 null mice also show an upregulation of expression of transposon reactivation. In these mice, intra-cisternal A-type particle (IAP) elements are increased in expression at postnatal day 2 and fail to establish methylation at these sites (Pastor et al., 2014). There is a concomitant opening of chromatin as assessed by the presence of H3K4me3 (which correlates with

transcriptionally active areas; this alone is sufficient to repel enzymes that establish DNA methylation such as DNMT3A) and ATAC-seq (Pastor et al., 2014). Analysis of LINE elements demonstrate that methylation is most impacted at transcriptional start site of transposons, suggesting that MORC1 exerts its effects locally (Pastor et al., 2014).

Mouse MORC3 was shown to be necessary for silencing a reporter transgene bearing a long terminal repeat in mouse erythrocytes. Through a technique called Proteomics of isolated chromatin segments PICh , mouse MORC3 was shown to localize to telomeres and be involved in maintaining telomeres through the alternative lengthening of telomere (ALT) pathway (Saksouk et al., 2014). Overexpression of MORC3 induces cellular senescence through activating p53. MORC3 has been shown to localize to PML bodies, but is capable of forming significant PML-independent bodies in the nucleus in multiple cell lines (Takahashi et al., 2007). MORC3 likely has a critical role in development, as MORC3 -/- mice die either *in utero* or shortly after birth (Takahashi et al., 2007). The role of mouse MORC3 in silencing has yet to be established, although chromatin immunoprecipitation experiments (ChIP) in embryonic stem cells show that MORC3 localizes to H3K4me3, and that the levels of MORC3 correlate directly with the abundance of H3K4me3. However, *morc3* embryonic stem cells (ESCs) display few changes in gene expression (Li et al., 2016 [See Chapter 4]).

Human and mouse MORC3 have a high degree of sequence conservation, and published reports in human cell lines suggest that human *morc3* cell lines take longer to exit mitosis (Zhang et al., 2019). Although there are many possibilities for the underlying mechanism, one possible explanation is that there is a failure to re-establish heterochromatin in *morc3* cells, suggesting that MORC3 may play a role in maintaining heterochromatin during the cell cycle. Human

MORC3 has also been demonstrated to be modified with sumo at multiple lysine residues, and mutations of these lysines abrogate the ability of human MORC3 to silence a Gal4-VP16-responsive luciferase reporter (Rosendorff et al., 2006). It is possible that the localization pattern in ESCs represent inactive MORC isoforms, and developmental signaling depends on sumoylation to establish specific silencing in particular regions of the genome (Figure 2-3).

In C. elegans, gene silencing in the germline is predominantly regulated by small RNAs. Small RNAs with sequences complementary to the gene of interest are processed and used as guides for RNA interference (RNAi). *morc1* mutants fail to silence a seam cell transgene reporter despite and intact RNAi pathway (Moissiard et al., 2012).

Human MORC2 was first identified from a CRISPR-Cas9 based silencing screen where a GFP reporter was previously silenced but reactivated upon the loss of MORC2. MORC2 was then shown to recruit a silencing complex known as the HUSH (human silencing hub) complex; immunoprecipitation experiments show that it specifically precipitates with TASOR, MPP8 and peripherin (Tchasovnikarova et al., 2017). The HUSH complex then recruits SETDB1, an enzyme known to deposit H3K9me3 (Tchasovnikarova et al., 2017). Through the use of genomics, it was shown that MORC2 recruits HUSH, and HUSH localization is dependent on MORC2 at the site (Tchasovnikarova et al., 2017). In this manner, MORC2 acts to recruit silencing machinery that establishes a silencing state in a specific region of the genome.

2.4 MORC forms higher order complexes

Multiple studies using immunoprecipitation and mass spectrometry demonstrate that MORCs are often found *in vivo* as multimers. AtMORC6 co-precipitates AtMORC1 and AtMORC2, but individually AtMORC1 or AtMORC2 could only precipitate AtMORC6 without

AtMORC1 (Moissiard et al., 2014 [See Chapter 3]). Peptide counts in mass spectrometry experiments suggested that in each of these experiments, either AtMORC6 and AtMORC1 or AtMORC6 and AtMORC2 were present in approximately the same stoichiometry (Moissiard et al., 2014 [See Chapter 3]). These results demonstrate that AtMORC6 could form mutually exclusive complexes with either AtMORC1 or AtMORC2, and these heteromers were responsible for transposon silencing.

Additionally, AtMORC4 and AtMORC7 were shown to regulate the expression of protein coding genes involved in stress response (Harris et al., 2016). However, these proteins were shown to act as homomers by immunoprecipitation/mass spectrometry, and seem to act redundantly on their target genes (Harris et al., 2016). Like other MORCs, AtMORC4 and AtMORC7 form nuclear bodies *in vivo* but appear to be localized near the nuclear periphery instead of at the heterochromatic chromosome centers called chromocenters (Harris et al., 2016; Moissiard et al., 2012). Because stress response genes are frequently found in proximity to the nuclear pore, the localization is expected within the context of genes that MORC4 and MORC7 regulate.

Human MORC3 experiments demonstrated that these proteins, like those in Arabidopsis and mouse, act as multimers (Takahashi et al., 2007). Human MORC3 was either HA or Flag tagged, then immunoprecipitated using the epitope of choice as target. When using Flag or the HA antibody to precipitate the fusion protein, western blot detected the presence of the other epitope. Detection of the other epitope was abrogated by deleting the coiled coil from the full-length protein, suggesting that the coiled coil domain was important for multimerization (Takahashi et al., 2007). Mutating the ATP hydrolysis site was not sufficient to abrogate this interaction (Takahashi et al., 2007), suggesting that any failure to induce ATP dependent

dimerization at the N terminus was decoupled from protein multimerization.

2.5 MORCs regulate gene silencing and genome condensation

In vivo, MORC6 was found to be localized to the periphery of the heterochromatic chromocenters (Moissiard et al., 2012), where transposons are predominantly localized in the genome of *Arabidopsis thaliana*. Most notably, *morc6* and *morc1* mutants show significant chromocenter decondensation measured by immunofluorescence and Hi-C, a chromosome conformation capture technique (Moissiard et al., 2012). This observation suggests that MORCs play a role in genome condensation, and activity is essential for transposable element silencing. This is because the other silencing markers such as DNA methylation or H3K9me2 (the silencing histone mark of deep heterochromatin) are largely unchanged at these sites where expression is upregulated in MORC mutants (Moissiard et al., 2012). Though these markers recruit proteins involved in establishing the requisite silencing pathways (such as CMT2 and CMT3), this is not sufficient to prevent transposable element expression.

Unlike other organisms, the worm *C. elegans* shows decondensation of the X chromosome (Weiser et al., 2017). In worms, heterochromatin is marked by H3K9me3 and is localized to the nuclear envelope. Additionally, in wild type worms transformed with a heterochromatic reporter array, the reporter is heavily coated with H3K9me3 and localized to the nuclear periphery. In *morc-1*, these arrays are decondensed and dissociated from the nuclear periphery and invade the nuclear space (Weiser et al., 2017). RNA-seq reveals that multiple genes become upregulated in *morc-1*, and these genes also exhibit a concomitant loss of H3K9me3, suggesting that these regions are no longer heterochromatic (Weiser et al., 2017). Additionally, a conditional mutant screen found that losing MET-1, a protein that deposits

H3K36me3 (an active mark that denotes regions that were recently transcribed), suppressed the germline sterility defect in *morc-1* mutants (Weiser et al., 2017). This finding suggests that there is an interplay between MORC-1 (silencing) and transcription. Together, these two processes regulate genome condensation.

2.6 MORC structure and function

MORCs have also been linked to gene silencing, with several of them being identified from forward genetic screens where a transgene reporter is reactivated upon loss of the MORC protein. Therefore, it stands to reason that MORCs are novel players in gene silencing. This function is unexpected based solely on the domain architecture (and the functions of its homologs and paralogues). However, given MORC's tight conservation of the ATPase domain, structural hypotheses that can be posed to explain its function. GHKL ATPases ubiquitously dimerize at the N terminus upon ATP binding, suggesting that there is a potential gating mechanism. Additionally, all MORCs have C terminal coiled coil domains. By homology, this domain is known to mediate multimerization (Mimura et al., 2010). Therefore, the model is that MORCs are constitutively complexed through the C terminus, and the N terminus opens and closes upon ATP binding (Figure 2-4).

Recent structural work by multiple groups have shown that both the mouse MORC3 and human MORC2 ATPase domain dimerize upon ATP binding (Li et al., 2016 [See Chapter 4], Douse et al., 2018). Interestingly, mutations that cause constitutive dimerization of the N terminal ATPase domain induce hyper-silencing (increase the kinetics of silencing) of a reporter transgene compared to wild-type, suggesting that the full ATP hydrolysis cycle is important for cycling MORC into an off state.

Some MORCs contain also contain CW zinc finger domains between the N and C terminal domains. The CW domain is structurally similar to a PHD histone mark reader domain, except it utilizes cysteines and tryptophans to coordinate the zinc ion required for proper folding. Although MORC3 and MORC4, like their mouse homologs, retain conserved tryptophans that allow them to recognize methylated lysines, MORC1 and MORC2 lack these residues and likely contain a CW domain that cannot bind to specific regions of the genome. MORC-1 in *C. elegans* also has CW domain without the conserved tryptophans required for methylated lysine histone tail recognition. Work using ectopically expressed CW domains and histone peptides demonstrate that without the tryptophans, the CW domain cannot recognize the H3K4me3 peptide (Hoppmann et al., 2011, Liu et al., 2016).

What then might be the role of the CW domain? In studies with mouse MORC3, it was demonstrated that the ATPase domain alone failed to bind DNA, but that a ATPase + CW domain construct could bind DNA (Andrews et al., 2016, Zhang et al., 2019). It was proposed that the CW domain bound to a histone peptide triggered a conformational change in the linker region between the ATPase and the CW domain. This change was important because the CW domain packing against the ATPase domain then triggered further conformational changes that allowed a construct bearing just the MORC3 ATPase+ CW domain to robustly bind DNA.

Figure 2-1: Domain schematic of the MORC ATPases

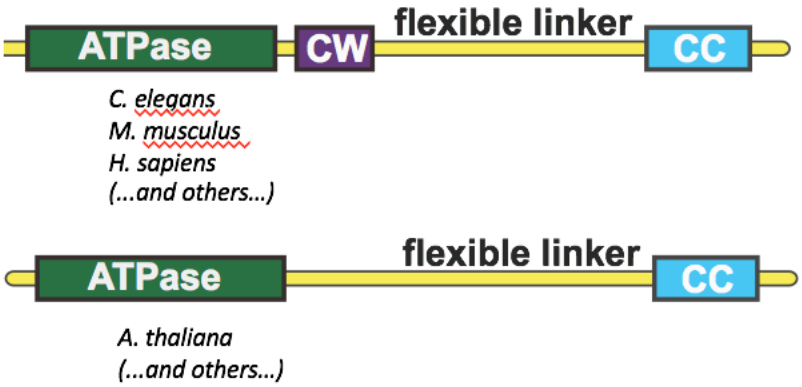
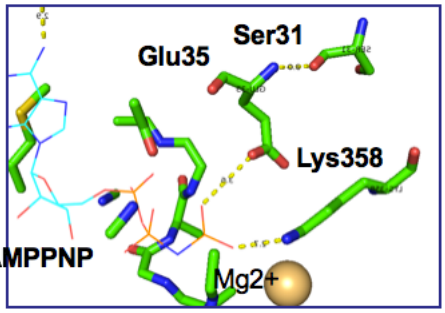


Figure 2-2: ATP Binding pocket of mouse MORC3



ATP hydrolysis
S31 (positions E for catalysis)
E35 (catalysis)
K358 (transition state stabilizer)

Figure 2-3: Hypothetical model of MORC3 post translational modification dependent targeting

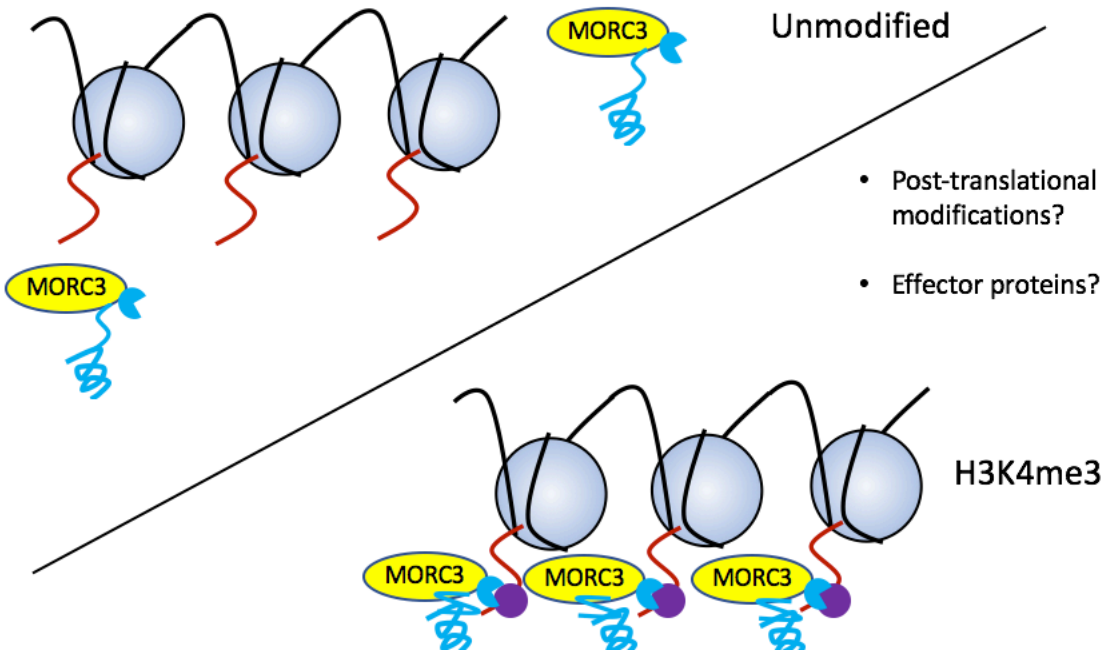
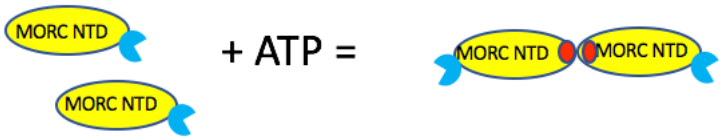


Figure 2-4: MORC ATPase domains dimerize upon ATP binding



2.7 References

Andrews, F.H., Tong, Q., Sullivan, K.D., Cornett, E.M., Zhang, Y., Ali, M., Ahn, J., Pandey, A., Guo, A.H., Strahl, B.D., et al., (2016). Multivalent Chromatin Engagement and Inter-domain Crosstalk Regulate MORC3 ATPase. Cell Reports 16, 3195-3207.

Corbett, K.D., and Berger, J.M. (2005). Structural Dissection of ATP Turnover in the Prototypical GHL ATPase TopoVI. Structure 13, 873-882.

Dong, W., Vannozzi, A., Chen, F., Hu, Y., Chen, Z., Zhang, L., and Pritham, E. (2018). MORC Domain Definition and Evolutionary Analysis of the MORC Gene Family in Green Plants. Genome Biology and Evolution 10, 1730-1744.

Douse, C.H., Bloor, S., Liu, Y., Shamin, M., Tchasovnikarova, I.A., Timms, R.T., Lehner, P.J., and Modis, Y. (2018). Neuropathic MORC2 mutations perturb GHKL ATPase dimerization dynamics and epigenetic silencing by multiple structural mechanisms. Nature communications 9.

Eberl, H.C., Spruijt, Cornelia G., Kelstrup, Christian D., Vermeulen, M., and Mann, M. (2013). A Map of General and Specialized Chromatin Readers in Mouse Tissues Generated by Label-free Interaction Proteomics. Molecular Cell 49, 368-378.

Engelen, E., Brandsma, J.H., Moen, M.J., Signorile, L., Dekkers, D.H.W., Demmers, J., Kockx, C.E.M., Ozgür, Z., van Ijcken, W.F.J., van den Berg, D.L.C., et al., (2015). Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. Nature communications 6.

Harris, C.J., Husmann, D., Liu, W., Kasmi, F.E., Wang, H., Papikian, A., Pastor, W.A., Moissiard, G., Vashisht, A.A., Dangl, J.L., et al., (2016). Arabidopsis AtMORC4 and AtMORC7 Form Nuclear Bodies and Repress a Large Number of Protein-Coding Genes. PLOS Genetics 12, e1005998.

Hoppmann, V., Thorstensen, T., Kristiansen, P.E., Veiseth, S.V., Rahman, M.A., Finne, K., Aalen, R.B., and Aasland, R. (2011). The CW domain, a new histone recognition module in chromatin proteins. The EMBO Journal 30, 1939-1952.

Inoue, N., Hess, K.D., Moreadith, R.W., Richardson, L.L., Handel, M.A., Watson, M.L., and Zinn, A.R. (1999). New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. Human Molecular Genetics 8, 1201-1207.

Iyer, L.M., Abhiman, S., and Aravind, L. (2008). MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. Biology direct 3, 8.

Kang, H.-G., Woo Choi, H., von Einem, S., Manosalva, P., Ehlers, K., Liu, P.-P., Buxa, S.V., Moreau, M., Mang, H.-G., Kachroo, P., et al., (2012). CRT1 is a nuclear-translocated MORC endonuclease that participates in multiple levels of plant immunity. Nature communications 3, 1297.

Li, S., Yen, L., Pastor, W.A., Johnston, J.B., Du, J., Shew, C.J., Liu, W., Ho, J., Stender, B., Clark, A.T., et al., (2016). Mouse MORC3 is a GHKL ATPase that localizes to H3K4me3 marked chromatin. Proceedings of the National Academy of Sciences of the United States of America 113, E5108-5116.

Li, X., Foley, E.A., Molloy, K.R., Li, Y., Chait, B.T., and Kapoor, T.M. (2012). Quantitative Chemical Proteomics Approach To Identify Post-translational Modification-Mediated Protein–Protein Interactions. Journal of the American Chemical Society 134, 1982-1985.

Liu, Y., Tempel, W., Zhang, Q., Liang, X., Loppnau, P., Qin, S., and Min, J. (2016). Family-wide Characterization of Histone Binding Abilities of Human CW Domain-containing Proteins. J Biol Chem 291, 9000-9013.

Mimura, Y., Takahashi, K., Kawata, K., Akazawa, T., and Inoue, N. (2010). Two-step colocalization of MORC3 with PML nuclear bodies. Journal of cell science 123, 2014-2024.

Moissiard, G., Bischof, S., Husmann, D., Pastor, W.A., Hale, C.J., Yen, L., Stroud, H., Papikian, A., Vashisht, A.A., Wohlschlegel, J.A., et al., (2014). Transcriptional gene silencing by Arabidopsis microrchidia homologues involves the formation of heteromers. Proceedings of the National Academy of Sciences of the United States of America 111, 7474-7479.

Moissiard, G., Cokus, S.J., Cary, J., Feng, S., Billi, A.C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B.R., McCord, R.P., et al., (2012). MORC family ATPases required for heterochromatin condensation and gene silencing. Science 336, 1448-1451.

Rosendorff, A., Sakakibara, S., Lu, S., Kieff, E., Xuan, Y., DiBacco, A., Shi, Y., Shi, Y., and Gill, G. (2006). NXP-2 association with SUMO-2 depends on lysines required for transcriptional repression. Proceedings of the National Academy of Sciences 103, 5308-5313.

Saksouk, N., Barth, Teresa K., Ziegler-Birling, C., Olova, N., Nowak, A., Rey, E., Mateos-Langerak, J., Urbach, S., Reik, W., Torres-Padilla, M.-E., et al., (2014). Redundant Mechanisms to Form Silent Chromatin at Pericentromeric Regions Rely on BEND3 and DNA Methylation. Molecular Cell 56, 580-594.

Takahashi, K., Yoshida, N., Murakami, N., Kawata, K., Ishizaki, H., Tanaka-Okamoto, M., Miyoshi, J., Zinn, A.R., Shime, H., and Inoue, N. (2007). Dynamic regulation of p53 subnuclear localization and senescence by MORC3. Molecular biology of the cell 18, 1701-1709.

Tchasovnikarova, I.A., Timms, R.T., Douse, C.H., Roberts, R.C., Dougan, G., Kingston, R.E., Modis, Y., and Lehner, P.J. (2017). Hyper-activation of HUSH complex function by Charcot-Marie-Tooth disease mutation in MORC2. Nat Genet 49, 1035-1044.

Watson, M.L., Zinn, A.R., Inoue, N., Hess, K.D., Cobb, J., Handel, M.A., Halaban, R., Duchene, C.C., Albright, G.M., and Moreadith, R.W. (1998). Identification of morc (microrchidia), a

mutation that results in arrest of spermatogenesis at an early meiotic stage in the mouse. Proceedings of the National Academy of Sciences 95, 14361-14366.

Weiser, N.E., Yang, D.X., Feng, S., Kalinava, N., Brown, K.C., Khanikar, J., Freeberg, M.A., Snyder, M.J., Csankovszki, G., Chan, R.C., et al., (2017). MORC-1 Integrates Nuclear RNAi and Transgenerational Chromatin Architecture to Promote Germline Immortality. Developmental cell 41, 408-423.e407.

Zhang, Y., Klein, B.J., Cox, K.L., Bertulat, B., Tencer, A.H., Holden, M.R., Wright, G.M., Black, J., Cardoso, M.C., Poirier, M.G., et al., (2019). Mechanism for autoinhibition and activation of the MORC3 ATPase. Proceedings of the National Academy of Sciences 116, 6111-6119.

Chapter 3: Transcriptional gene silencing by Arabidopsis *microrchidia* homologues involves the formation of heteromers.

ABSTRACT

Epigenetic gene silencing is of central importance to maintain genome integrity and is mediated by an elaborate interplay between DNA methylation, histone posttranslational modifications, and chromatin remodeling complexes. DNA methylation and repressive histone marks usually correlate with transcriptionally silent heterochromatin, however there are exceptions to this relationship. In Arabidopsis, mutation of Morpheus Molecule 1 (MOM1) causes transcriptional derepression of heterochromatin independently of changes in DNA methylation. More recently, two Arabidopsis homologues of mouse microrchidia (MORC) genes have also been implicated in gene silencing and heterochromatin condensation without altering genome-wide DNA methylation patterns. In this study, we show that Arabidopsis microrchidia (AtMORC6) physically interacts with AtMORC1 and with its close homologue, AtMORC2, in two mutually exclusive protein complexes. RNA-sequencing analyses of high-order mutants indicate that AtMORC1 and AtMORC2 act redundantly to repress a common set of loci. We also examined genetic interactions between AtMORC6 and MOM1 pathways. Although AtMORC6 and MOM1 control the silencing of a very similar set of genomic loci, we observed synergistic transcriptional regulation in the mom1/atmorc6 double mutant, suggesting that these epigenetic regulators act mainly by different silencing mechanisms.

INTRODUCTION

DNA methylation and histone posttranslational modifications are essential for silencing of transposable elements (TEs) and other repeat sequences. In the plant model organism Arabidopsis thaliana, DNA methylation sites are found in three different cytosine contexts: CG, CHG, and CHH (in which H is A, T, or C) (Law et al., 2010). Symmetric CG and CHG

methylations are mediated by DNA Methyltransferase 1 (MET1) and Chromomethylase 3 (CMT3), respectively (Jackson et al., 2002, Ronemus et al., 1996). Asymmetric CHH methylation is maintained at nonoverlapping sites by CMT2 and Domains Rearranged Methyltransferase 2 (DRM2) (Stroud et al., 2014, Zemach et al., 2015). In the RNA-directed DNA methylation (RdDM) pathway, de novo methylation of CHH sites is established by DRM2 and involves 24-nucleotide small interfering RNAs and long noncoding RNAs (Cao et al., 2002, Chan et al., 2004, Herr et al., 2005, Law et al., 2013, Onodera et al., 2005, Johnson et al., 2011). Genome-wide studies revealed that DNA methylation and repressive histone modifications such as dimethylation of histone 3 lysine 9 (H3K9me2) correlate with transcriptionally silent chromatin (Bernatavichute et al., 2008, Cokus et al., 2008, Du et al., 2012, Johnson et al., 2002, Tariq et al., 2003). Furthermore, transcriptional derepression of silenced methylated loci is accompanied by loss of DNA methylation. A prominent exception to this interdependence is the Morpheus Molecule 1 (MOM1).

MOM1 is unique to the plant kingdom and was identified in a random transfer-DNA (T-DNA) insertion screen reporting the derepression of a silenced transgene (Amedeo et al., 2000). The mom1 mutant shows a loss of transcriptional gene silencing at loci located predominantly in the pericentromeric regions of the chromosomes (Yokthongwattana et al., 2010). Interestingly, these transcriptional gene-silencing defects occur without major changes in DNA methylation or histone marks (Amedeo et al., 2000, Yokthongwattana et al., 2010, Habu et al., 2006, Probst et al., 2003, Vaillant et al., 2006).

To date, the extent to which MOM1 is implicated in RdDM as well as its molecular mechanism of action remain poorly understood. Because MOM1 shows partial sequence similarities to chromodomain–helicase–DNA binding proteins, it has been proposed that MOM1

is involved in heterochromatin compaction (Amedeo et al., 2000, Calkowski et al., 2008). However, the mom1 mutant does not show any heterochromatin decondensation (Probst et al., 2003, Mittelsten et al., 2002).

Recently, members of the Arabidopsis microrchidia (AtMORC) ATPase family have also been shown to be involved in transposon repression and gene silencing (Brabbs et al., 2013, Lorkovic et al., 2012a, Moissiard et al., 2012). The MORC1 gene was originally described in mice, where it was found to be essential for male primordial germ cell development (Inoue et al., 1999, Watson et al., 1998). The Arabidopsis genome contains seven MORC homologs, which were termed AtMORC1 [NP_568000; AT4G36290; Compromized Recognition of Turnip Crinkle Virus 1 (CRT1)], AtMORC2 [NP_195351; AT4G36280; CRT1–Homolog 1 (CRH1)], AtMORC3 (NP_195350; AT4G36270; CRH2), AtMORC4 (NP_199891; AT5G50780; CRH4), AtMORC5 (NP_196817; AT5G13130; CRH5), AtMORC6 [NP_173344; AT1G19100; CRH6; Defective in Meristem Silencing 11 (DMS11)], and AtMORC7 (NP_194227; AT4G24970; CRH3) (Lorkovic et al., 2012a, Langen et al., 2014, Kang et al., 2008a, Kang et al., 2008b, Kang et al., 2010). AtMORC1 and AtMORC2 are the most closely related homologs and share 80.9% amino acid sequence identity (Langen et al., 2014, Kang et al., 2008a, Kang et al., 2008b, Kang et al., 2010) (Figure 3-1). AtMORC6 has been identified in four independent forward genetic screens (Brabbs et al., 2013, Lorkovik et al., 2012, Moissiard et al., 2012, Kang et al., 2008) as required for gene silencing and maintenance of heterochromatin integrity. AtMORC1 is also required for gene silencing (Moissiard et al., 2012), although it was originally described as a master regulator in plant disease resistance signaling (Kang et al., 2008a, Kang et al., 2008b, Kang et al., 2010, Kang et al., 2012).

Currently, the molecular mechanisms by which the different AtMORC homologs achieve gene silencing remain to be elucidated. AtMORC proteins carry a gyrase, Hsp90, histidine kinase, and MutL (GHKL) domain together with an S5 domain that constitute an active adenosine triphosphatase (ATPase) module (Inoue et al., 1999, Kang et al., 2008b, Dutta et al., 2000). They also carry a putative C-terminal coiled-coil domain (Inoue et al., 1999). In vitro assays showed that both AtMORC1 and AtMORC6 are bona fide ATPases (Moissiard et al., 2012, Kang et al., 2008b). A modest reduction of DNA methylation and repressive histone marks at specific RdDM target sites in atmorc6 mutant suggested that AtMORC6 could also play a role in RdDM (Brabbs et al., 2013, Lorkovic et al., 2012). However, whole genome sequencing analyses of DNA methylation and H3K9me2 in atmorc1 and atmorc6 did not reveal significant differences compared with the wild-type level either in the genome at large or at sites of the highest level of gene derepression in atmorc mutants (Moissiard et al., 2012). Therefore, it is unlikely that the predominant function of AtMORC proteins is maintenance of DNA methylation and H3K9me2, although some interaction with the RdDM pathway seems likely.

In this study, we describe the physical interactions between three different AtMORC homologs and their functional implication in gene silencing. Biochemical analyses indicate that AtMORC6 forms mutually exclusive heteromers with AtMORC1 and its close homolog, AtMORC2. RNA-sequencing (RNA-seq) analyses of high-order mutants show that AtMORC1 and AtMORC2 act redundantly to repress a set of TEs similar to AtMORC6. Furthermore, we also examined the relationship between AtMORC6- and MOM1-mediated silencing as both pathways have only minor impacts on genome-wide DNA methylation. Interestingly, we observed a synergistic effect on transposon derepression, suggesting that these epigenetic regulators act by independent silencing mechanisms.

RESULTS AND DISCUSSION

AtMORC6 Interacts in Vivo with AtMORC1 and AtMORC2 to Form Distinct Heteromers.

Previous analyses showed similar transcriptional derepression between the single *atmorc6* single mutant and the *atmorc1/atmorc6* double mutant, suggesting that *AtMORC1* and *AtMORC6* could interact to enforce gene silencing (Moissiard et al., 2012). To test this hypothesis, FLAG epitope-tagged *AtMORC1* and *AtMORC6* under their respective endogenous promoters were introduced into *cmt3/atmorc1-3* and *atmorc6-1* lines, respectively. Western blotting analyses confirmed that both AtMORC1-FLAG and AtMORC6-FLAG were expressed in their respective mutant background and could complement the *suppressor of drm2 cmt3* (*SDC*)*::GFP* silencing defects (Figure 3-1). These lines were subsequently used to immunoprecipitate FLAG-tagged AtMORC proteins from leaf tissue, and mass spectrometry (MS) analyses were performed to determine potential interacting proteins. MS analyses indicated that AtMORC1 was strongly immunoprecipitated with AtMORC6-FLAG and vice versa (Table 3-1). This interaction was validated by coimmunoprecipitation (co-IP) using F_1 transgenic plant lines expressing complementing AtMORC1-myelocytomatosis (MYC) (Moissiard et al., 2012) and AtMORC6-FLAG (Figure 3-2).

To further characterize the interaction between AtMORC1 and AtMORC6, we performed gel filtration experiments. Leaf protein extracts from epitope-tagged lines were separated on a Superdex 200 10/300GL column, and the eluted fractions were probed by immunoblotting. We observed that both AtMORC1-FLAG and AtMORC6-FLAG were predominantly eluting around 200–300 KDa, suggesting that AtMORC proteins are primarily existing in vivo as dimers (Figure

3-3). Together with the co-IP experiments, these results indicate that AtMORC1 and AtMORC6 are primarily found in vivo as heteromers, most likely as heterodimers. Nevertheless, it cannot be completely ruled out that AtMORC proteins might also form heterotetramers or higher molecular weight complexes, as we observed some signal in fractions with predicted sizes up to several hundred kilodaltons.

MS analysis of FLAG-tagged AtMORC6 IPs revealed an additional interaction with the closest homolog of AtMORC1, AtMORC2 (Table 3-1). This result is consistent with a recent independent study that also found peptides of AtMORC1 and AtMORC2 in an IP–MS of AtMORC6 in flowers (Liu et al., 2014). Interestingly, AtMORC2 was not immunoprecipitated with AtMORC1, suggesting that AtMORC6 was interacting with AtMORC1 and AtMORC2 in two distinct complexes (Table 3-1). To validate the heteromerization between AtMORC6 and AtMORC2, we engineered a complementing transgenic line expressing FLAG-tagged AtMORC2 in an *atmorc1*/*atmorc2* background (Figure 3-1) and performed IP followed by MS. MS analysis showed that AtMORC6 was immunoprecipitated with FLAG-AtMORC2 (Table 3-1). Consistent with this interaction, gel filtration analysis of FLAG-AtMORC2 leaf extracts showed that FLAG-AtMORC2 was principally present in the elution fractions around 200–300 KDa, corresponding to similar elution fractions as AtMORC6-FLAG (Figure 3-2). In summary, our biochemical analyses indicate that AtMORC6 physically interacts with AtMORC1 and AtMORC2 in the form of two mutually exclusive heteromers.

AtMORC6 was shown to interact in vitro with DMS3 when both proteins were coexpressed in *Escherichia coli*, providing a physical link to the RdDM pathway (Lorkovic et al., 2012a). DMS3 is a structural maintenance of chromosomes hinge domain-containing protein that lacks an ATPase domain (Kanno et al., 2008). Based on the stimulation of AtMORC6

ATPase activity by in vitro interaction with DMS3, it was proposed that AtMORC6 and DMS3 cooperate to promote transcriptional repression. DMS3 has also been shown to interact with additional components of the DRD1-DMS3-RDM1 (DDR) complex including Defective in RNA-Directed DNA Methylation 1 (DRD1) or RDM1 as well as with the largest subunit of PolV (Law et al., 2010). Furthermore, genome-wide association of PolV to chromatin and thus the production of PolV-dependent transcripts and subsequent DNA methylation are dependent on all members of the DDR complex (Law et al., 2010, Zhong et al., 2012). However, we did not detect DMS3 or other components of the DDR complex in our IP–MS experiments. Also, previous IP–MS experiments using FLAG-tagged DRD1 and DMS3 proteins as bait did not immunoprecipitate AtMORC6 (Zhong et al., 2012). Nevertheless, we cannot rule out that the interactions between components of the DDR complex and AtMORC6 are weak or ephemeral and could not be detected under our IP conditions.

A recent study found that AtMORC6 was immunoprecipitated in flowers in very small amounts with SUVH9, an SRA- (SET [suppressor of variegation 3–9 [Su(var)3–9], enhancer of zeste [E(z)], and trithorax (Trx)] and RING [really interesting new gene] associated)- and SET-domain-containing protein (Liu et al., 2014). SUVH9 and its closest homolog, SUVH2, were shown to bind methylated DNA and recruit PolV to chromatin through an interaction with the DDR complex (Johnson et al., 2014, Liu et al., 2014, Johnson et al., 2008). Yeast two-hybrid assays further indicated that the interactions between AtMORC proteins and SUVH proteins were direct (Liu et al., 2014). These data, together with the slight changes observed in DNA methylation of certain RdDM target loci (Brabbs et al., 2013, Lorkovic et al., 2012a, Lorkovic et al, 2012b), suggest that AtMORC proteins modulate RdDM through interactions with the DDR complex and SUVH proteins. Nevertheless, the mild changes of small RNAs and DNA

methylation genome-wide in *atmorc* mutants (Moissiard et al., 2012) suggest that AtMORCs are unlikely to be canonical RdDM factors. It is also plausible that AtMORCs contribute to processing of target loci transcripts, thus leading to posttranslational silencing. Future experiments are needed to clarify the precise function in gene silencing and degree of involvement of AtMORCs in the RdDM pathway.

AtMORC2 Acts Redundantly with AtMORC1 to Achieve Gene Silencing.

To further study the role of *AtMORC2* in gene silencing and its functional relationship with *AtMORC1* and *AtMORC6*, we generated high-order mutants and performed transcriptional profiling analyses. Real-time PCR (RT-PCR) from RNA extracted from leaf tissue indicated that *SDC* was derepressed in *atmorc1* but not *atmorc2* (Figure 3-3b) consistent with the fact that *AtMORC2* was not identified in the genetic screens that identified *AtMORC1*and *AtMORC6* (Brabbs et al., 2013, Lorkovic et al., 2012, Moissiard et al., 2012, Kang et al., 2008). RT-PCR also showed an increased derepression of two transposons, *AtCopia28* and *RomaniaT5*, in the *atmorc1*/*atmorc2* double mutant compared with *atmorc1* and *atmorc2* single mutants (Figure 3-3b) indicating that *AtMORC1* and *AtMORC2* act redundantly in transposon silencing. Further genome-wide characterization of the transcriptome by RNA-seq indicated that only two transposons was significantly up-regulated in *atmorc2* compared with wild type [using a very stringent cutoff of fold change ≥4; false discovery rate (FDR) < 0.05], whereas nine TEs were up-regulated in *atmorc1* (Figure 3-3c) Transcriptional derepression of protein-coding genes was also more pronounced in *atmorc1* compared with *atmorc2* (Figure 3-4a). Publicly available microarray data indicate that expression of *AtMORC1* is higher than *AtMORC2* in most tissues and

developmental stages (Figure 3-5a), providing a plausible explanation for the stronger silencing defects observed in *atmorc1* compared with *atmorc2*. Interestingly, combined deletion of *AtMORC1* and *AtMORC2* led to significantly higher transcription of TEs and protein-coding genes compared with both single mutants (Figure 3-3c, e, f and Figure 3-4a, c, d), confirming that *AtMORC1* and *AtMORC2* are functionally redundant. In addition, the overexpression of FLAG-AtMORC2 succeeded in complementing transcriptional derepression in the *atmorc1*/*atmorc2* double mutant (Figure 3-1d).

The observed redundancy between *AtMORC1* and *AtMORC2* and their physical interaction with AtMORC6 in two mutually exclusive heteromers predict that a loss of *AtMORC6* should be phenotypically comparable to the combined loss of *AtMORC1* and *AtMORC2*. To test this hypothesis, we compared the transcriptomes of *atmorc1/atmorc2* with the *atmorc6* single mutant. RNA-seq revealed a high overlap of transcriptional derepression between *atmorc1/atmorc2* and *atmorc6* (Figure 3-3 D-F and Figure 3-4 B-D), supporting the notion that *AtMORC6* function is epistatic to both *AtMORC1* and *AtMORC2* combined. Derepressed transposons were not restricted to a specific family in any of the mutant backgrounds analyzed (Figure 3-5B) Finally, the observed transcriptional derepression did not significantly increase in a triple mutant lacking *AtMORC1*, *AtMORC2*, and *AtMORC6* (Figure 3-3D-F and Figure 3-4 B-D). These results are consistent with the model that AtMORC6 interacts exclusively with either AtMORC1 or AtMORC2 to achieve gene silencing and that *AtMORC1* is functionally redundant with *AtMORC2*.

It appeared that up-regulated genes were preferentially localized in H3K9me2-enriched heterochromatin (Bernatavichute et al., 2008) even though they are protein-coding (Figure 3-4E).

This is in agreement with the previous observations that *AtMORC1* and *AtMORC6* are mainly involved in silencing and compaction of heterochromatin (Moissiard et al., 2012). Gene ontology term analysis using AmiGO (Carbon et al., 2009) of all up-regulated protein-coding genes indicated enrichments ($P < 6e-4$) in response to chitin and in response to organonitrogen compounds in *atmorc1*/*atmorc2* and in *atmorc1*/*atmorc2*/*atmorc6*. It is interesting to note that chitin has been recognized as a general elicitor of plant defense responses (Boller 1995), which is in agreement with the reported implication of *AtMORC1* in plant immunity (Kang et al., 2008b). To assess if protein-coding genes up-regulated in *atmorc6* were also targets of the RdDM machinery, we looked at their expression in a mutant lacking the methyltransferases *DRM1* and *DRM2* that is thus defective in RdDM (Stroud et al., 2014). These were not significantly up-regulated in *drm1*/*drm2* (Figure 3-4F) indicating that AtMORCs are unlikely to be canonical RdDM factors.

Our combined genetics and RNA-seq data show that the simultaneous absence of *AtMORC1* and *AtMORC2* in *atmorc1*/*atmorc2* cannot be functionally compensated by the presence of *AtMORC6* alone (Figure 3-3 and Figure 3-4). Also, the loss of *AtMORC6* in *atmorc6* cannot be compensated by the presence of *AtMORC1* and *AtMORC2* (Figure 3-3 and Figure 3-4). Furthermore, the *atmorc1*/*atmorc2*/*atmorc6* triple mutant does not have a stronger phenotype than the *atmorc1*/*atmorc2* double mutant (Figure 3-3 B and D-F and Figure 3-4 B-D). Together with the observation that AtMORC1 and AtMORC2 did not interact, these results lead to the conclusion that AtMORCs function as heteromers and not as homomers.

AtMORC6 and MOM1 Act Synergistically to Silence a Common Set of Transposons.

*AtMORC1*and *AtMORC6* were identified in a forward genetic screen reporting the derepression of an *SDC::GFP*transgene in wild type or in the *cmt3* mutant background (Moissiard et al., 2012). Further screening of ethyl methanesulfonate (EMS) mutagenized seeds followed by deep genome resequencing identified two new alleles of *AtMORC6*in the *cmt3* background. In the first line, *cmt3* 262, glycine 212 was mutated to glutamic acid, and in *cmt3*379, a guanine (chr1:6599258) was mutated to adenine in the splice site before exon 14. Interestingly, we also identified three loss-of-function alleles of the *MOM1* gene in the same genetic screen. The EMS mutations in these new *mom1* alleles were a stop codon introduced at amino acid 603 (line 337 in a wild-type background), a stop codon introduced at amino acid 586 (*cmt3* 265), and a substitution of Leucine 656 to Phenylalanine (*cmt3* 113).

MOM1 is unique to the plant kingdom and has no homologs in the *Arabidopsis* genome. Previous studies showed that DNA methylation in *mom1* mutants was similar to the wild-type level (Amadeo et al., 2000, Yokthongwattana et al., 2010, Habu Y, et al., 2006, Vaillant et al., 2006). This observation was recently confirmed by genome-wide bisulfite-sequencing (BS-seq) analyses (Stroud et al., 2013). RNA-seq analyses showed that 52 TEs were significantly up-regulated in *mom1* using similarly stringent cutoffs as for *atmorc* mutants (Figure 3-5A), and we found that the DNA methylation levels of these TEs also remained unchanged in *mom1* compared with wild type (Figure 3-5D). Nineteen transposons were significantly derepressed in *atmorc6* in this experiment, and most of these were also derepressed in *mom1* (Figure 3-5A). The numbers of TEs significantly up-regulated in *atmorc6* slightly vary between the two RNA-seq experiments performed (Figure 3-3D and Figure 3-5A) because both experiments were done independently. As shown previously, DNA methylation was not significantly changed in TEs up-regulated in *atmorc6* (Moissiard et al., 2012) (Figure 3-5D).

These data indicate that overall transcriptional derepression is higher in *mom1* compared with *atmorc6* and that *MOM1*and *AtMORC6* mediate the silencing of a subset of common targets as well as of a number of independent loci.

To further understand the relationship between *MOM1*- and *AtMORC6*-mediated transcriptional silencing, we generated a double mutant lacking *MOM1* and *AtMORC6*. RNA-seq analyses in *mom1*/*atmorc6* showed a significant increase in derepression of TEs and to a smaller extent of protein-coding genes compared with both single mutants (Figure 3-5A-C and Figure 3-6 A–C*)*. RT-PCR analyses corroborated the synergistic derepression of *SDC* and *RomaniatT5* (Figure 3-7A). Overexpressed TEs in all three genotypes profiled by RNA-seq are predominantly located in the pericentromeric heterochromatin and belong to diverse families, consistent with previous reports (Yokthongwattana et al., 2010, Moissiard et al., 2012) (Figure 3-7 B and C). Genome-wide BS-seq analysis showed that DNA methylation was unchanged in TEs up-regulated in *mom1*/*atmorc6* (Figure 3-7D). Similar to *AtMORC6* target loci, protein-coding genes significantly up-regulated in *mom1* were preferentially located in heterochromatin (Figure 3-8D). Furthermore, transcription of these was not affected in the *drm1*/*drm2* mutant, suggesting a limited role of *MOM1* in RdDM (Figure 3-8E). Altogether, these results indicate that *AtMORC6* and *MOM1* act synergistically to silence a largely common set of heterochromatic DNA elements through two independent pathways.

CONCLUSION

In this study, we combined biochemistry, genetics, and genomics to understand further the mode of action of the recently discovered *Arabidopsis* MORC homologs. We found that AtMORC6-mediated transcriptional silencing requires the formation of mutually exclusive heteromers with

AtMORC1 and its closest homolog, AtMORC2. Further biochemical studies involving domain deletions or point mutations should uncover the molecular mechanisms of the AtMORC proteins and the implication of heteromerization for ATPase activity. It is interesting to note the similarities between AtMORCs and the structural maintenance of chromosome proteins cohesin and condensin (Wood et al., 2010). These three protein families are ATPases that function in vivo as heteromers and modulate chromatin superstructure to regulate proper expression and maintenance of genomic integrity.

Genetic and RNA-seq analyses showed that *AtMORC6* acts synergistically with the putative chromatin remodeler *MOM1* to silence a common set of heterochromatin-localized loci. The synergistic effect observed in the *mom1*/*atmorc6* double mutant suggests that *AtMORC6* and *MOM1* act in two convergent pathways that are both required for the proper silencing of pericentromeric heterochromatin. It has been previously shown that AtMORC6 and AtMORC1 accumulate in the nucleus as discrete nuclear bodies that localize in the vicinity of the heterochromatic chromocenters (Moissiard et al., 2012). It will be interesting to determine in the future whether MOM1 accumulates in a similar fashion in the nucleus to form distinct nuclear bodies. The identification of MOM1 interactors will also be crucial to understanding its mode of action.

Figure 3-1: Phylogenetic analysis and epitope-tagging of Arabidopsis Microrchidia

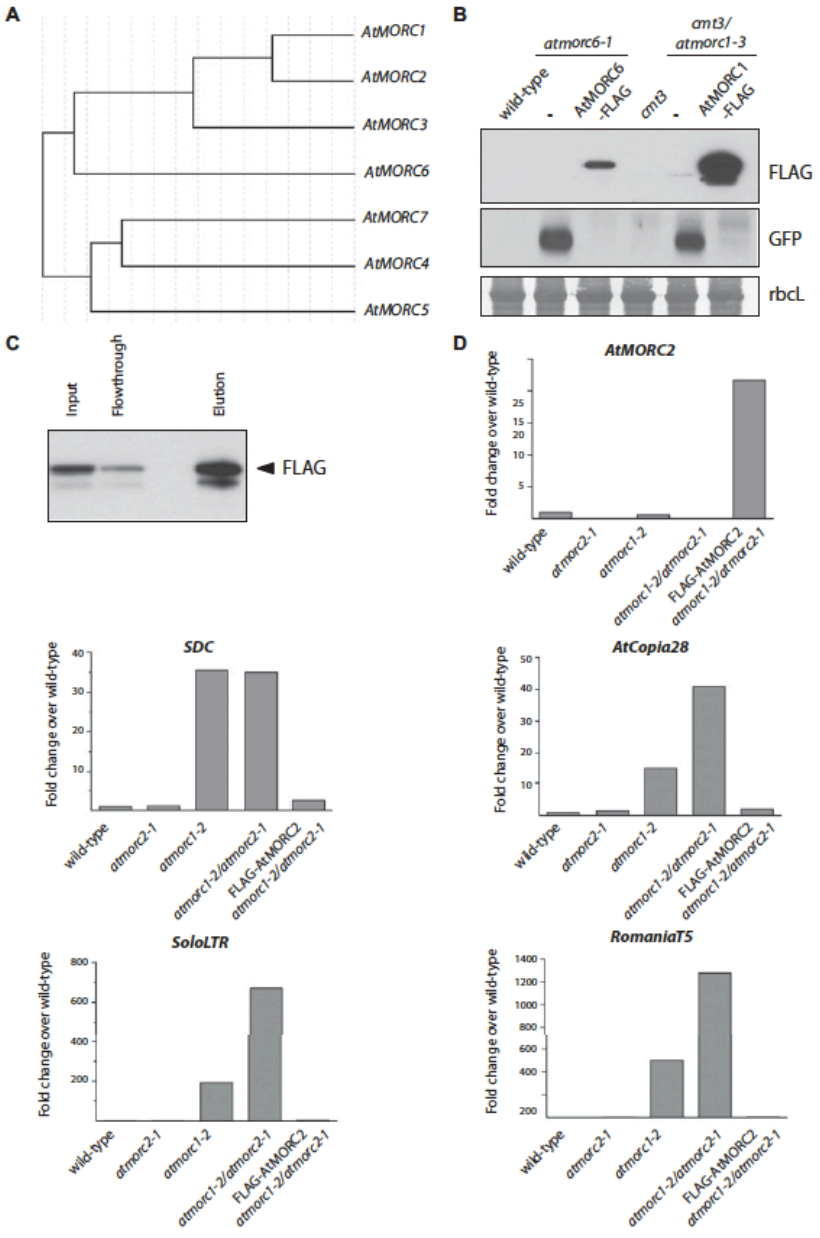


Figure 3-2: AtMORC1, AtMORC2, and AtMORC6 co-purify by gel filtration.

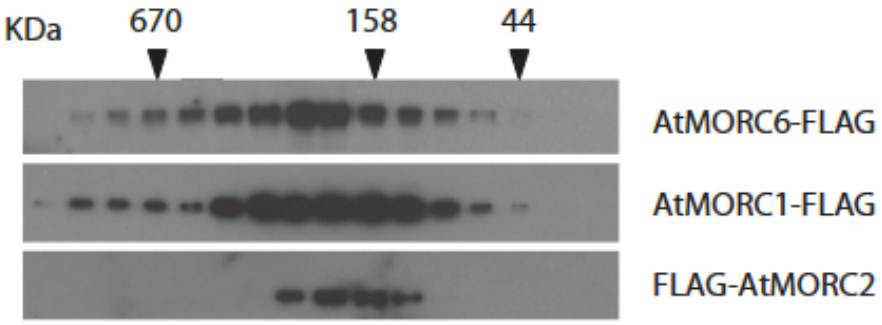


Figure 3-3: Redundancy of AtMORC1 and AtMORC2 in transposon silencing.

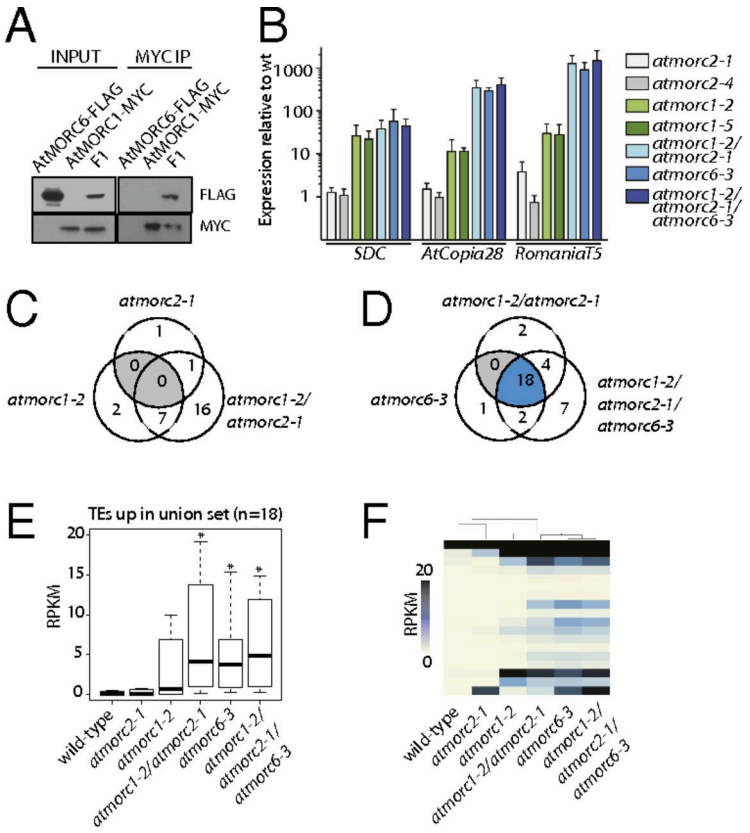


Figure 3-4: Redundancy of AtMORC1 and AtMORC2 in gene silencing.

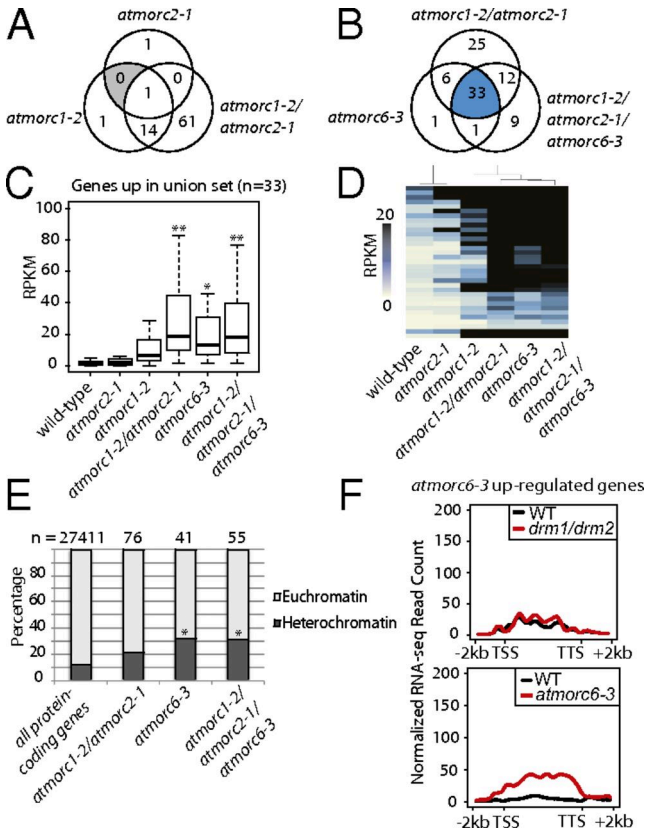


Figure 3-5: Synergy of AtMORC6 and MOM1 in transposon silencing.

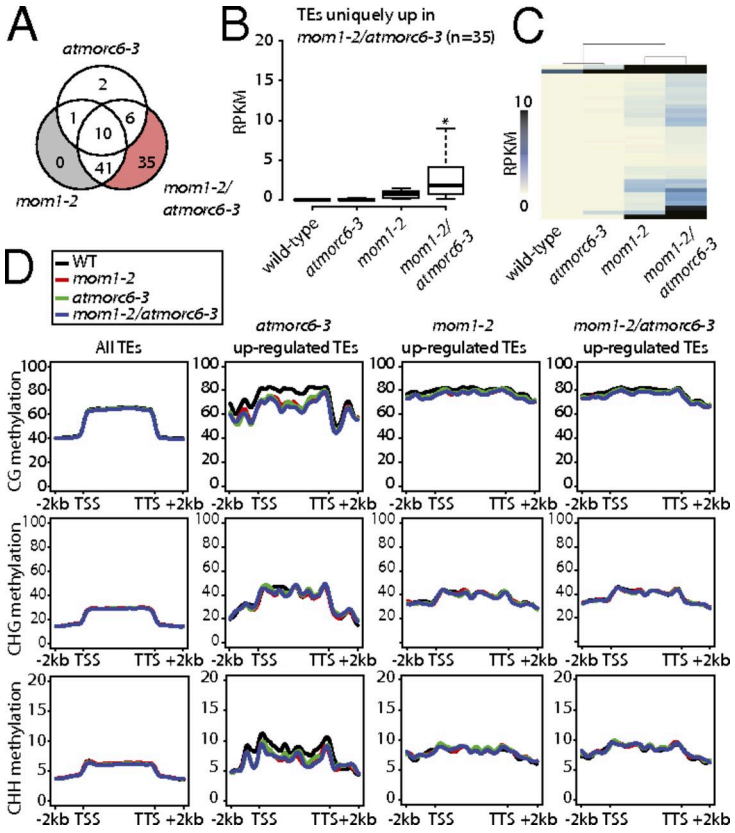


Figure 3-6: Synergy of AtMORC6 and MOM1 in gene silencing.

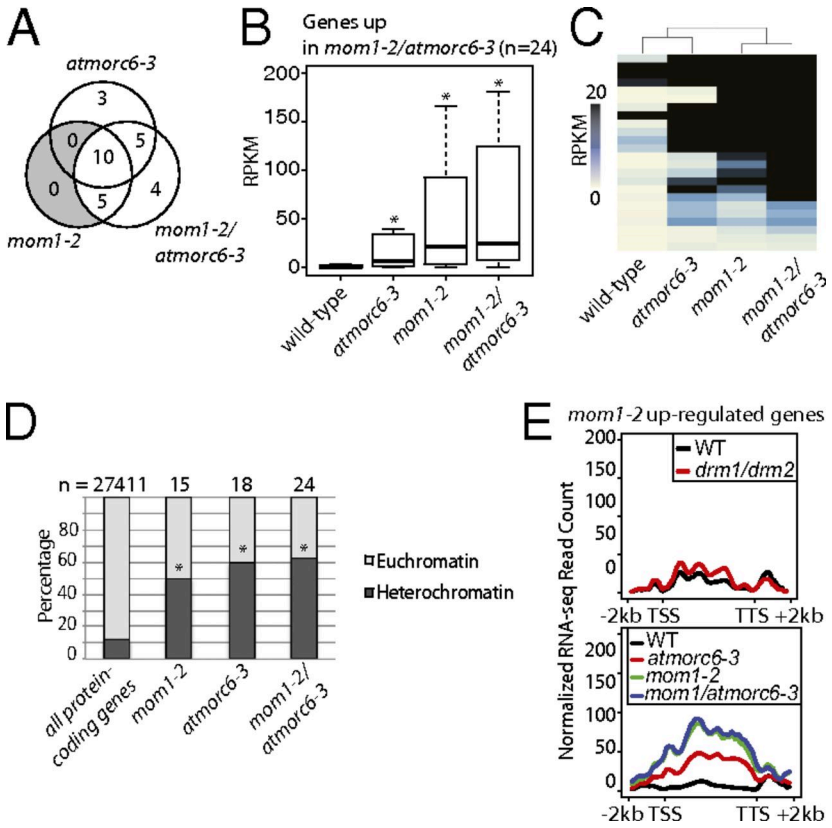


Figure 3-7: Analysis of AtMORCs' expression and derepressed transposon families in *atmorc* mutants.

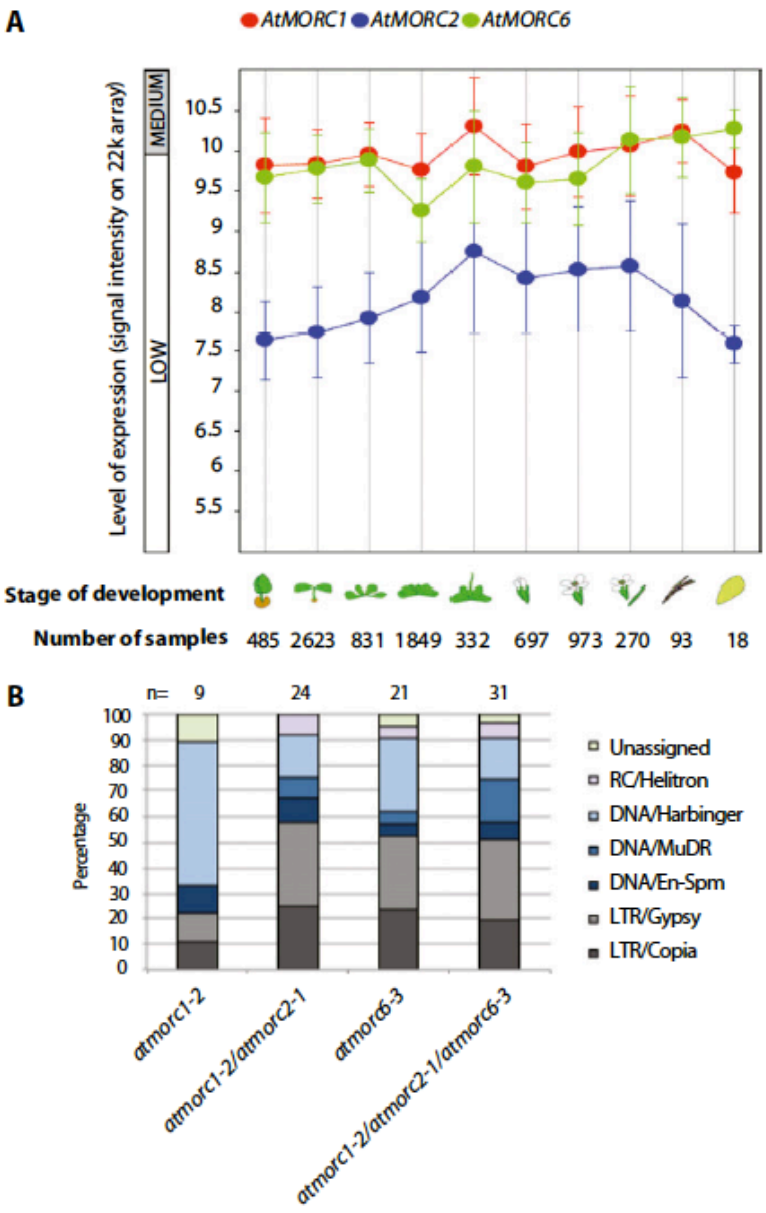


Figure 3-8: Synergy of AtMORC6 and Morpheus Molecule 1 (MOM1) in transposon silencing.

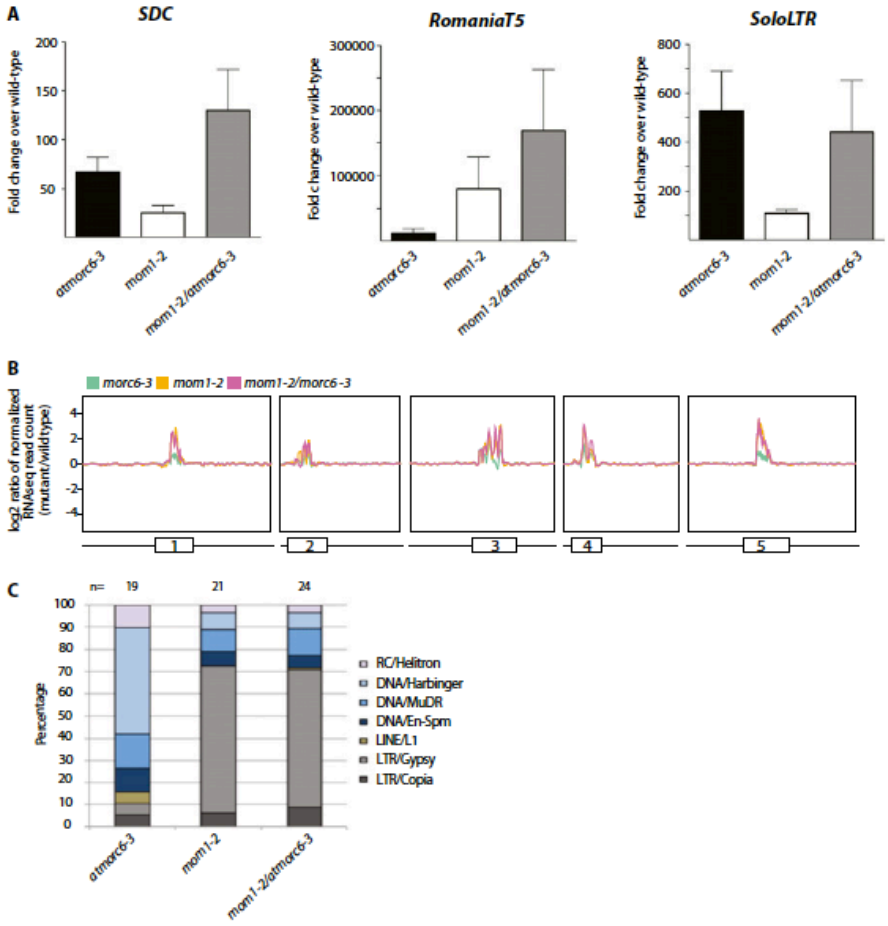


Figure legends

Figure 3-1. Phylogenetic analysis and epitope-tagging of Arabidopsis Microrchidia AtMORC1, AtMORC2, and AtMORC6.

(A) Phylogenetic analysis of the AtMORC gene family in Arabidopsis thaliana. The alignment was made with Genebee using the default parameters. (B) AtMORC1 and AtMORC6 transgenic lines. FLAG epitope-tagged AtMORC1 and AtMORC6 were expressed under their respective endogenous promoter in their respective mutant background. Protein expression and complementation of the SDC::GFP silencing defects were probed by Western blotting. The large subunit of rubisco (rbcL) was used as the loading control. (C) Transgenic AtMORC2 line. FLAG epitope-tagged AtMORC2 was expressed under its respective endogenous promoter in the *atmorc1/atmorc2* double mutant background. FLAG-AtMORC2 is enriched in the elution fraction after immunoprecipitation. (D) Complementation of transcriptional derepression by expression of FLAG-AtMORC2 in atmorc1/atmorc2. RT-PCR shows increased levels of FLAG-AtMORC2 transcripts compared with wild-type and atmorc1-2 despite being expressed under its respective endogenous promoter. Derepression of suppressor of drm2 cmt3 (SDC)::GFP, AtCopia28, Solo long terminal repeat (LTR), and RomaniaT5 is suppressed by overexpression of FLAG-AtMORC2.

Figure 3-2. AtMORC1, AtMORC2, and AtMORC6 co-purify by gel filtration.

AtMORC6-FLAG, AtMORC1-FLAG, and FLAG-AtMORC2 leaf protein extracts were separated on a Superdex 200 10/300GL column. Eluting fractions were immunoblotted with antibodies directed against FLAG. Sizing standards are shown.

Figure 3-3. Redundancy of AtMORC1 and AtMORC2 in transposon silencing.

(A) AtMORC1 physically interacts with AtMORC6. AtMORC6-FLAG was coimmunoprecipitated with AtMORC1-MYC in F1 plants expressing both epitope tagged proteins. Epitope-tagged proteins were detected by Western blotting. (B) RT-PCR assessing endogenous expression of SDC, AtCopia28, and RomaniaT5. Three biological replicates were performed for each tested genotype. Two individual alleles were used for atmorc1 and atmorc2. (C and D) Venn diagrams of overlap between TEs up-regulated (fourfold increase; FDR, 0.05; Fisher's exact test) in each genotype. Gray regions represent categories with no TEs counted. Blue shading represents the union set of TEs up-regulated in *atmorc* mutants. (E) Boxplot and (F) heatmap of average reads per kilo base per million (RPKM) values between two biological replicates for TEs in a union set for different genotypes. An asterisk indicates a significant increase relative to wild-type samples ($P < 1e\text{-}3$, Mann–Whitney U test).

Figure 3-4. Redundancy of AtMORC1 and AtMORC2 in gene silencing.

(A and B) Venn diagrams showing relationships between sets of protein-coding genes called up-regulated (fourfold increase in expression; FDR < 0.05) for different genotypes. Gray regions represent categories with no gene counted. Blue shading represents the union set of genes up-regulated in *atmorc* mutants. (C) Boxplot and (D) heatmap of average RPKM values for different genotypes (two biological replicates) for protein-coding genes in a union set for different genotypes. An asterisk indicates a significant increase relative to wildtype samples ($P < 1e\text{-}8$, Mann–Whitney U test). Two asterisks represent a significant increase relative to wild-type samples and the *atmorc1* single mutant ($P < 1e\text{-}2$, Mann–Whitney U test). (E)

Overrepresentation in H3K9me2-enriched heterochromatin of protein-coding genes significantly up-regulated in *atmorc1-2/atmorc2-1, atmorc6-3,* or *atmorc1-2/atmorc2-1/atmorc6-3* mutants. An asterisk indicates a significant increase relative to all protein-coding genes (P < 1e-3, Fisher's exact test). (F) Metagene analysis of RNA-seq reads over protein-coding genes called up-regulated in *atmorc1-2/atmorc2-1, atmorc6-3, or atmorc1-2/atmorc2-1/atmorc6-3* mutants. Reads are derived from previously published RNA-seq libraries for two replicates of the *drm1/drm2* double mutant and the corresponding wild type (WT).

Figure 3-5. Synergy of AtMORC6 and MOM1 in transposon silencing.

(A) Venn diagram showing relationships between sets of TEs called up-regulated (fourfold increase in expression; FDR < 0.05) for different genotypes. Grayed regions highlight sets with no elements, and red shading highlights TEs uniquely called up-regulated in the higher order mutant. (B) Boxplot and (C) heat map of average RPKM values between two biological replicates for TEs uniquely called up-regulated in the *mom1/atmorc6* mutant background for different genotypes. An asterisk indicates a significant increase relative to all other genotypes (P < 1e-8, Mann–Whitney U test). (D) Metagene analysis of DNA methylation levels across all Arabidopsis TEs for the *atmorc6-3, mom1-2, mom1-2/atmorc6-3*, and wild-type genotypes. Also shown are the methylation levels at TEs up-regulated in mutant genotypes.

Figure 3-6. Synergy of AtMORC6 and MOM1 in gene silencing.

(A) Venn diagram showing relationships between sets of protein-coding genes called up-regulated (four-fold increase in expression; FDR < 0.05) for different genotypes. Grayed regions highlight sets with no elements. (B) Boxplot and (C) heat map of average RPKM values for different genotypes (two biological replicates) for protein-coding genes uniquely called up-

regulated in the *mom1/atmorc6* mutant background. An asterisk represents a significant increase relative to wild-type samples (P < 1e-2, Mann–Whitney U test). (D) Overrepresentation in H3K9me2-enriched heterochromatin of protein-coding genes significantly up-regulated in *atmorc6-3, mom1-2*, or *mom1-2/atmorc6-3* mutants. An asterisk indicates a significant increase relative to all protein-coding genes (P < 1e-3, Fisher's exact test). (E) Metagene analysis of RNA-seq reads over protein coding genes called up-regulated in *atmorc6-3, mom1-2,* or *mom1-2/atmorc6-3* mutants. Reads are derived from previously published RNA-seq libraries for two replicates of the *drm1/drm2* double mutant and the corresponding wild type (WT).

Figure 3-7. Analysis of AtMORCs' expression and derepressed transposon families in atmorc mutants. (A) Transcript levels of AtMORC1, AtMORC2, and AtMORC6 over stages of development. Gene expression data were retrieved from Genevestigator (www.genevestigator.com). (B) Transposons significantly derepressed in the different genotypes grouped according to their family.

Figure 3-8. Synergy of AtMORC6 and Morpheus Molecule 1 (MOM1) in transposon silencing. (A) RT-PCR assessing endogenous expression of SDC, RomaniaT5, and SoloLTR. Two biological replicates were performed for each tested genotype. (B) Chromosomal view of percent change in read abundance between higher order mutants and other genotypes. (C) Transposons significantly derepressed in the different genotypes grouped according to their family.

Table 3-1. FLAG-tagged AtMORC proteins were immunoprecipitated and interacting proteins were analyzed by mass spectrometry. The total numbers of identified spectra, the normalized spectral abundance factor (NSAF) and the percentage relative to the bait protein are given for two biological replicates.

Table 3-2. Primers used in this study

Table 3-1:

AtMORC6-FLAG IP

Name	Accession	Spectra		NSAF		% AtMORC6	
AtMORC6	AT1G19100	77	75	2060	539	100	100
AtMORC1	AT4G36290	62	31	1732	233	84	43
AtMORC2	AT4G36280	35	20	992	152	48	28

AtMORC1-FLAG IP

Name	Accession	Spectra		NSAF		% AtMORC1	
AtMORC1	AT4G36290	76	71	6273	765	100	100
AtMORC6	AT1G19100	11	42	870	434	14	57

FLAG-AtMORC2 IP

Name	Accession	Spectra		NSAF		% AtMORC2	
AtMORC2	AT4G36280	65	-	370	-	100	-
AtMORC6	AT1G19100	32	-	172	-	47	-

Table 3-2: primers used in this study

Genotyping

Genotype	Insertion line	Left border primer	Right border primer
atmorc1-2	SAIL_893_B06	TTGCAGTTTGGAACCAAAATC	TGAGTTTTGACGACGATGATG
atmorc1-4	SAIL_1239_C08	CGTATCTCAGCCGCTAACTTG	AAGCAGCTGCAGTGGATTATG
atmorc1-5	SAIL_131_H11	CGTATCTCAGCCGCTAACTTG	AAGCAGCTGCAGTGGATTATG
atmorc2-1	SALK_072774C	CTACTCAGAGCGTTGGCATTC	GTTGTAGCTGTATGGGGCTTG
atmorc2-4	SALK_021267C	TTTCGTCATCATTGCTTTTCC	GGTTGACTCTTCCACTGCTTG
atmorc6-3	GABI_599B06	GGAAAGCTGGAAGCTATAATGATG	GATGACATCTGCCCCAAGTCTC
cmt3-11	SALK_148381	TAACGGAAGGATGCCAGATT	CAAGAAATGGGCTGTTGACAT
mom1-2	SAIL_610_G01	ACAATGCAGGAGCAAACACTC	GGAAAGGAGATACTTCACCGG
LB3	SAIL T-DNA	TAGCATCTGAATTTCATAACCAATCTCGATACAC	—
LBA1	SALK T-DNA	TGGTTCACGTAGTGGGCCATCG	—
LB O8409	GABI-KAT	ATATTGACCATCATACTCATTGC	—

RT-PCR

Locus	Left border primer	Right border primer
Actin 7	TCGTGGTGGTGAGTTTGTTAC	CAGCATCATCACAAGCATCC
SDC	AATGTAAGTTGTAAACCATTTGAACGTGACC	CAGGCATCCGTAGAACTCATGAGC
AtCopia28	AGTCCTTTTGGTTGCTGAACA	CCGGATGTAGCAACATTCACT
RomaniaT5	GTATCCTTTGGCCCGGTATT	GCCTCTTCGAAATGCCATAA
SoloLTR	AACTAACGTCATTACATACACATCTTG	AATTAGGATCTTGTTTGCCAGCTA

MATERIALS AND METHODS

Plant Material and Growing Conditions.

Wild-type and all mutant lines are from the ecotype Columbia and were grown under continuous light. Plant lines used include *atmorc1-2* (SAIL_893_B06; crt1-2), *atmorc1-4* (SAIL_1239_C08), *atmorc1-5* (SAIL_131_H11; crt1-5), *atmorc2-1* (SALK_072774C; crh1-1), *atmorc2-4* (SALK_021267C; crh1-4), *atmorc6-3* (GABI_599B06), *cmt3-11* (SALK_148381), and *mom1-2*(SAIL_610_G01). EMS mutagenized *atmorc6-1* and *cmt3/morc1-3* lines and complementing AtMORC1-MYC and AtMORC6-MYC lines are described in Moissiard et al., 2012. T-DNA insertions were confirmed by PCR-based genotyping. Primer sequences are described in Table 3-2.

Cloning of pAtMORC1::AtMORC1-FLAG, pAtMORC2::FLAG-AtMORC2, and pAtMORC16::AtMORC6-FLAG.

Cloning was done according to Moissiard et al., 2012. Briefly, AtMORC1 and AtMORC6 genomic regions were PCR amplified and the FLAG epitope was added to the C terminus of AtMORC1 and AtMORC6 and at the N terminus of AtMORC2. The amplified region includes a ~1 Kb promoter sequence upstream of the respective transcriptional start site.

IP and MS Analysis.

Ten grams of 2-wk-old seedling tissue of each epitope-tagged line were ground in liquid nitrogen and resuspended in 45 mL ice-cold IP buffer [50 mM Tris·HCl pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 0.1% Nonidet P-40, 10% (vol/vol) glycerol, 1× Protease Inhibitor Mixture (Roche)] and centrifuged for 10 min at 4 °C at 16,000 × *g*. We added 200 µL M2 magnetic FLAG-beads

(SIGMA, M8823) to the supernatants and incubated it for 60 min rotating at 4 °C. M2 magnetic FLAG-beads were washed five times in ice-cold IP buffer for 5 min rotating at 4 °C, and immunoprecipitated proteins were eluted three times with 100 µL 3×-FLAG peptides (SIGMA, F4799) for 15 min at 25 °C. The eluted protein complexes were precipitated by trichloroacetic acid and subjected to MS analyses as previously described Du et al., 2012.

Co-IP and Immunoblotting.

We ground 1.5 g of 2-wk-old seedling tissue of each epitope-tagged line in liquid nitrogen, resuspended it in 12 mL ice-cold IP buffer [50 mM Tris·HCl pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 0.1% Nonidet P-40, 10% (vol/vol) glycerol, 1× Protease Inhibitor Mixture (Roche)], and centrifuged it for 10 min at 4 °C at 16,000 × *g*. We added 100 µL M2 magnetic FLAG-beads (SIGMA, M8823) or 150 µL MYC-conjugated agarose beads (COVANCE, AFC-150P-1000) to the supernatants and incubated it for 60 min rotating at 4 °C. Beads were washed five times in ice-cold IP buffer for 5 min rotating at 4 °C, and immunoprecipitated proteins were eluted in 1× Lämmli buffer for 15 min at 80 °C. Western blots were performed as previously described Moissiard et al., 2012 with GFP-specific antibody (Invitrogen, AA1122), HRP-coupled FLAG-specific antibody (SIGMA, A8592), and MYC-specific antibody (Pierce, MA1-980).

Gel Filtration.

Gel filtration experiments were performed according to Law et al., 2010. Briefly, 0.5 g of 2-wk-old seedling tissue of each epitope-tagged line were ground in liquid nitrogen and resuspended in 1 mL of ice-cold IP buffer [50 mM Tris·HCl pH 8.0, 150 mM NaCl, 0.1% Nonidet P-40, 10% (vol/vol) glycerol, 1× Protease Inhibitor Mixture (Roche)] and centrifuged for 10 min at 4 °C at

16,000 × *g*. The supernatants were centrifuged again for 10 min at 4 °C at 16,000 × *g*. The supernatants were then centrifuged through a 0.2 μm filter (Millipore), 500 μL were loaded onto a Superdex 200 10/300GL column (GE Healthcare, 17–5175-01) column, and 250 μL fractions were collected. We ran 20 μL of every collected fraction on a 4–12% SDS/PAGE. Before use, the column was equilibrated and calibrated with gel filtration standards (Biorad, 151–1901).

RNA Extraction.

We froze 100 mg of 20-d-old leaf tissue in liquid nitrogen. The frozen leaves were then added to a mortar containing liquid nitrogen. Immediately after the liquid nitrogen boiled off, the leaf tissue was crushed to powder using a pestle. We immediately added 1.2 mL of TRIzol Reagent (Life Technologies 15596) to the cold powder, and then it was pulverized further until a clear, dark brown solution was visible. The solution was transferred to a chilled Eppendorf tube, and 400 μL of chloroform was added. The tube was vortexed for 5 s at maximum power, then spun in a centrifuge at 16,000 × *g* (4 °C) for 10 min to separate the aqueous and organic phases. We collected 700 μL of the aqueous (top) phase. To precipitate the RNA, 700 μL of isopropanol was added to the aqueous material, the solution was vortexed for 5 s at maximum power, and then it was centrifuged for 10 min at 16,000 × *g* (4 °C). The supernatant was removed, and 500 μL of room temperature 80% (vol/vol) ethanol was added to the pellet, which was then spun for 5 min at 16,000 × *g* (4 °C). The supernatant was removed and the pellet was air-dried for 5 to 10 min. The pelleted RNA was resuspended in 100 μL water and then purified using the Qiagen RNeasy Mini (Qiagen 74104) "RNA Cleanup Protocol" according to manufacturer's instructions. RNA was quantified using Nanodrop.

RT-PCR.

We treated 1 μg of input RNA with DNase I (Life Technologies, 18068) according to the manufacturer's protocol. Of the 11 μL final reaction volume, 3 μL was set aside as a negative control for RT-PCR, whereas 8 μL was converted to cDNA using SuperScript III (Life Technologies 18080). We used 5% of cDNA for each RT-PCR. RT-PCR was performed using IQ SYBR Green Supermix (Bio-Rad 170–8880), with 375 nM final primer concentration using a Stratagene Mx3005p instrument. Amplification conditions were as follows: 95 °C 10:00; 40 cycles, 95 °C, 30 s, 55 °C 1:00, 72 °C 1:00; melting curve. At least two technical replicates were performed per biological replicate, and three biological replicates were used in all experiments. Relative abundance of transcripts was calculated using the difference of squares method. Primer sequences are described in Table 3-2.

BS-Seq, RNA-Seq, and Accession Codes.

BS-seq was done according to Moissiard *et al.,* 2012. RNA-seq libraries were generated using 2 μg of input RNA using TruSeq RNA Sample Preparation Kit v2 (Illumina RS-122-2001) according to the manufacturer's protocols. Sequencing data were deposited into Gene Expression Omnibus under accession no. GSE54677.

References

Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010;11(3):204–220.

Jackson JP, Lindroth AM, Cao X, Jacobsen SE. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature. 2002;416(6880):556–560.

Ronemus MJ, Galbiati M, Ticknor C, Chen J, Dellaporta SL. Demethylation-induced developmental pleiotropy in Arabidopsis. Science. 1996;273(5275):654–657.

Stroud H, et al., Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol. 2014;21(1):64–72.

Zemach A, et al., The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell. 2013;153(1):193–205.

Cao X, Jacobsen SE. Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. Curr Biol. 2002;12(13):1138–1144.

Chan SW, et al., RNA silencing genes control de novo DNA methylation. Science. 2004;303(5662):1336.

Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. RNA polymerase IV directs silencing of endogenous DNA. Science. 2005;308(5718):118–120.

Law JA, et al., Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature. 2013;498(7454):385–389.

Onodera Y, et al., Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell. 2005;120(5):613–622.

Johnson LM, et al., SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature. 2014;507(7490):124–128.

Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS ONE. 2008;3(9):e3156.

Cokus SJ, et al., Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature. 2008;452(7184):215–219.

Du J, et al., Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell. 2012;151(1):167–180.

Johnson L, Cao X, Jacobsen S. Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. Curr Biol. 2002;12(16):1360–1367.

Tariq M, et al., Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. Proc Natl Acad Sci USA. 2003;100(15):8823–8827.

Amedeo P, Habu Y, Afsar K, Mittelsten Scheid O, Paszkowski J. Disruption of the plant gene MOM releases transcriptional silencing of methylated genes. Nature. 2000;405(6783):203–206.

Yokthongwattana C, et al., MOM1 and Pol-IV/V interactions regulate the intensity and specificity of transcriptional gene silencing. EMBO Journal. 2010;29:340–351.

Habu Y, et al., Epigenetic regulation of transcription in intermediate heterochromatin. EMBO Rep. 2006;7(12):1279–1284.

Probst AV, Fransz PF, Paszkowski J, Mittelsten Scheid O. Two means of transcriptional reactivation within heterochromatin. Plant J. 2003;33(4):743–749.

Vaillant I, Schubert I, Tourmente S, Mathieu O. MOM1 mediates DNA-methylation-independent silencing of repetitive sequences in Arabidopsis. EMBO Rep. 2006;7(12):1273–1278.

Caikovski M, et al., Divergent evolution of CHD3 proteins resulted in MOM1 refining epigenetic control in vascular plants. PLoS Genet. 2008;4(8):e1000165.

Mittelsten Scheid O, Probst AV, Afsar K, Paszkowski J. Two regulatory levels of transcriptional gene silencing in Arabidopsis. Proc Natl Acad Sci USA. 2002;99(21):13659–13662.
Brabbs TR, et al., The stochastic silencing phenotype of Arabidopsis morc6 mutants reveals a role in efficient RNA-directed DNA methylation. Plant J. 2013;75(5):836–846.

Lorković ZJ, Naumann U, Matzke AJ, Matzke M. Involvement of a GHKL ATPase in RNA-directed DNA methylation in Arabidopsis thaliana. Curr Biol. 2012;22(10):933–938.

Moissiard G, et al., MORC family ATPases required for heterochromatin condensation and gene silencing. Science. 2012;336(6087):1448–1451.

Inoue N, et al., New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. Hum Mol Genet. 1999;8(7):1201–1207.

Watson ML, et al., 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. 1998;95(24):14361–14366.

Langen G, et al., The compromised recognition of turnip crinkle virus1 subfamily of microrchidia ATPases regulates disease resistance in barley to biotrophic and necrotrophic pathogens. Plant Physiol. 2014;164(2):866–878.

Kang HG, Klessig DF. The involvement of the Arabidopsis CRT1 ATPase family in disease resistance protein-mediated signaling. Plant Signal Behav. 2008;3(9):689–690.

Kang HG, Kuhl JC, Kachroo P, Klessig DF. CRT1, an Arabidopsis ATPase that interacts with diverse resistance proteins and modulates disease resistance to turnip crinkle virus. Cell Host Microbe. 2008;3(1):48–57.

Kang HG, et al., Endosome-associated CRT1 functions early in resistance gene-mediated defense signaling in Arabidopsis and tobacco. Plant Cell. 2010;22(3):918–936.

Kang HG, et al., CRT1 is a nuclear-translocated MORC endonuclease that participates in multiple levels of plant immunity. Nat Commun. 2012;3:1297.

Dutta R, Inouye M. GHKL, an emergent ATPase/kinase superfamily. Trends Biochem Sci. 2000;25(1):24–28.

Liu ZW, et al., The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. PLoS Genet. 2014;10(1):e1003948.

Kanno T, et al., A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. Nat Genet. 2008;40(5):670–675.

Law JA, et al., A protein complex required for polymerase V transcripts and RNA-directed DNA methylation in Arabidopsis. Curr Biol. 2010;20(10):951–956.

Zhong X, et al., DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. Nat Struct Mol Biol. 2012;19(9):870–875.

Johnson LM, Law JA, Khattar A, Henderson IR, Jacobsen SE. SRA-domain proteins required for DRM2-mediated de novo DNA methylation. PLoS Genet. 2008;4(11):e1000280.

Lorković ZJ. MORC proteins and epigenetic regulation. Plant Signal Behav. 2012;7(12):1561–1565.

Carbon S, et al., AmiGO Hub Web Presence Working Group AmiGO: Online access to ontology and annotation data. Bioinformatics. 2009;25(2):288–289.

Boller T. Chemoperception of microbial signals in plant-cells. Annu Rev Plant Physiol. 1995;46:189–214.

Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell. 2013;152(1-2):352–364.

Wood AJ, Severson AF, Meyer BJ. Condensin and cohesin complexity: The expanding repertoire of functions. Nat Rev Genet. 2010;11(6):391–404.

Chapter 4: Mouse MORC3 is a GHKL ATPase that localizes to H3K4me3 marked chromatin

Abstract

Microrchidia (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.

Introduction

The Microrchidia (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 (Watson et al., 1998). This arrest was later shown to be associated with transposon derepression, implicating murine MORC1 as a crucial mediator of transposon silencing (Pastor et al., 2014). *Arabidopsis thaliana* MORC1 and MORC6 were shown to mediate silencing of transposons in a manner largely independent of changes in DNA methylation (Lorkovic et al., 2012, Moissiard et al., 2012, Brabbs et al., 2013). Studies in *Caenorhabditis elegans,* which lack DNA methylation,

also concluded that the single MORC gene in this organism plays a role in transgene silencing (Moissiard et al., 2012). 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 (Moissiard et al., 2012, Mimura et al., 2010, Moissiard et al., 2014 [See Chapter 3]). The coiled-coil region has been proposed to promote constitutive dimerization, whereas N-terminal ATPase head dimerization occurs only on ATP binding (Mimura et al., 2010). This is consistent with other GHKL ATPases described in the literature, many of which have been reported to undergo ATP-dependent dimerization (Dutta and Inouye 2000, Corbett and Berger 2003, Corbett and Berger 2005). 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 (Mimura et al., 2010).

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) (He et al., 2010, Li et al., 2012, Eberl et al., 2013). Consistent with this hypothesis, human and murine MORC3 have been identified in mass spectrometry screens as H3K4me3 readers (Engelen et al., 2015, Hoppmann et al., 2011, Liu et al., 2016). 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 (Mimura et al., 2010).

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 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 (**Figure 4-1*A***). The CW-type zinc finger has been identified as an H3K4me reader module as assayed by structural and biochemical methods (He et al., 2010, Li et al., 2012, Eberl et al., 2013, Hoppmann et al., 2011). 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 (**Figure 4-1*A***) was refined to 2.6-Å resolution, yielding an *R* factor of 21.8% and a free *R* factor of 23.0% (**Table 4-1**). 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 (**Figure 4-1*A***). 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) (Liu et al., 2016). 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) (Ali et al., 2006). 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 Å2 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 (Figure 4-1B).

Notably, Ile9 and Leu14 of one monomer extended their side chains into two small hydrophobic pockets located on the other monomer (**Figure 4-1*B***). 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 (**Figure 4-1*C***). 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 (**Figure 4-1*D***). 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 (**Figure 4-1*E*** and **Figure 4-2**).

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-Mg2+. The ATPase domains formed a symmetrical dimer in the crystal structure, consistent with the observation that GHKL ATPases dimerize on ATP binding (Corbett and Berger 2005, Iyer et al., 2008) 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 (Figure 4-3 and Table 4-2). 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 (Figure 4-1A) (Hoppmann et al., 2011). 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 (Figure 4-3A). The MORC3 (7-456) holoenzyme showed a ratio of roughly 70% monomer/30% dimer without added ligand (Figure 4-3 and Table 4-2). Incubation of substrate (ATP-Mg2+) resulted in a small shift in the equilibrium ratio to 59% monomer/41% dimer, similar to that of the ADP-Mg2+ condition, where we observed a ratio of 64% monomer/36% dimer. Presumably, the rapid hydrolysis of ATP resulted in ADP-Mg2+ 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 (Figure 4-4A).

Furthermore, protein cross-linking experiments showed that the population of dimers increased on incubation with ATP or AMPPNP compared with the protein without nucleotide (Figure 4-4B). 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 (Figure 4-3 and Table 4-2). 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-Mg2+ could be bound to the monomer and that both one and two molar equivalents were bound to the dimer (Table 4-3). With the nonhydrolyzable analogs, we observed two molar equivalents of each analog binding to the dimer form (Figure 4-5). With AMPPNP, we observed the expected one molar equivalent of AMPPNP-Mg2+ bound per monomer, but for ATP- γ-S-Mg2+, we did not detect a monomer-ligand complex. The observed dimer bound to two molar equivalents of AMPPNP-Mg2+ 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 (Figure 4-1B). A similar isoleucine residue was previously found to significantly contribute to dimer stabilization in other GHKLs (Corbett and Berger 2005). 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 (Figure 4-3B, Figs.4-4B and 4-6, and Tables 4-4 and 4-5).

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 (Figure 4-1 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, 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 (Figure 4-4B).

The I9A monomer and dimer were found in complex with one and two molar equivalents of nucleotide, respectively (Fig. 4-6 and Table 4-5), 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 also suggest that the wild-type and I9A proteins have similar overall structures (Figure 4-3*A* and Fig. 4-6). 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. 4-7*A*). The ATPase domain formed minimal contacts (as small as 102 $Å^2$ 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 (Krissinel and Henrick 2007). The peptide was bound within a negatively charged surface of the CW domain (Figure 4-7B 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 (Figure 4-8A), there were additional side chain interactions involved in the specific recognition of the bound H3K4me3 peptide (Figure 4-7B 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. 4-8D). 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. 4-8E), similar to other classical methyl-lysine recognition modules (Patel and Wang 2013). 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 (Hoppmann et al., 2011, Adams et al., 2010), formed extensive hydrogen bonds and salt bridge interactions with the CW domain (Fig. 4-8F). 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. 4-8F).

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% (Table 4-1). 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. 4-8 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 (Figure 4-9B). The MORC3 CW domain exhibited a significant preference for H3K4me2 and H3K4me3 over H3K4me0, consistent with previous reports of human MORC3 binding preferences (Li et al., 2012, Eberl et al., 2013).

We used isothermal titration calorimetry (ITC) to calculate the binding affinity of the MORC3 CW domain for both unmodified H3K4me0 and H3K4me3 H3 peptides (Figure 4-8C). We used H3K9me3 peptide as a negative control, because it appeared to be anticorrelated with MORC3 in the genome (Engelen et al., 2015). The MORC3 CW domain bound to the H3K4me3 peptide

with a high affinity of 0.49 μM (Figure 4-8C). 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) (Figure 4-8C), 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 (Figure 4-10*A*). 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 (Figure 4-10B). We called 14,032 MORC3 peaks, of which 11,071 were located in promoters (Fig. 4-9A 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 (Figure 4-9C). In contrast, we observed a strong anticorrelation between MORC3 signals and silencing features, such as H3K9 trimethylation (H3K9me3) (Figure 4-9C). 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 (Figure 4-9C).

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 (Figure 4-9D). Consistent with the heat maps (Figure 4-9C), 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 (Calo and Wysocka 2013), were found together with MORC3 at promoters (Figure 4-9E). 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 (Figure 4-9E) (Heintzman et al., 2007). 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 (Iyer et al., 2008). 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 (Mimura et al., 2010). 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 (Novilos and Sherratt 2014, Hirano 2016).

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 (Li et al., 2012, Engelen et al., 2015, Hoppmann et al., 2011, Liu et al., 2016), 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 (Mimura et al., 2010). 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.

Figure legends:

Figure 4-1. 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.

Figure 4-2: Schematic representation of specific recognition of AMPPNP by the ATPase domain of MORC3.

Intermolecular hydrogen-bonding interactions between AMPPNP and the ATPase domain in the complex are shown. Hydrogen bonds are indicated by dashed lines, together with water molecules and Mg2+ cations.

Figure 4-3: 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.

Figure 4-4: MORC3 (7-456) is an active ATPase that productively forms nucleotide-dependent dimers.

(A) Hydrolysis assays were performed at 25 °C for 60 min in 50 mM ammonium acetate buffer, pH 7.5. Images were quantified using ImageQuant TL (GE Healthcare). Percent hydrolysis was calculated as (ADP/ADP + ATP) × 100. (B) Cross-linking assays demonstrating that nucleotide binding is sufficient for dimerization. Dimerization is compromised when a key dimerization residue, I9, is mutated to an alanine. The assay was done with a final protein concentration of 1.36 μM.

Figure 4-5: Native mass spectra of MORC3 (7-456) with nonhydrolyzable analogs.

Expanded views of the dimer charge envelopes of MORC3 with AMPPNP (A) and ATP-γ-S (B). Shown are the +20 to +24 charge states of MORC3 dimer (red) and resulting splitting due to one (gray) and two (green) molar equivalents of the respective nonhydrolyzable analogs bound to the dimer.

Figure 4-6: Native mass spectra of the MORC3 (7-456) I9A mutant.

The charge envelope on the left corresponds to the MORC3 (7-456) I9A monomer at +16, +15, +14, and +13 ions and the MORC3 dimer at +24, +23, +22, +21, and +20 ions. Black arrows

correspond to the peak splitting observed when the protein or protein complex is bound with one nucleotide; blue represents the protein complex bound with two nucleotides.

Figure 4-7: 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).

Figure 4-8: MORC3 binds H3K4me3.

(A) The H3K4me3 peptide forms a β-strand–like interaction with the CW of MORC3. (Upper) The intermolecular hydrogen bonds between the two β-strands are highlighted. (Lower) Schematic of side-chain interactions between the H3K4me3 peptide and the CW domain of

MORC3 in the complex. (B) Quantification of GST or GST-MORC3 CW domain retained by H3 (1–20), H3(1–20)K4me1, H3(1–20)K4me2, and H3(1–20)K4me3 peptides. (C) ITC binding curves between the MORC3 CW domain and different H3 peptides, including H3(1–15), H3(1–15)K4me3, and H3(5–19)K9me3, indicating that MORC3 CW preferentially binds H3K4me3.

Figure 4-9: 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. Log2 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.

Figure 4-10: MORC3 peaks co-localize with H3K4me3.

(A) MORC3 antibody validation. Western blots of wild-type and morc3 knockout lysates. The anti-MORC3 antibody is represented in the green channel, and anti-actin antibody (loading control) is represented in the red channel. A band corresponding to MORC3 is absent in the knockout lysate. (B) Representative browser track screenshot of a MORC3 peak and an H3K4me3 peak. Both tracks are normalized to their respective input.

Figure 4-1: Structure of MORC3 ATPase-CW cassette in complex with AMP-PNP and H3K4me3 peptide

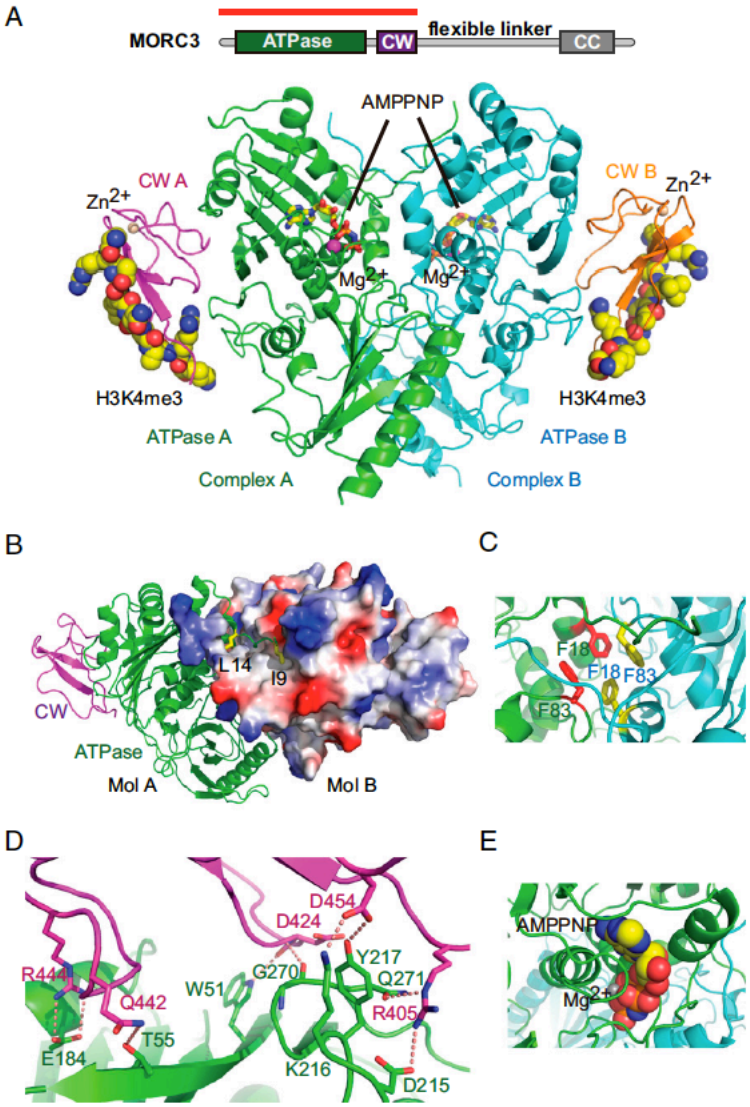


Figure 4-2: Schematic representation of specific recognition of AMPPNP by the ATPase domain of MORC3.

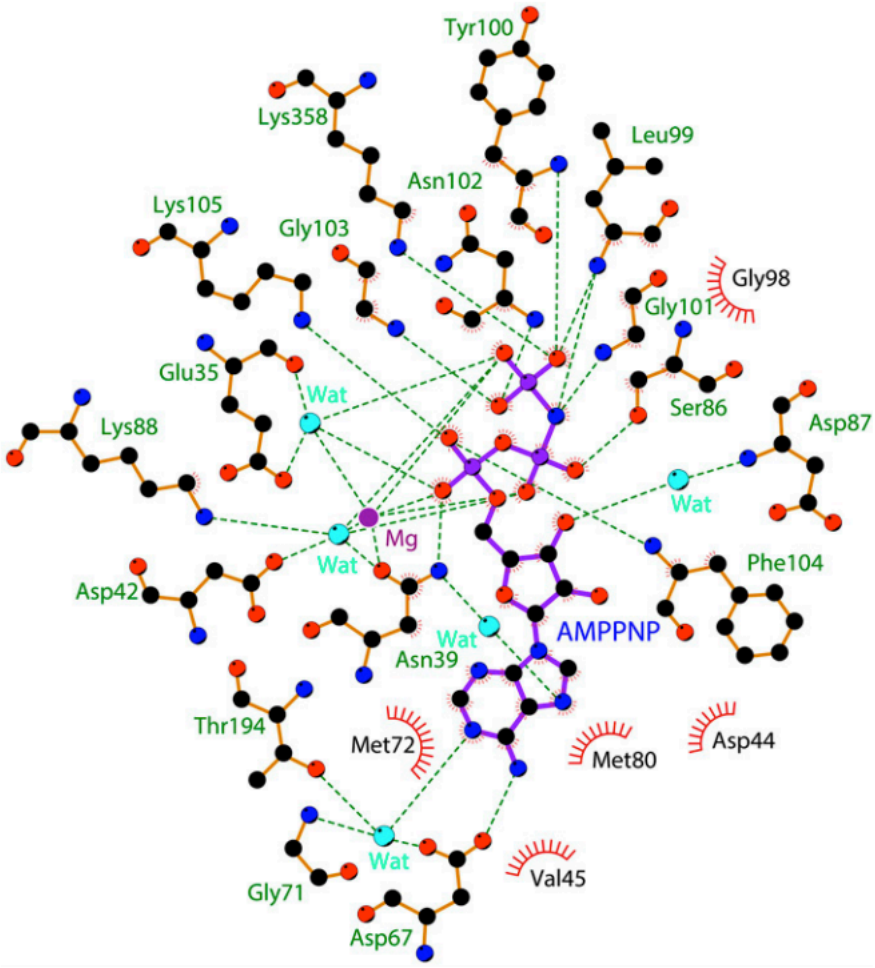


Figure 4-3: The N terminus of MORC3 forms ATP-dependent dimers.

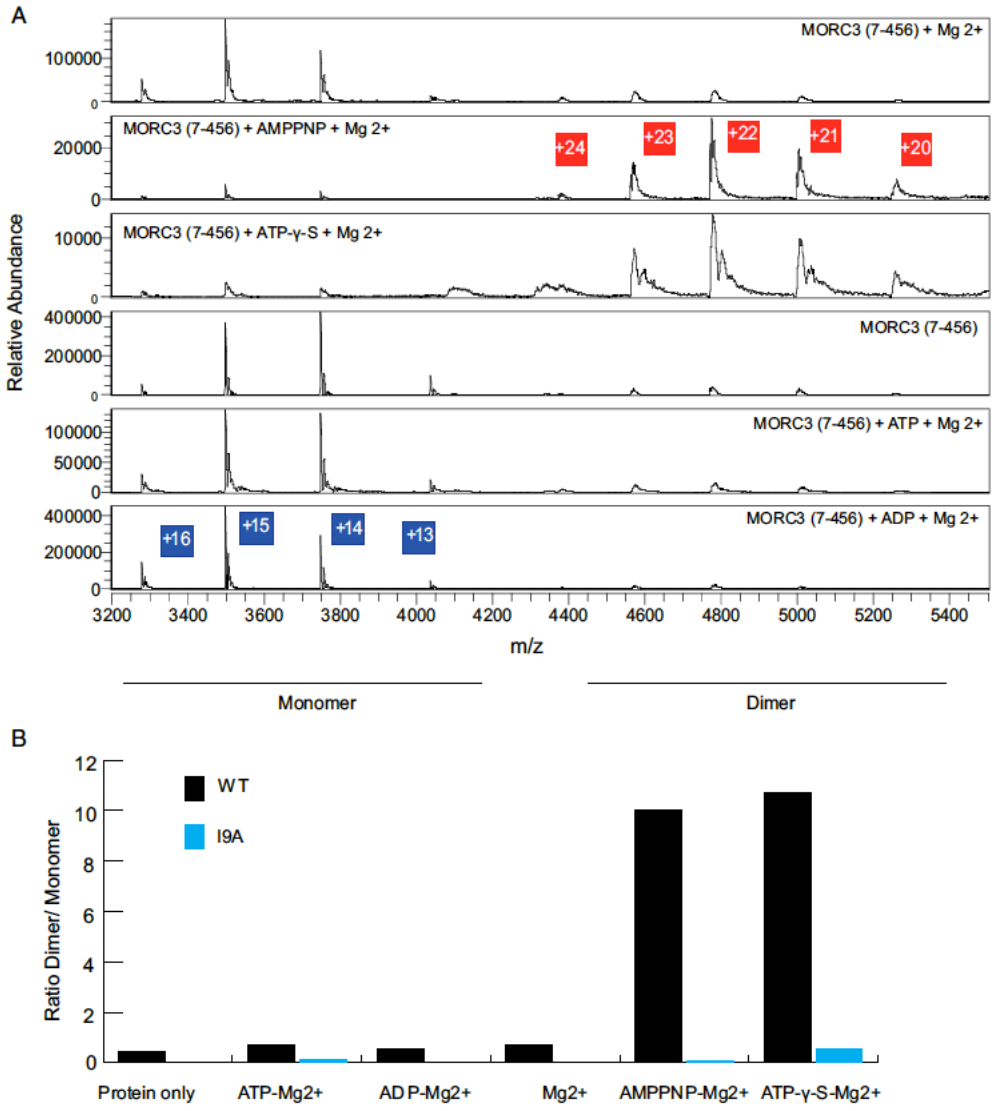


Figure 4-4: MORC3 (7-456) is an active ATPase that productively forms nucleotide-dependent dimers.

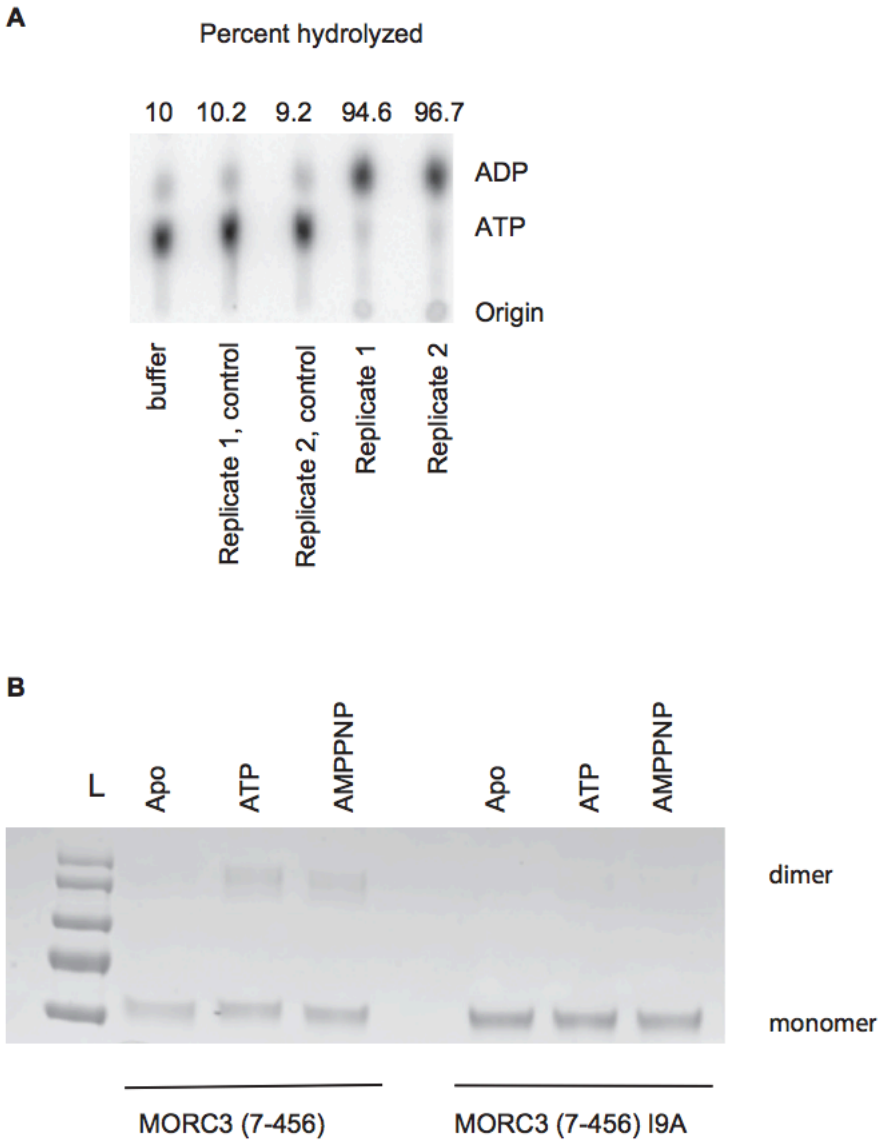


Figure 4-5: Native mass spectra of MORC3 (7-456) with non-hydrolyzable analogs.

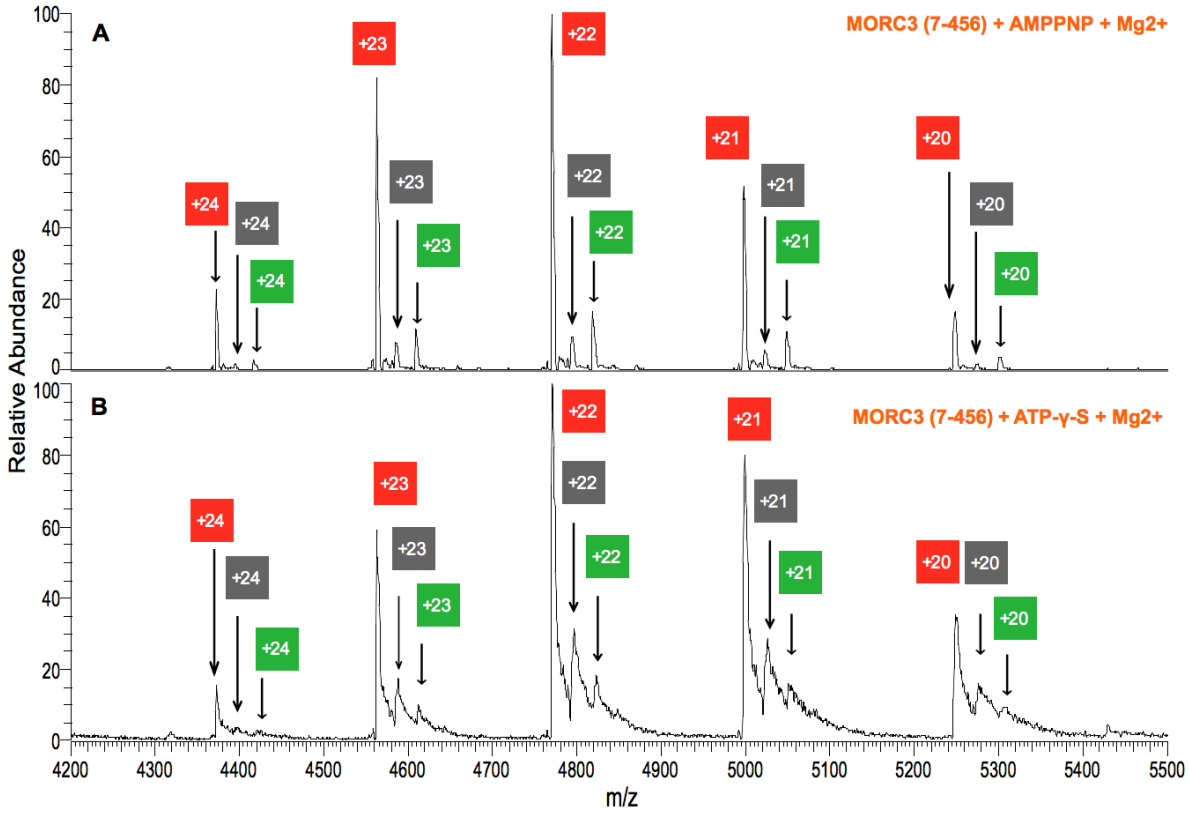


Figure 4-6: Native mass spectra of the MORC3 (7-456) I9A mutant.

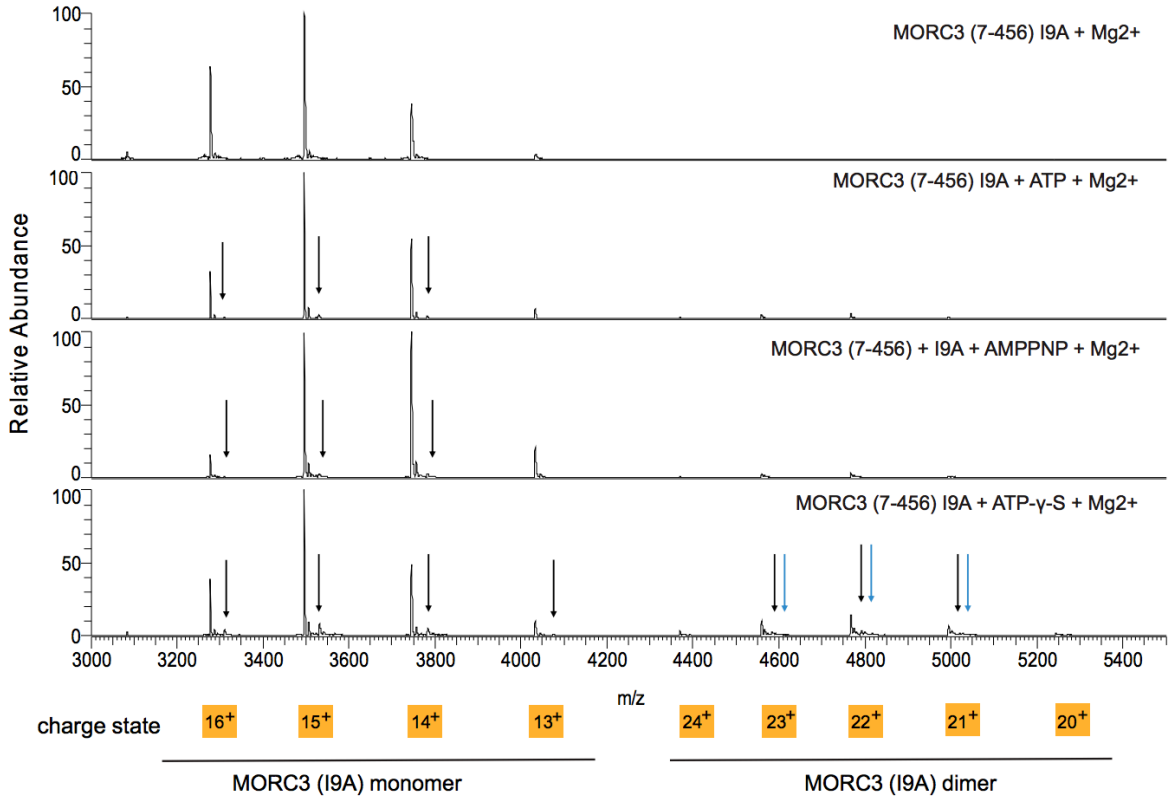


Figure 4-7: Intermolecular interactions between MORC3 and the H3K4me3 peptide

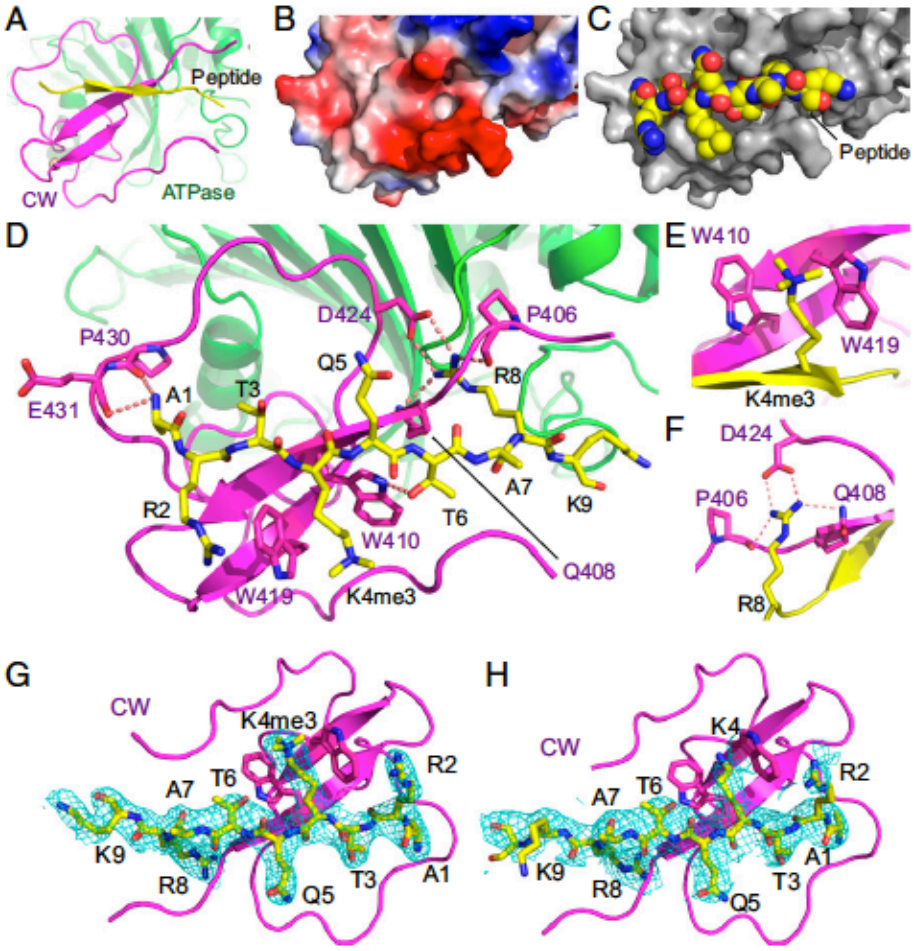


Figure 4-8: MORC3 binds H3K4me3.

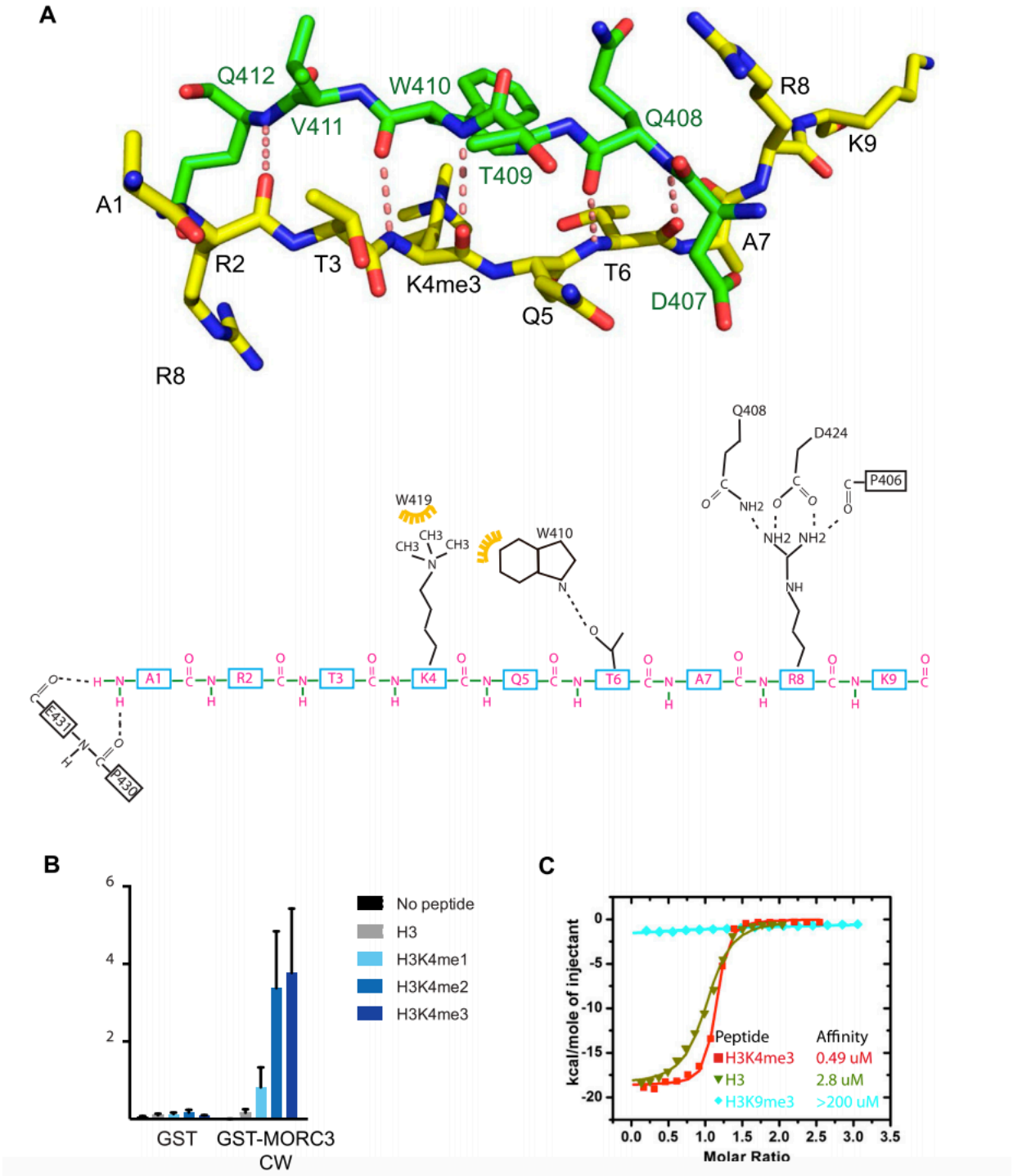


Figure 4-9: MORC3 is localized to active gene promoters *in vivo*

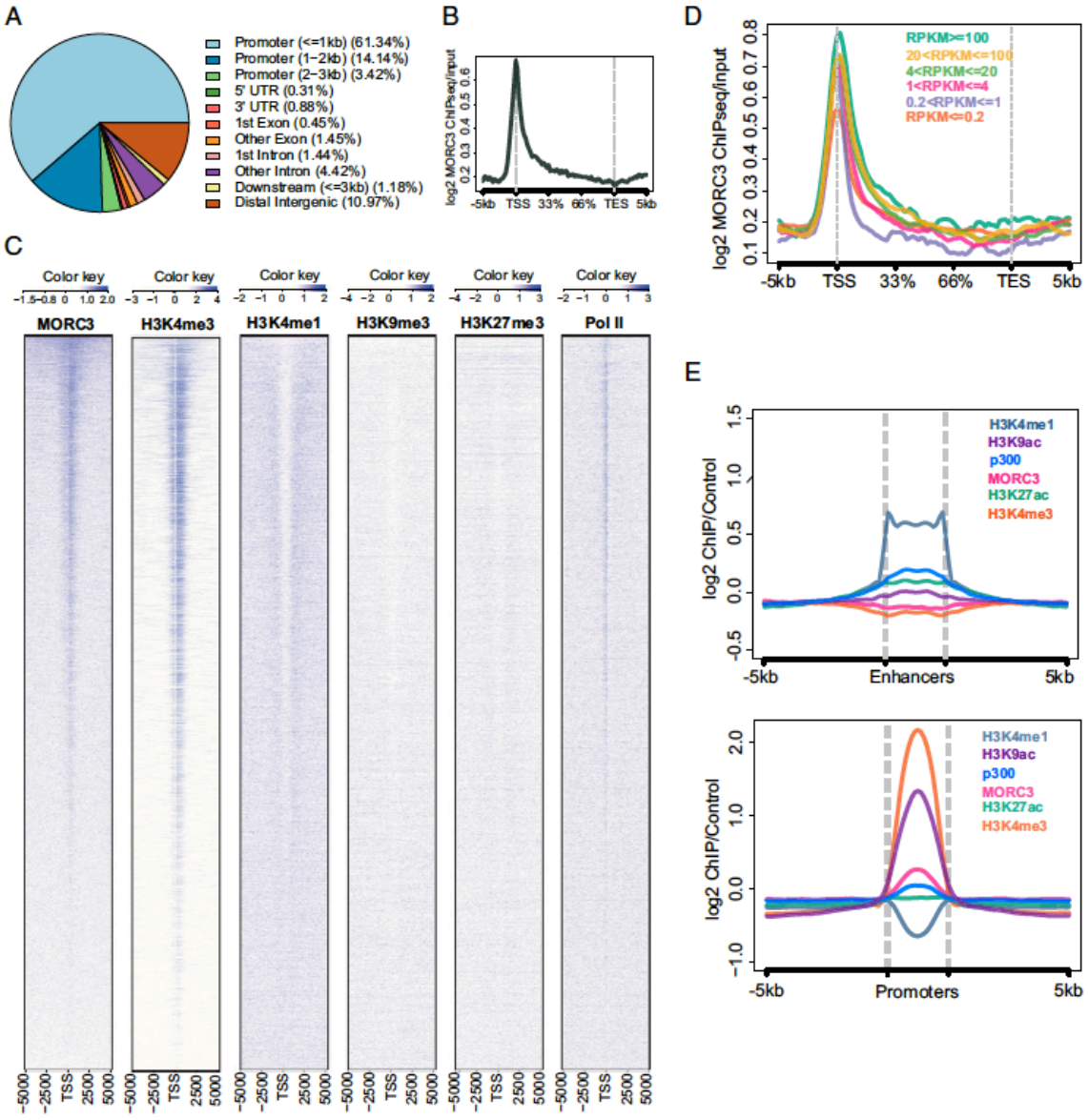


Figure 4-10: MORC3 peaks colocalize with H3K4me3.

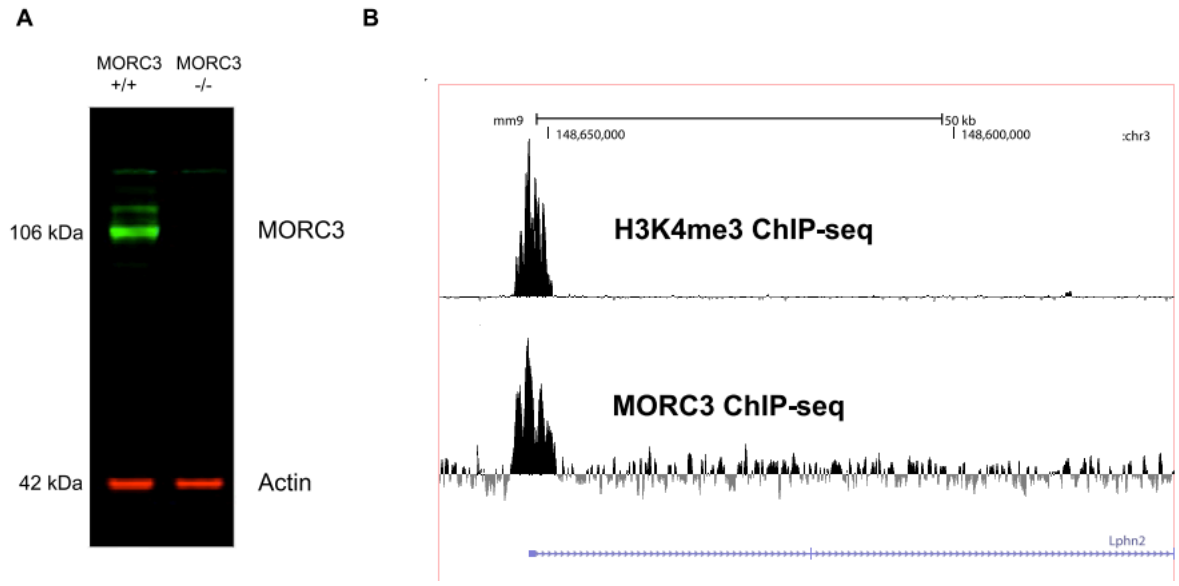


Table 4-1:

Crystal	Hg derivate	MORC3+AMPPNP+H3(1–15)K4me3	MORC3+AMPPNP+H3(1–32)
Summary of diffraction data			
Beamline	SSRF-BL17U1	SSRF-BL17U1	SSRF-BL19U
PDB ID code		5IX1	5IX2
Wavelength, Å	0.9928	0.9792	0.9792
Space group	*I*222	*I*222	*I*222
Cell parameters, Å			
a	110.9	111.4	112.2
b	151.6	149.1	147.5
c	174.1	173.5	173.6
Resolution, Å	50.0–2.9 (3.00–2.90)*	20.0–2.6 (2.74–2.60)	20.0–2.9 (3.06–2.90)
R_{merge}, %	13.4 (70.1)	8.5 (86.6)	9.1 (70.1)
Observed reflections	479,105	205,990	206,638
Unique reflections	33,002	44,534	32,131
Average I/σ(I)	34.6 (4.2)	11.8 (2.0)	12.5 (2.6)
Completeness, %	99.9 (100.0)	99.7 (99.9)	99.6 (99.8)
Redundancy	14.5 (14.9)	6.4 (5.9)	6.3 (6.9)
Refinement and structure model			
R/Free *R* factor, %		21.8/23.0	22.0/23.6
Non-H atoms		7,042	7,049
Protein/peptide		6,751/158	6,825/158
AMPPNP/water		62/67	62/-
Zn^{2+}/Mg^{2+}		2/2	2/2
B factor, $Å^2$		75.7	80.2
Protein/peptide		75.4/93.0	79.9/101.0
AMPPNP/water		65.5/72.7	62.9/-
Zn^{2+}/Mg^{2+}		87.9/59.1	88.5/58.9
Rmsd			
Bond lengths, Å		0.020	0.022
Bond angles, °		1.790	1.871

SSRF, Shanghai Synchrotron Radiation Facility.
*Values in parentheses are for the highest shell.

Table 4-2:

Sample	Monomer mass, Da (SD)	Dimer mass, Da (SD)	Ratio monomer/dimer (% error)
MORC3 (7-456)	52,517 (±10)	105,148 (±122)	69.5/29.5 (±0.7)
MORC3 (7-456) + ATP-Mg^{2+}	52,514 (±31)	105,208 (±13)	59/41 (±1.4)
MORC3 (7-456) + ADP-Mg^{2+}	52,466 (±2)	105,166 (±25)	63.5/36.5 (±5)
MORC3 (7-456) + Mg^{2+}	52,496 (±23)	105,199 (±20)	58/42 (±5.7)
MORC3 (7-456) + AMPPNP-Mg^{2+}	52,464 (±11)	105,100 (±22)	9/91 (±1.4)
MORC3 (7-456) + ATP-γ-S-Mg^{2+}	52,500 (±8)	105,134 (±42)	8.5/91.5 (±3.5)

The predicted mass of a wild-type monomer with one bound zinc is 52,461 Da, and that of a dimer with no nucleotides and with two zinc is 104,922 Da. Two samples were analyzed.

Table 4-3:

Sample	Monomer-ligand 1 mass, Da	Dimer-ligand 1 mass, Da	Dimer-ligand 2 mass, Da
MORC3 (7-456) + ATP-Mg^{2+}	52,902	104,982	105,592
MORC3 (7-456) + Mg^{2+}	ND	ND	ND
MORC3 (7-456) + AMPPNP-Mg^{2+}	52,910	105,030	105,546
MORC3 (7-456) + ATP-γ-S-Mg^{2+}	ND	105,036	105,615

Predicted mass of a WT monomer with one bound zinc, one magnesium, and one ATP is 52,993 Da; of a dimer with one zinc, one magnesium, and one ATP is 105,454 Da; of a dimer with two zinc, two magnesium, and two ATPs is 105,985 Da; of a monomer with one bound zinc, one magnesium, and one AMPPNP is 52,991 Da; of a dimer with two zinc, one magnesium, and one AMPPNP is 105,453 Da; of a dimer with two zinc, two magnesium, and two AMPPNPs is 105,984 Da; of a monomer with one bound zinc and one ATP-γ-S is 53,009 Da; of a dimer with two zinc, one magnesium, and one ATP-γ-S is 105,470 Da; and of a dimer with two zinc, two magnesium, and two ATP-γ-S is 106,018 Da. Two samples were analyzed. ND, not detected.

Table 4-4:

Sample	Monomer mass, Da	Dimer mass, Da	Ratio monomer/dimer, %
MORC3 (7-456) I9A	52,485 ± 49	ND	100/ND
MORC3 (7-456) I9A + ATP-Mg^{2+}	52,435 ± 10	104,852 ± 13	84/16
MORC3 (7-456) I9A + ADP-Mg^{2+}	52,416 ± 114	ND	100/ND
MORC3 (7-456) I9A + Mg^{2+}	52,485 ± 16	104,866 ± 10	97/3
MORC3 (7-456) I9A + AMPPNP-Mg^{2+}	52,427 ± 45	104,877 ± 26	96/4
MORC3 (7-456) I9A + ATP-γ-S-Mg^{2+}	52,657 ± 352	105,016 ± 152	65/35

Predicted mass of a mutant monomer with one bound zinc is 52,419 Da; and of a mutant dimer with no nucleotides and with two zinc is 104,838 Da. Two samples were analyzed. ND, not detected.

Table 4-5:

Sample	Monomer-ligand 1 mass, Da	Dimer-ligand 1 mass, Da	Dimer-ligand 2 mass, Da
MORC3 (7-456) I9A + ATP-Mg^{2+}	52,867	ND	ND
MORC3 (7-456) I9A + Mg^{2+}	ND	ND	ND
MORC3 (7-456) I9A + AMPPNP-Mg^{2+}	52,913	ND	ND
MORC3 (7-456) I9A + ATP-γ-S-Mg^{2+}	52,906	105,123	105,491

Predicted mass of a MORC3 (7-456) I9A monomer with one bound zinc, one magnesium, and one ATP is 52,951 Da; of a MORC3 (7-456) I9A monomer with one bound zinc, one magnesium, and one AMPPNP is 52,950 Da; of a dimer with two zinc, one magnesium, and one AMPPNP is 105,369 Da; of a dimer with two zinc, two magnesium, and two AMPPNPs is dimer with two zinc, one magnesium, and one ATP-γ-S is 105,386 Da; and of a dimer with two zinc, two magnesium, and two ATP-γ-S is 105,934 Da. Two samples were analyzed. ND, not detected.

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 cyroprotected 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 Otwinowski and Minor 1997). The mercury derivate single-wavelength anomalous diffraction (SAD) data were used for phasing using Phenix (Adams et al., 2010). The model was built using the program Coot (Emsley et al., 2010) and refined using Phenix (Adams et al., 2010). All of the molecular graphics were generated with PyMol (DeLano Scientific) and LigPlus (Laskowski and Swindells 2011). Diffraction data and structure refinement statistics are summarized in Table 4-1.

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 $1e^6$. 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 (Zhang and Marshall 1998) and PeakSeeker (Lu et al., 2015) 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 (Langmead et al., 2009) by allowing up to two mismatches and keeping only uniquely mapped reads. PCR duplicated reads were removed with SAMtools (Li et al., 2009). Heatmaps and metaplots were generated in R with the ngs.plot package (Shen et al., 2014). MORC3 ChIPseq peaks were identified with MACS version 2.1.0 (Zhang et al., 2008). Defined MORC3 peaks were then annotated to the mm9 genome with the ChIPseeker R package (Yu et al., 2015). 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 (Heintzman et al., 2007, Bernstein et al., 2005, Shen et al., 2012). Enhancer lists were identified as done previously (Heintzman et al., 2007, Shen et al., 2012, Heintzman et al., 2009) 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 (Anders et al., 2015) with default parameters. Expression levels were determined by reads per kilobase of exons per million aligned reads.

SI Experimental Procedures

Cross-Linking Experiments.

Protein at a final concentration of 1.4 μM was cross-linked in 25 mM Hepes pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT, and 1 mM nucleotide. BS3 cross-linker (Life Technologies) was resuspended to a final concentration of 20 mM and then used at 0.5 mM final concentration. Cross-linking was allowed to proceed for 5 min. Cross-linked samples were quenched in a final concentration of 200 mM Tris pH 8.0, and then separated by SDS/PAGE and stained with Coomassie blue.

ATPase Assay.

Protein was incubated with a mixture of cold ATP and trace amounts of ^{32}P-α-ATP in 150 mM AmAc pH 7.5 and 2 mM $MgCl_2$. The reaction was quenched with 50 mM Tris pH 8.0, 0.1% SDS, and 5 mM EDTA, followed by spotteding on PEI-F plates (J.T. Baker). Heat-killed enzyme

was used in control reactions. Plates were run in 0.5 M LiCl/1 M formic acid, dried, and exposed to a phosphor screen. The images were scanned and quantified by ImageQuant TL (GE Healthcare).

Peptide Pull-Down Assays.

Peptide pull-down assays were performed as described previously (Hung et al., 2009), but with 20 μL of streptavidin magnetic resin (65001; Life Technologies). Biotinylated histone H3(1–20) K4 monomethylated, dimethylated, trimethylated, and unmodified peptides were purchased from EpiCypher (12-0007, 12-0008, 12-0009, and 12-0001). Samples were rotated for 1 h at 4 °C. The beads were washed three times with 1 mL of cold binding buffer, resuspended in 60 μL of Laemmli buffer, and then boiled for 10 min. Protein was detected using anti-GST (sc-459; Santa Cruz Biotechnology). The percent pull-down was calculated by dividing the pull-down signal by the input signal and adjusting for the difference in percentage volume loaded.

ITC.

The ITC experiments were conducted using a MicroCal calorimeter ITC 200 instrument. The MORC3 CW domain protein was dialyzed against a buffer of 100 mM NaCl, 20 mM Hepes pH 7.0, and 2 mM β-mercaptoethanol. The peptide was dissolved into the same buffer. All of the titrations were performed at 20 °C. The data were analyzed using Origin 7.0.

Generation of MORC3-Deficient mESCs.

To target MORC3, the genomic DNA sequence GCAACTGCACTGAACGGCC, corresponding to bases 101–83 of the MORC3 coding sequence, was cloned into the pX330 vector (Cong et al., 2013). This construct and a pMaxGFP transient GFP expression vector (Lonza) were then cotransfected into V6.5 cells in suspension using Lipofectamine 2000 (Life Technologies). After

48 h in culture, GFP^+ cells were sorted by FACS, plated as single colonies on a 96-well plate, and expanded for analysis. Control lines were generated from cells transfected only with a GFP expression vector. Targeted lines were screened using an anti-MORC3 antibody generated in collaboration with Rockland Immunochemicals. Lines that showed a loss of MORC3 were screened by amplifying and sequencing the targeted region. Lines that showed frameshift mutations of MORC3 on both alleles were then subcloned by FACS sorting to ensure a homogenous genotype.

- Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 5IX1 and 5IX2). The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE78258).

References

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.

Pastor WA, et al., (2014) MORC1 represses transposable elements in the mouse male germline. Nat Commun 5:5795.

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.

Moissiard G, et al., (2012) MORC family ATPases required for heterochromatin condensation and gene silencing. Science 336(6087):1448–1451.

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.

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.

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.

Dutta R, Inouye M (2000) GHKL, an emergent ATPase/kinase superfamily. Trends Biochem Sci 25(1):24–28.

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.

Corbett KD, Berger JM (2005) Structural dissection of ATP turnover in the prototypical GHL ATPase TopoVI. Structure 13(6):873–882.

He F, et al., (2010) Structural insight into the zinc finger CW domain as a histone modification reader. Structure 18(9):1127–1139.

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.

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.

Engelen E, et al., (2015) Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. Nat Commun 6:7155.

Hoppmann V, et al., (2011) The CW domain, a new histone recognition module in chromatin proteins. EMBO J 30(10):1939–1952.

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.

Ali MM, et al., (2006) Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. Nature 440(7087):1013–1017.

Iyer LM, Abhiman S, Aravind L (2008) MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. Biol Direct 3:8.

Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372(3):774–797.

Patel DJ, Wang Z (2013) Readout of epigenetic modifications. Annu Rev Biochem 82:81–118.

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.

Calo E, Wysocka J (2013) Modification of enhancer chromatin: What, how, and why? Mol Cell 49(5):825–837.

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.

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.

Hirano T (2016) Condensin-based chromosome organization from bacteria to vertebrates. Cell 164(5):847–857.

Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276:307–326.

Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66(Pt 4):486–501.

Laskowski RA, Swindells MB (2011) LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model 51(10):2778–2786.

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.

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.

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.

Li H, et al., 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079.

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.

Zhang Y, et al., (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9(9):R137.

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.

Bernstein BE, et al., (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 120(2):169–181.

Shen Y, et al., (2012) A map of the cis-regulatory sequences in the mouse genome. Nature 488(7409):116–120.

Heintzman ND, et al., (2009) Histone modifications at human enhancers reflect global cell type-specific gene expression. Nature 459(7243):108–112.

Anders S, Pyl PT, Huber W (2015) HTSeq: A Python framework to work with high-throughput sequencing data. Bioinformatics 31(2):166–169.

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.

Cong L, et al., (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339(6121):819–823.

Chapter 5: The gene silencing protein MORC-1 topologically entraps DNA and forms multimeric assemblies to cause DNA compaction.

Summary

Microrchidia (MORC) ATPases are critical for gene silencing and chromatin compaction in multiple eukaryotic systems but the mechanisms by which MORC proteins act are poorly understood. Here we apply a series of biochemical, single-molecule and cell-based imaging approaches to better understand the function of the C. elegans MORC-1 protein. We find that MORC-1 binds to DNA in a length-dependent but sequence non-specific manner, and compacts DNA by forming DNA loops. MORC-1 molecules diffuse along DNA, but become static as they grow into foci that are topologically entrapped on DNA. Consistent with the observed MORC-1 multimeric assemblies, MORC-1 forms nuclear puncta in cells, and can also form phase-separated droplets in vitro. We also demonstrate that MORC-1 compacts nucleosome templates. Together these results suggest that MORCs impact genome structure and gene silencing by forming multimeric assemblages to topologically entrap and progressively loop and compact chromatin.

INTRODUCTION

MORC proteins are members of a highly conserved family of GHKL (Gyrase, HSP90, Histidine Kinase, MutL) type ATPases that are found in prokaryotic and eukaryotic organisms (Dong et al., 2018; Iyer et al., 2008). In Arabidopsis thaliana, MORC1 and MORC6 are required for silencing of DNA methylated genes and for large-scale compaction of pericentromeric heterochromatin regions (Moissiard et al., 2012). In mouse, MORC1 is involved in compacting and silencing transposons during male germline development (Pastor et al., 2014). In humans, MORC2 is part of the HUSH complex which is also implicated in transposon silencing (Douse et al., 2018; Tchasovnikarova et al., 2017). In Caenorhabditis elegans (C. elegans), MORC-1 is

required for transgene silencing and for maintaining the transgenerational silencing of siRNA-targeted genes in the germline, and morc-1 mutants show visible genome decondensation (Weiser et al., 2017). Despite these genetic and genomic studies demonstrating a role of MORCs in chromatin compaction processes (Douse et al., 2018; Harris et al., 2016; Moissiard et al., 2014 [See Chapter 3]; Moissiard et al., 2012; Tchasovnikarova et al., 2017; Weiser et al., 2017), the molecular mechanisms by which MORCs act to compact the genome are not understood.

MORCs generally share a similar domain arrangement, possessing N-terminal GHKL ATPase and S5 domains and a large unstructured region followed by a C-terminal coiled coil domain that is thought to be important in homomer or heteromer formation (Douse et al., 2018; Iyer et al., 2008; Li et al., 2016 [See Chapter 4]; Mimura et al., 2010; Takahashi et al., 2007). The ATPase cassette adopts a prototypical GHKL ATPase (Bergerat) fold and dimerizes upon ATP binding, much like other GHKL family members such as bacterial Topoisomerase VIb (Corbett and Berger, 2005; Douse et al., 2018; Dutta and Inouye, 2000; Li et al., 2016 [See Chapter 4]). These properties have led to the proposal that MORCs might act as molecular clamps (Li et al., 2016 [See Chapter 4]; Mimura et al., 2010), however very little is known about how MORCs interact with DNA or with chromatin.

Here we present the first evidence that MORCs bind and compact DNA by progressively trapping DNA loops, resulting in large protein assemblies that are topologically trapped on DNA. Compaction in vitro does not require ATP, although it is stimulated by the addition of ATP and especially by addition of a non-hydrolyzable ATP analog. C. elegans MORC-1 is also found in nuclear bodies, and in vitro will undergo liquid-liquid phase separation. We also demonstrate that MORC-1 can compact DNA that has been assembled into chromatin. Together

these observations provide mechanistic insight into how MORC family members function across eukaryotes to compact genomes and regulate gene silencing.

RESULTS

MORC-1 binds DNA with little sequence preference and prefers longer DNAs over shorter DNAs.

To study C. elegans MORC-1 in an in vitro reconstituted system, we purified bacterially expressed full length MORC-1, containing both the conserved N-terminal GHKL ATPase domain known to dimerize upon ATP binding and the C-terminal coiled coil domain thought to be important in multimerization (Figure 5-1A) (Harris et al., 2016; Inoue et al., 1999; Iyer et al., 2008; Li et al., 2016 [See Chapter 4]; Mimura et al., 2010; Moissiard et al., 2014 [See Chapter 3]; Moissiard et al., 2012; Tchasovnikarova et al., 2017). Native mass spectrometry analysis of the purified protein (Figure 5-1B) revealed it was well ordered and folded, and was predominantly in equilibrium as a monomer or dimer, with minor populations of trimers and tetramers (Figure 5-1C), consistent with the observation that plant and mammalian MORCs exist as homomers or heteromers in vivo in (Harris et al., 2016; Mimura et al., 2010; Moissiard et al., 2014 [See Chapter 3]; Moissiard et al., 2012).

A. thaliana MORC1 was shown to bind DNA (Kang et al., 2012), and we therefore utilized gel shift assays to test the DNA binding activity of full length C. elegans MORC-1. MORC-1 bound radiolabeled 250 base pair (bp) double stranded DNA (Figure 5-2A). Both cold competitor DNA of the same sequence, or a 250 bp DNA with a different sequence competed equally for binding to the labeled probe, suggesting that MORC-1 has little sequence specificity (Figure 5-2B).

However, a 50 bp DNA derived from the original 250 bp sequence was unable to efficiently compete for binding, suggesting that MORC-1 prefers longer DNAs over shorter DNAs (Figure 5-2B). To confirm this observation, we performed gel shift assays with a DNA ladder ranging from 100 to 1500 bp. At lower MORC-1 concentrations (< 400 nM), only longer DNAs (sizes larger than 1000 bp) were shifted, but at higher concentrations (> 800 nM) MORC-1 could also shift shorter DNAs (Figure 5-3). Therefore, MORC-1 is capable of binding both long and short DNAs, but exhibits a preference for longer DNAs under conditions where MORC-1 is limiting. In addition, the observation that MORC-1 could bind ladder DNAs derived from multiple plasmids of varying DNA sequences again suggests MORC-1 likely has little sequence preference, although we cannot rule out that MORC-1 may prefer particular DNA sequences not tested here.

MORC-1 robustly compacts DNA.

C. elegans morc-1 mutants exhibit X chromosome decondensation (Weiser et al., 2017). MORC mutants in other species also exhibit chromosome decompaction phenotypes (Moissiard et al., 2012). To directly test if MORC-1 might be able to compact DNA, we imaged flow-stretched DNAs tethered to a functionalized coverslip via a 3' biotin neutravidin linkage and labeled at the 5' end with a quantum dot (Figure 5-4A) (Graham et al., 2014; Kim and Loparo, 2016). Addition of MORC-1 into the flow cell induced DNA compaction, and the rate of compaction increased linearly with MORC-1 concentration (Figures 5-4B-C; Video S1). Compaction occurred in 150 mM NaCl, but did not occur at NaCl concentrations above 300mM (Figure 5-5A). Importantly, compaction was largely reversible at a high salt concentration (500 mM NaCl) (Figure 5-5B),

suggesting that the observed DNA compaction was not the result of non-specific protein aggregation.

MORCs are ATPases, and we therefore assayed the effect of adding ATP to the compaction assays. While the addition of ATP was not strictly required for compaction, the addition of ATP stimulated the rate of compaction by ~30% (Figures 5-4C and 5-5C-D; File 5-1). Moreover, the non-hydrolysable ATP analog adenylyl-imidodiphosphate (AMP-PNP) further stimulated DNA compaction by approximately two-fold relative to ATP (Figure 5-4D). ATP binding is known to induce GHKL head dimerization (Corbett and Berger, 2005; Douse et al., 2018; Li et al., 2016 [See Chapter 4]), and thus this observation suggests that it is ATP binding, rather than hydrolysis, that enables MORC-1 to most efficiently compact DNA.

MORC-1 uses a loop trapping mechanism to compact DNA.

We used a DNA motion capture assay to analyze the mechanism of MORC-1-mediated DNA compaction by following how segments along the DNA length were condensed (Graham et al., 2014; Kim and Loparo, 2016). λ-DNA was labeled at its five EcoRI binding sites by quantum-dot labeled catalytically inactive EcoRI (EcoRIE111Q) (Figure 5-4E). Compaction initiated from the free end of the DNA and moved sequentially down to the tether point (Figure 5-4E), consistent with either a DNA loop-trapping mechanism as demonstrated for bacterial Spo0J (ParB), or with a loop extrusion mechanism as shown for yeast condensin (Ganji et al., 2018; Graham et al., 2014). Proteins that compact DNA using either loop forming mechanism are sensitive to the force applied to DNA by flow, and therefore preferentially initiate compaction from the free end of the DNA, which experiences substantially lower drag force compared to the DNA segments closer to the tether point (Graham et al., 2014; Kim and Loparo, 2016).

To directly test whether MORC-1 extrudes DNA loops, we tethered both ends of λ-DNA, allowing for slack in the DNA, and then flowed MORC-1 and ATP into the flow cell orthogonal to the direction of DNA tethering. DNA was imaged using the intercalating DNA stain SYTOX Orange. DNA compaction events occurred on both single and double tethered DNA in the same field of view. Singly tethered DNA was compacted down to the tether point, while doubly tethered became taut over time, showing that MORC-1 could compact both singly and doubly tethered DNAs. In contrast to what was previously shown for condensin that acts via a loop extrusion mechanism (Ganji et al., 2018), we did not see evidence for extruded DNA loops on MORC-1 compacted doubly tethered DNA (Figure 5-6; Video S2). The absence of extruded loops implies that MORC-1 uses a loop trapping mechanism rather than a loop extrusion mechanism to condense DNA. This interpretation is also consistent with the observation that MORC-1 can compact DNA without the addition of ATP, since loop extrusion is an active ATP-requiring process driven by directional motor activity (Ganji et al., 2018).

MORC-1 forms discrete foci as it compacts DNA.

To investigate MORC-1 behavior as it compacts DNA, we non-specifically labeled MORC-1 with Cy3, at an average of ~1 Cy3 per MORC-1 dimer. Fluorophore labeling did not perturb the ability of MORC-1 to compact DNA (Figure 5-7A). At concentrations below 2 nM, where the majority of DNA-bound proteins were single MORC-1 dimers, we observed MORC-1 diffusing along DNA (Figure 5-8A). However, DNA compaction was not observed. Addition of higher concentrations of labeled MORC-1 (> 10 nM) to the flow cell resulted in progressive DNA compaction down to the tether point, along with the formation of static MORC-1 foci on the DNA that grew brighter with time (Figures 3A and S5B; Video S3). Additionally, the number of foci per substrate increased as a function of MORC-1 concentration (Figure 5-8C). We

frequently observed collision events between mobile/static MORC-1 proteins, suggesting that the growth of static foci was in part due to the addition of additional DNA bound MORC-1 proteins (Video S4). However, the relatively fast growth of MORC-1 foci suggests that additional MORC-1 proteins were likely also added to the growing foci from the population in solution. As DNA compaction proceeded, static MORC-1 foci came into close proximity and merged into larger foci (Figure 5-7A; Video S3). There was a linear relationship between the concentration of MORC-1 and the average number of MORC-1 molecules in each body, as estimated by Cy3 fluorescence (Pearson Correlation coefficient: > 0.99) (Figures 5-7B and 5-8B-D). These results mirror the linear relationship between MORC-1 concentration and DNA compaction rate, and are consistent with a model in which additional MORC-1 molecules added to existing foci contribute to further DNA compaction. Additionally, the growth of existing MORC-1 foci suggests that initial MORC-1 binding events seed additional MORC-1 binding in part via protein-protein interactions.

MORC-1 forms discrete nuclear foci in vivo and forms phase separated droplets *in vitro.*

The MORC-1 foci observed on DNA during compaction are reminiscent of in vivo nuclear bodies observed with plant MORCs (Harris et al., 2016; Moissiard et al., 2012). C. elegans MORC-1 was previously observed to be localized adjacent to dense areas of heterochromatin (Weiser et al., 2017). We used Structured Illumination Microscopy (SIM) to visualize the nuclear distribution of C-terminal 3xFlag tagged MORC-1 expressed under its endogenous promoter in vivo (Weiser et al., 2017). We observed MORC-1 punctate bodies in both mitotically dividing germline stem cells as well as in gonadal nuclei undergoing meiosis (Figure 5-10A). Consistent with previously published data, MORC-1 bodies were mostly adjacent to, but not overlapping with, the most DAPI-dense areas corresponding to heterochromatin (Figure

5-10A) (Weiser et al., 2017). Imaging these same cell types at higher resolution using stochastic optical reconstruction microscopy (STORM) (Köhler et al., 2017) showed that, in addition to large MORC-1 puncta, a series of smaller bodies could also be seen throughout the nucleus (Figure 4-9A). These data are also consistent with observations that MORCs in multiple species form punctate bodies in vivo (Harris et al., 2016; Mimura et al., 2010; Moissiard et al., 2012). Together with the MORC-1 foci formation on flow stretched DNAs, these results suggest that MORC-1 acts in part via higher order oligomerization to cause DNA compaction.

Some proteins found in nuclear bodies like the heterochromatic protein HP1 have properties of phase separated condensates (Alberti et al., 2019; Boija et al., 2018; Larson et al., 2017; Sabari et al., 2018; Strom et al., 2017). MORC-1 contains a long predicted intrinsically disordered region (IDR) as determined by IUPred2A, and IDRs are known to be able to mediate phase separation (Figure 5-1A) (Lin et al., 2015). We therefore tested whether purified MORC-1 protein may be able to form phase separated droplets in vitro. Using both confocal and differential interference contrast (DIC) microscopy, we observed MORC-1 phase separated droplets at 75 mM NaCl in the presence of 10% polyethylene glycol 3350 (PEG 3350), a crowding agent that can mimic a crowded cellular environment (Lin et al., 2015; Protter et al., 2018) (Figure 5-9B). We then tested for droplet fusion to confirm that these were phase-separated droplets (Lin et al., 2015; Protter et al., 2018). Cy3 or Cy5-labeled MORC-1 droplets were pre-formed in 10% PEG 3350, then mixed and observed to form droplets containing both Cy3 and Cy5 dye (Figure 5-9B). Furthermore, addition of λ-DNA to MORC-1 in 75 mM NaCl without any PEG crowding agent caused MORC-1 to coalesce into fiber-like conformations that are also likely representative of a phase separated state (Figure 5-9B) (Lin et al., 2015). Thus MORC-1 is able to form phase separated droplets in vitro.

A subset of proteins found in cellular bodies with properties of phase transitioned condensates show dispersion in vivo after exposure to 1,6 hexanediol, a solvent that can interfere with weak hydrophobic interactions (Alberti et al., 2019; Boija et al., 2018; Larson et al., 2017; Sabari et al., 2018; Strom et al., 2017). We thus subjected the same mitotic and meiotic germ cells to 5% 1,6 hexanediol treatment and quantified whether the MORC-1 bodies remained punctate or became diffuse upon treatment. We found that the proportion of MORC-1 bodies that were punctate remained largely unchanged upon 5% 1,6 hexanediol treatment, even though ribonucleoprotein bodies that represented P granules became largely dispersed in the presence of 1,6 hexanediol (Updike et al., 2011) (Figure 5-10B). Thus, while the in vivo MORC-1 bodies may represent intracellular condensates, these bodies are likely held together by additional complex interactions other than weak hydrophobic interactions Kroschwald et al., 2015; Kroschwald et al., 2017).

MORC-1 topologically entraps DNA

The observation that MORC-1 nuclear bodies were largely 1,6 hexanediol resistant in vivo suggested that like HP1, these MORC-1 bodies may engage chromatin to produce a more metastable phase separated state in vivo. The presence of additional scaffolding materials like nucleic acids can impact the physical properties of nuclear bodies (Kroschwald et al., 2015; Larson and Narlikar, 2018; Shin and Brangwynne, 2017; Strom et al., 2017; Wheeler et al., 2016). In addition, our in vitro phase transition experiments showed that MORC-1 formed granular fibrils upon incubation with lambda DNA (Figure 5-9B). Thus, a possible explanation for MORC-1 body resistance to 1,6 hexanediol in vivo is that it is stably engaged with DNA (Alberti et al., 2019; Larson and Narlikar). MORC ATPases have also been proposed to act like other GHKLs that can gate and encircle their substrates, and we therefore sought to test whether

MORC-1 could topologically entrap DNA using single molecule experiments (Corbett and Berger, 2005; Douse et al., 2018; Li et al., 2016 [See Chapter 4]; Mimura et al., 2010).

We observed that pre-formed, static Cy3-labeled MORC-1 bodies on flow stretched DNAs without quantum dots became mobile when subjected to a high salt wash (500 mM NaCl). Rather than slowly disintegrating, they were frequently observed to move to the free end of DNA as a single unit before disengaging from the DNA and traveling in the direction of buffer flow (Video S5). This implies that MORC-1 may be topologically entrapping DNA, and thus needs to slide off the free end to be released. To test this hypothesis, we attached quantum dots to the free end of DNAs to determine whether this blocked the dissociation of MORC-1 foci. Indeed, quantum dot-labeled DNAs retained MORC-1 foci at a much higher rate (82% retained, n=28) than unlabeled DNAs (16% retained, n=32) (Figure 5-11; Video S6, Video S7). In addition, MORC-1 foci often accumulated at the free end when quantum dots were present (Figure 5-11). Static MORC foci could similarly block the movement of mobile foci (Figure 5-12A; Video S8). These observations are consistent with MORC-1 topologically entrapping DNA.

To further confirm topological entrapment, we incubated 8xHis-tagged MBP-MORC-1 with plasmid DNAs of various topologies (supercoiled, open circular, or linear), captured MORC-1–DNA complexes with nickel resin, and then washed with either a low salt or high salt wash (Cuylen et al., 2011; Kanno et al., 2015; Murayama and Uhlmann, 2013). MORC-1 remained bound to all forms of DNA after low salt washes (150 mM NaCl) (Figures 5-2A and 5-13A-5-13B). However, upon high salt washes, MORC-1 retained open circular and supercoiled forms of DNA but did not retain significant amounts of linear DNA (Figures 5-13A-5-13B). This occurred without added ATP, and addition of ATP or AMP-PNP had only minor effects on retention

preferences (Figures 5-14B-E). We also confirmed the preference of MORC-1s for retention of circular DNA by first incubating MORC-1 with nicked circular plasmid DNA, and then linearizing the plasmid via restriction enzyme digest. We found that MORC-1 retained the circular plasmid but lost the linearized DNA (Figure 5-13B). Together, these results indicate that MORC-1, like eukaryotic SMC proteins (Cuylen et al., 2011; Kanno et al., 2015; Murayama and Uhlmann, 2013), are able to topologically entrap DNAs.

MORC-1 can compact nucleosomal DNA.

Because MORCs likely act on chromatin in vivo, we tested whether MORC-1 could compact a chromatinized template in our flow-based single-molecule experiments. We assembled nucleosomes on λ DNA with a single biotinylated end using HeLa core histones, the histone chaperone NAP1, and ATP-utilizing chromatin assembly and remodeling factor protein complex (ACF). The nucleosomal template was then tethered to the flow cell via the biotin linkage and visualized with SYTOX Orange staining. We observed that the length of the chromatinized substrates (mode = 2.86 nm) was approximately 3.5 fold shorter than mock-treated unassembled λ DNA (mode = 10.08 nm) (Figure 5-14A). This degree of DNA condensation is consistent with a previous report that showed that DNA with repeated positioning sequences with nucleosomes was lengthened by approximately 3.6 fold upon removal of the histone octamers (Brower-Toland et al., 2002). Introduction of 75 nM MORC-1 into the flow cell resulted in robust compaction (Figures 5-14B and C). Notably, independent of initial substrate length and thus the amount of nucleosomal compaction, the chromatinized substrates compacted to less than 1 micron in length, demonstrating that MORC-1 can efficiently compact nucleosomal arrays (Figure 5-14C).

DISCUSSION

Here we demonstrate that MORC-1 can bind to, compact, and topologically entrap DNA. From single molecule flow stretched DNA experiments, we observed that after MORC-1 binds DNA, it seeds the formation of foci containing multiple MORC-1 proteins, which mirrors the MORC-1 nuclear puncta seen in vivo. MORC-1 is also able to compact DNA that has been assembled into nucleosome arrays.

The finding that MORC-1 compaction activity is stimulated by AMP-PNP more than by ATP is consistent with a model in which ATP binding by MORC-1, rather than ATP hydrolysis, promotes a conformation that can loop and compact DNA. MORC family members are known to dimerize through their N-terminal ATPase domain upon ATP binding (Douse et al., 2018; Li et al., 2016 [See Chapter 4]), and MORC-1 bound to AMP-PNP should therefore be constitutively dimerized at its N terminus. AMP-PNP likely promotes MORC-1 stability on DNA because it disfavors the opening of the ATPase dimers, preventing topologically entrapped DNA from dissociating from MORC-1. It was recently reported that mutations in human MORC2 that create constitutive dimerization within the N-terminal ATPase domain were also more efficient than wild-type at silencing a transgene reporter, whereas mutations that disrupted ATP binding or dimer formation could not re-establish silencing (Douse et al., 2018). Together with our data, this supports a model in which MORCs utilize ATP metabolism to stimulate the conformational changes that regulate their stability and turnover on DNA.

MORC-1 demonstrates a preference for binding long DNAs over shorter DNAs. This preference can be explained by our observation that MORC-1 dimers can diffuse along DNA at low concentrations, and forms assemblages of multiple MORC-1 molecules on DNA at higher

concentrations. Thus longer DNA substrates are able to, on average, load more MORC-1 molecules. The observation that MORC dimers can freely diffuse, but then become static before growing into larger foci suggests that a conformational change occurs, likely coupled with initial DNA loop trapping, at the time that MORC-1 becomes statically localized on DNA. We propose that this conformational change also stimulates the interaction of MORC-1 with other freely diffusing MORC-1 molecules, recruiting them to static MORC-1 foci to initiate further DNA compaction. Both DNA compaction and the number of MORC-1 molecules per body scale linearly with MORC-1 concentration, suggesting an ordered process in which additional MORC-1 molecules recruited to static foci contribute to additional steps of DNA looping and compaction. This process eventually leads to assemblages of MORC-1 molecules that stably compact DNA.

Though MORC-1 preferred longer over shorter DNAs, it appeared to display little preference for particular sequences, as it was able to bind to different DNA templates, including a DNA ladder composed of many different sequences. It therefore seems unlikely that MORC-1 would use DNA sequence as the basis for selecting its chromatin targets in vivo, which is consistent with the view that MORCs in different organisms act as epigenetic regulators that silence targets such as transposons that are composed of a diverse array of sequences (Moissiard et al., 2012; Pastor et al., 2014). It seems more likely that MORCs are guided to their chromatin targets by other proteins or by epigenetic marks. In this regard, one well understood example is mouse MORC3 which contains a CW domain that binds to H3K4 trimethylation in vitro, and which shows co-localization with H3K4 trimethylation in vivo (Li et al., 2016 [See Chapter 4]). C. elegans MORC-1 also contains a CW domain, but it is currently unknown whether it serves to bind methylated histones.

We observed that MORC-1 formed nuclear foci in vivo, and that the MORC-1 protein contains a large intrinsically disordered region and could form phase separated droplets in vitro. Plant MORCs also form prominent nuclear bodies in vivo (Harris et al., 2016; Moissiard et al., 2012), and also contain predicted intrinsically disordered regions, suggesting that these features may be generally important for the function of eukaryotic MORC. Several chromatin regulators such as HP1 and Mediator have been shown to assemble into condensates to create an epigenetic nuclear microenvironment that enforces a particular epigenetic state (Boija et al., 2018; Larson et al., 2017; Sabari et al., 2018; Shin et al., 2018; Strom et al., 2017). Similarly, we propose that MORC-1 may first bind, topologically entrap, and locally compact its chromatin targets, but also form intracellular condensates to concentrate MORC-1 target loci into the observed in vivo MORC-1 bodies within nuclei.

In summary, the results of this study suggest that MORC-1 acts via a novel mechanism to trap a loop of DNA and then multimerize upon DNA binding to form topologically entrapped foci to cause stable chromatin compaction (Figure 5-15). Given that eukaryotic MORC proteins are highly conserved and evolved from ancient prokaryotic restriction-modification systems (Iyer et al., 2008), it is likely that other plant and animal MORCs use a similar mechanism to enforce chromatin compaction.

References:

Alberti, S., Gladfelter, A., and Mittag, T. (2019). Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. Cell 176, 419-434.

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al., (2018). Transcription Factors Activate Genes

through the Phase-Separation Capacity of Their Activation Domains. Cell 175, 1842-1855.e1816.

Brower-Toland, B.D., Smith, C.L., Yeh, R.C., Lis, J.T., Peterson, C.L., and Wang, M.D. (2002). Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA. Proceedings of the National Academy of Sciences 99, 1960-1965.

Corbett, K.D., and Berger, J.M. (2005). Structural Dissection of ATP Turnover in the Prototypical GHL ATPase TopoVI. Structure 13, 873-882.

Cuylen, S., Metz, J., and Haering, C.H. (2011). Condensin structures chromosomal DNA through topological links. Nature Structural & Molecular Biology 18, 894-901.

Dong, W., Vannozzi, A., Chen, F., Hu, Y., Chen, Z., Zhang, L., and Pritham, E. (2018). MORC Domain Definition and Evolutionary Analysis of the MORC Gene Family in Green Plants. Genome Biology and Evolution 10, 1730-1744.

Douse, C.H., Bloor, S., Liu, Y., Shamin, M., Tchasovnikarova, I.A., Timms, R.T., Lehner, P.J., and Modis, Y. (2018). Neuropathic MORC2 mutations perturb GHKL ATPase dimerization dynamics and epigenetic silencing by multiple structural mechanisms. Nature communications 9.

Dutta, R., and Inouye, M. (2000). GHKL, an emergent ATPase/kinase superfamily. Trends in Biochemical Sciences 25, 24-28.

Edelstein, A.D., Tsuchida, M.A., Amodaj, N., Pinkard, H., Vale, R.D., and Stuurman, N. (2014). Advanced methods of microscope control using μManager software. Journal of Biological Methods 1.

Ganji, M., Shaltiel, I.A., Bisht, S., Kim, E., Kalichava, A., Haering, C.H., and Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. Science 360, 102-105.

Graham, T.G., Wang, X., Song, D., Etson, C.M., van Oijen, A.M., Rudner, D.Z., and Loparo, J.J. (2014). ParB spreading requires DNA bridging. Genes & development 28, 1228-1238.

Graham, T.G.W., Walter, J.C., and Loparo, J.J. (2017). Ensemble and Single-Molecule Analysis of Non-Homologous End Joining in Frog Egg Extracts. In DNA Repair Enzymes: Cell, Molecular, and Chemical Biology, B.F. Eichman, ed. (Academic Press), pp. 233-270.

Harris, C.J., Husmann, D., Liu, W., Kasmi, F.E., Wang, H., Papikian, A., Pastor, W.A., Moissiard, G., Vashisht, A.A., Dangl, J.L., et al., (2016). Arabidopsis AtMORC4 and AtMORC7 Form Nuclear Bodies and Repress a Large Number of Protein-Coding Genes. PLOS Genetics 12, e1005998.

Inoue, N., Hess, K.D., Moreadith, R.W., Richardson, L.L., Handel, M.A., Watson, M.L., and Zinn, A.R. (1999). New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. Human Molecular Genetics 8, 1201-1207.

Iyer, L.M., Abhiman, S., and Aravind, L. (2008). MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. Biology direct 3, 8.

Kang, H.-G., Woo Choi, H., von Einem, S., Manosalva, P., Ehlers, K., Liu, P.-P., Buxa, S.V., Moreau, M., Mang, H.-G., Kachroo, P., et al., (2012). CRT1 is a nuclear-translocated MORC endonuclease that participates in multiple levels of plant immunity. Nature communications 3, 1297.

Kanno, T., Berta, Davide G., and Sjögren, C. (2015). The Smc5/6 Complex Is an ATP-Dependent Intermolecular DNA Linker. Cell Reports 12, 1471-1482.

Kim, H., and Loparo, J.J. (2016). Multistep assembly of DNA condensation clusters by SMC. Nature communications 7, 10200.

Kim, H., and Loparo, J.J. (2018). Observing Bacterial Chromatin Protein-DNA Interactions by Combining DNA Flow-Stretching with Single-Molecule Imaging. In Bacterial Chromatin, pp. 277-299.

Köhler, S., Wojcik, M., Xu, K., and Dernburg, A.F. (2017). Superresolution microscopy reveals the three-dimensional organization of meiotic chromosome axes in intact Caenorhabditis elegans tissue. Proc Natl Acad Sci USA 114, E4734-E4743.

Kroschwald, S., Maharana, S., Mateju, D., Malinovska, L., Nüske, E., Poser, I., Richter, D., and Alberti, S. (2015). Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules. eLife 4.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. Nature 547, 236-240.

Larson, A.G., and Narlikar, G.J. (2018). The Role of Phase Separation in Heterochromatin Formation, Function, and Regulation. Biochemistry 57, 2540-2548.

Li, S., Yen, L., Pastor, W.A., Johnston, J.B., Du, J., Shew, C.J., Liu, W., Ho, J., Stender, B., Clark, A.T., et al., (2016). Mouse MORC3 is a GHKL ATPase that localizes to H3K4me3 marked chromatin. Proceedings of the National Academy of Sciences of the United States of America 113, E5108-5116.

Lin, Y., Protter, D.S.W., Rosen, M.K., and Parker, R. (2015). Formation and Maturation of Phase Separated Liquid Droplets by RNA Binding Proteins. Molecular cell 60, 208-219.

Mimura, Y., Takahashi, K., Kawata, K., Akazawa, T., and Inoue, N. (2010). Two-step colocalization of MORC3 with PML nuclear bodies. Journal of cell science 123, 2014-2024.

Moissiard, G., Bischof, S., Husmann, D., Pastor, W.A., Hale, C.J., Yen, L., Stroud, H., Papikian, A., Vashisht, A.A., Wohlschlegel, J.A., et al., (2014). Transcriptional gene silencing by Arabidopsis microrchidia homologues involves the formation of heteromers. Proceedings of the National Academy of Sciences of the United States of America 111, 7474-7479.

Moissiard, G., Cokus, S.J., Cary, J., Feng, S., Billi, A.C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B.R., McCord, R.P., et al., (2012). MORC family ATPases required for heterochromatin condensation and gene silencing. Science 336, 1448-1451.

Murayama, Y., and Uhlmann, F. (2013). Biochemical reconstitution of topological DNA binding by the cohesin ring. Nature 505, 367-371.

Pastor, W.A., Stroud, H., Nee, K., Liu, W., Pezic, D., Manakov, S., Lee, S.A., Moissiard, G., Zamudio, N., Bourc'his, D., et al., (2014). MORC1 represses transposable elements in the mouse male germline. Nature communications 5, 5795.

Protter, D.S.W., Rao, B.S., Van Treeck, B., Lin, Y., Mizoue, L., Rosen, M.K., and Parker, R. (2018). Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly. Cell reports 22, 1401-1412.

Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., et al., (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science 361.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al., (2012). Fiji: an open-source platform for biological-image analysis. Nature Methods 9, 676-682.

Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. Science 357.

Shin, Y., Chang, Y.-C., Lee, D.S.W., Berry, J., Sanders, D.W., Ronceray, P., Wingreen, N.S., Haataja, M., and Brangwynne, C.P. (2018). Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. Cell 175, 1481-1491.e1413.

Song, D., Graham, T.G.W., and Loparo, J.J. (2016). A general approach to visualize protein binding and DNA conformation without protein labelling. Nature communications 7.

Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. Nature 547, 241-245.

Takahashi, K., Yoshida, N., Murakami, N., Kawata, K., Ishizaki, H., Tanaka-Okamoto, M., Miyoshi, J., Zinn, A.R., Shime, H., and Inoue, N. (2007). Dynamic regulation of p53 subnuclear localization and senescence by MORC3. Molecular biology of the cell 18, 1701-1709.

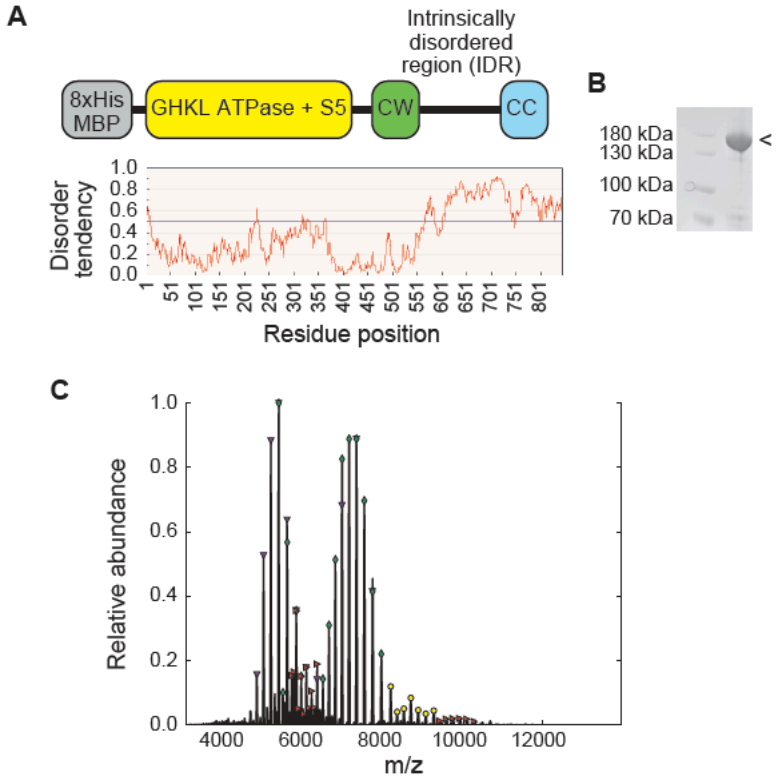
Tchasovnikarova, I.A., Timms, R.T., Douse, C.H., Roberts, R.C., Dougan, G., Kingston, R.E., Modis, Y., and Lehner, P.J. (2017). Hyper-activation of HUSH complex function by Charcot-Marie-Tooth disease mutation in MORC2. Nat Genet 49, 1035-1044.

Updike, D.L., Hachey, S.J., Kreher, J., and Strome, S. (2011). P granules extend the nuclear pore complex environment in theC. elegansgerm line. The Journal of Cell Biology 192, 939-948.

Weiser, N.E., Yang, D.X., Feng, S., Kalinava, N., Brown, K.C., Khanikar, J., Freeberg, M.A., Snyder, M.J., Csankovszki, G., Chan, R.C., et al., (2017). MORC-1 Integrates Nuclear RNAi and Transgenerational Chromatin Architecture to Promote Germline Immortality. Developmental cell 41, 408-423.e407.

Wheeler, J.R., Matheny, T., Jain, S., Abrisch, R., and Parker, R. (2016). Distinct stages in stress granule assembly and disassembly. eLife 5.

Figure 5-1: MORC-1 Native Mass Spectrometry



	Expected mass	Observed mass	Percent error	Rel. percent abundance
Monomer	138.41 kDa	138.56 kDa	0.1 %	41.7 %
Expected	276.82 kDa	277.36 kDa	0.2 %	52.5 %
Trimer	415.23 kDa	416.16 kDa	0.2 %	1.8 %
Tetramer	553.64 kDa	544.76 kDa	1.6 %	4.0 %

Figure 5-2: MORC-1 is a DNA binding protein.

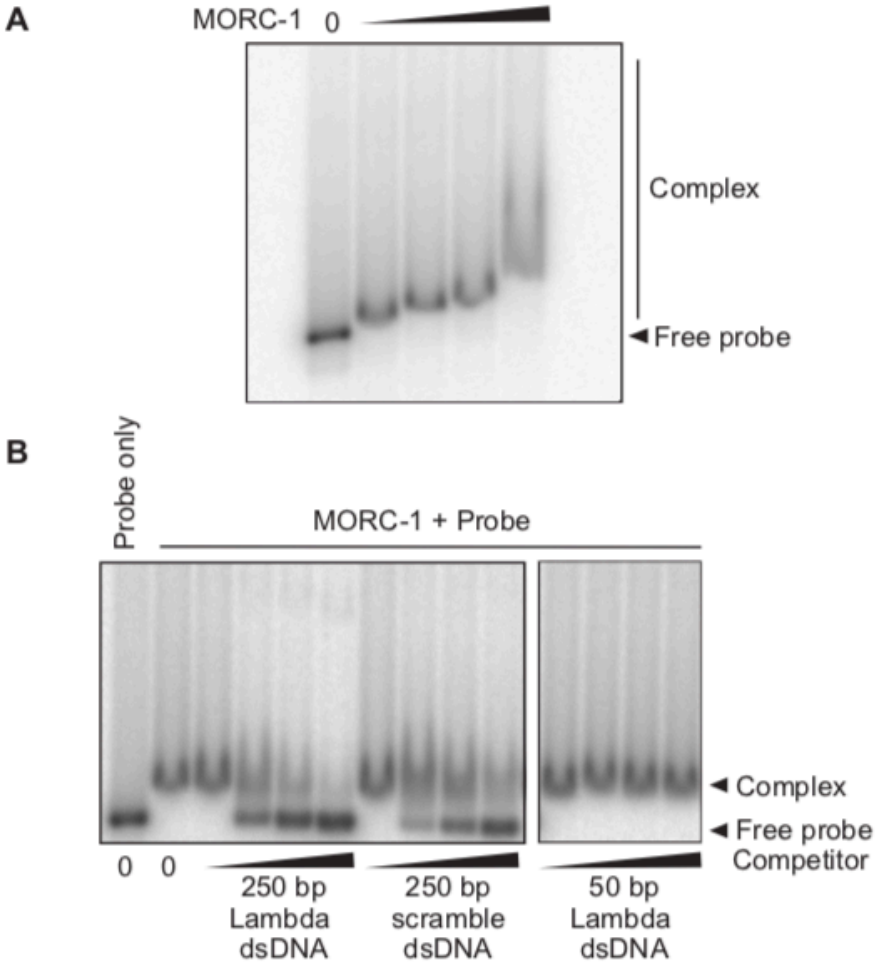


Figure 5-3: MORC-1 binds long DNAs over short DNAs

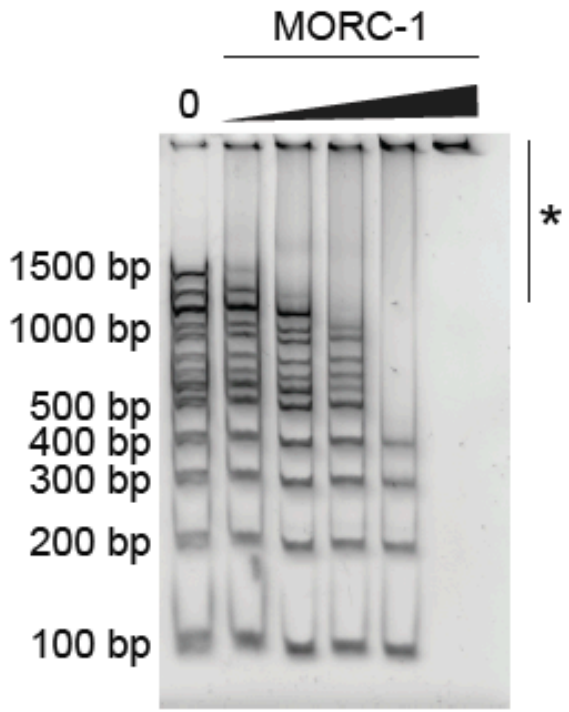


Figure 5-4: MORC-1 compacts DNA using a loop trapping mechanism.

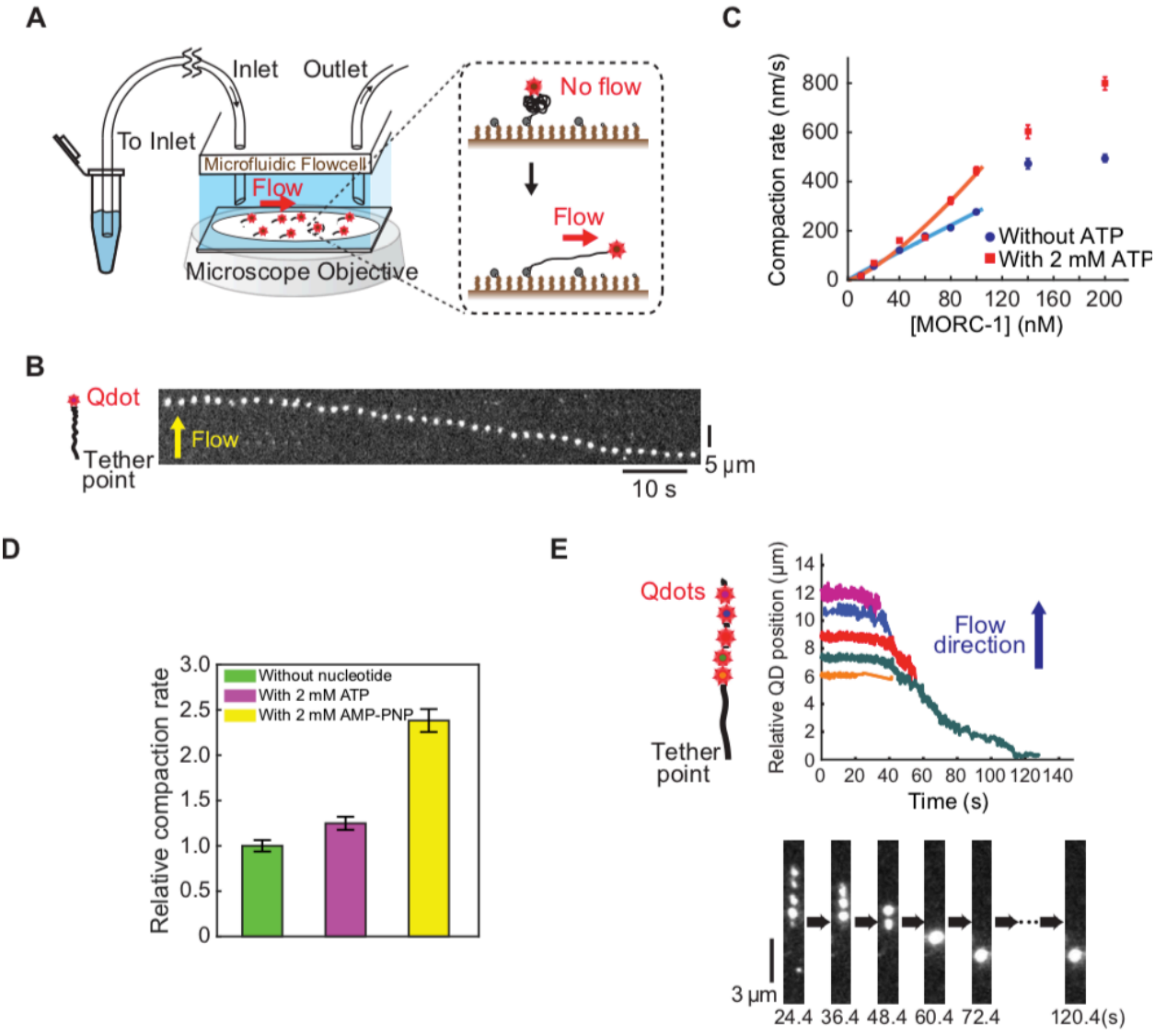


Figure 5-5: MORC-1 compaction is nucleotide responsive.

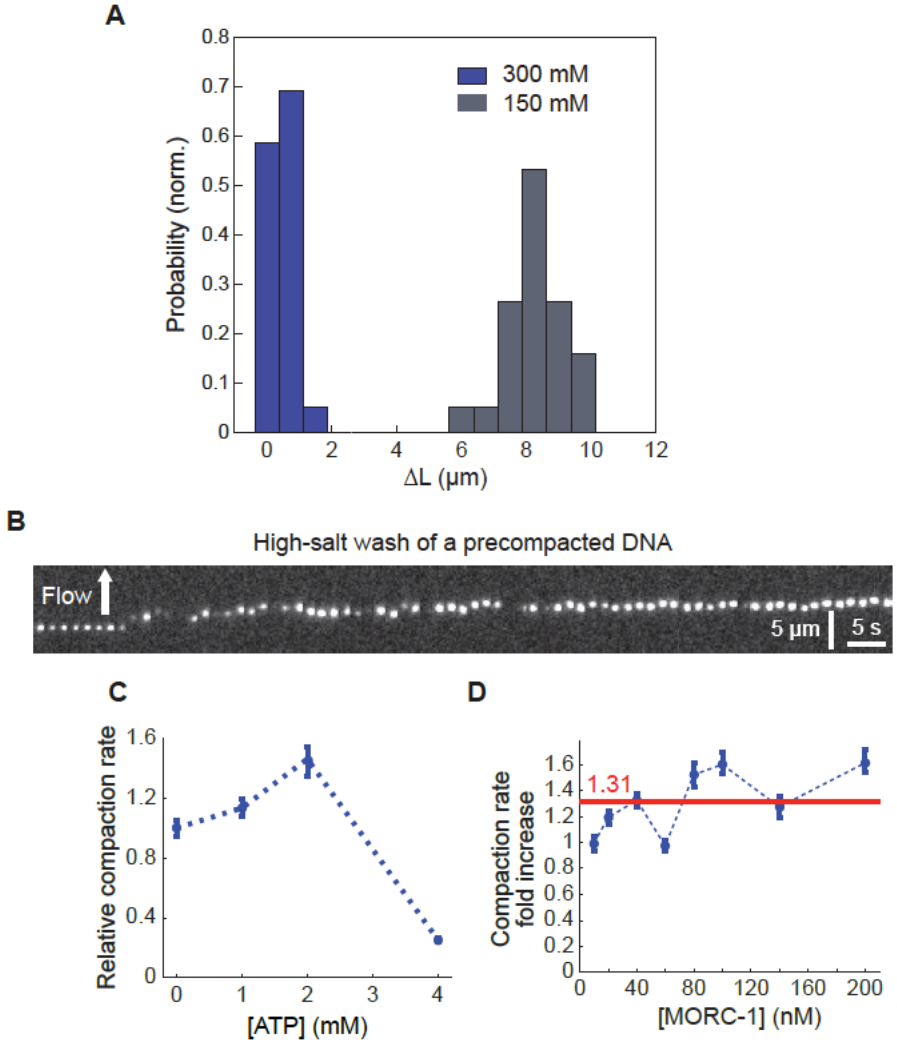


Figure 5-6: MORC-1 compacts doubly-tethered DNA without actively extruding a loop.

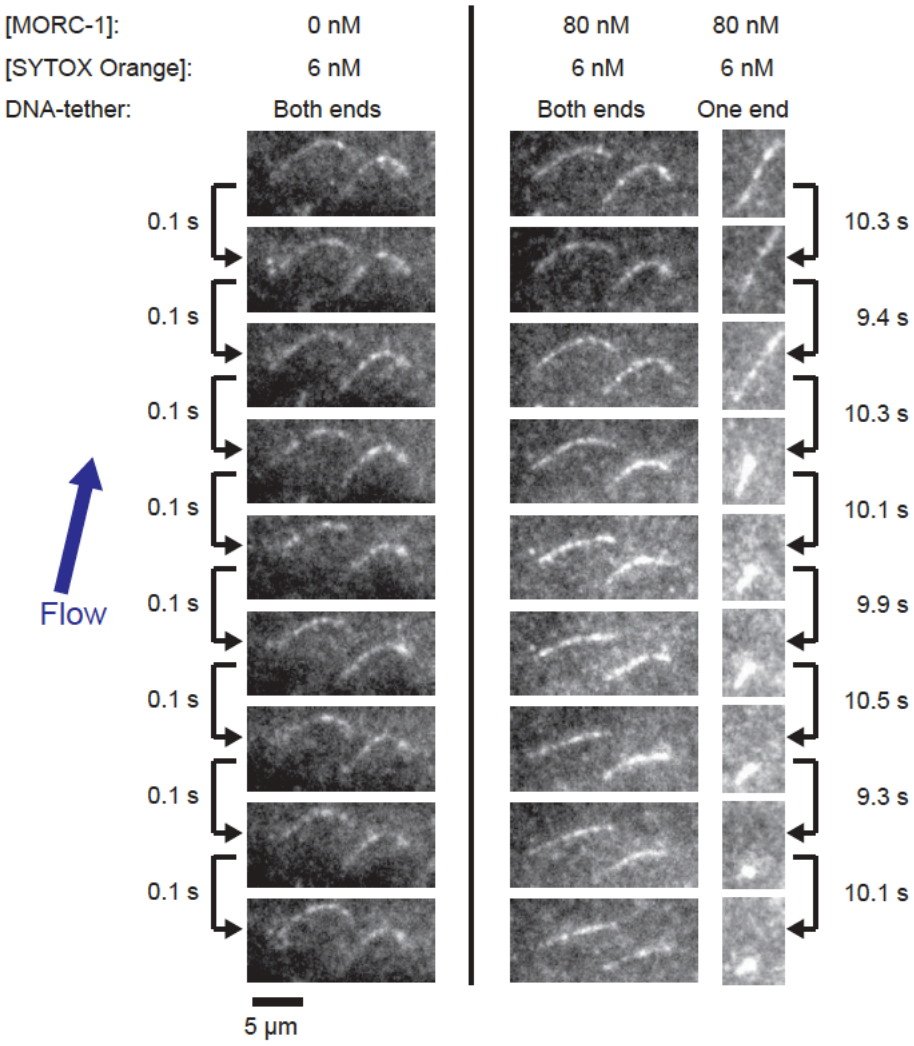


Figure 5-7A: MORC-1 forms bodies that grow in size with concentration.

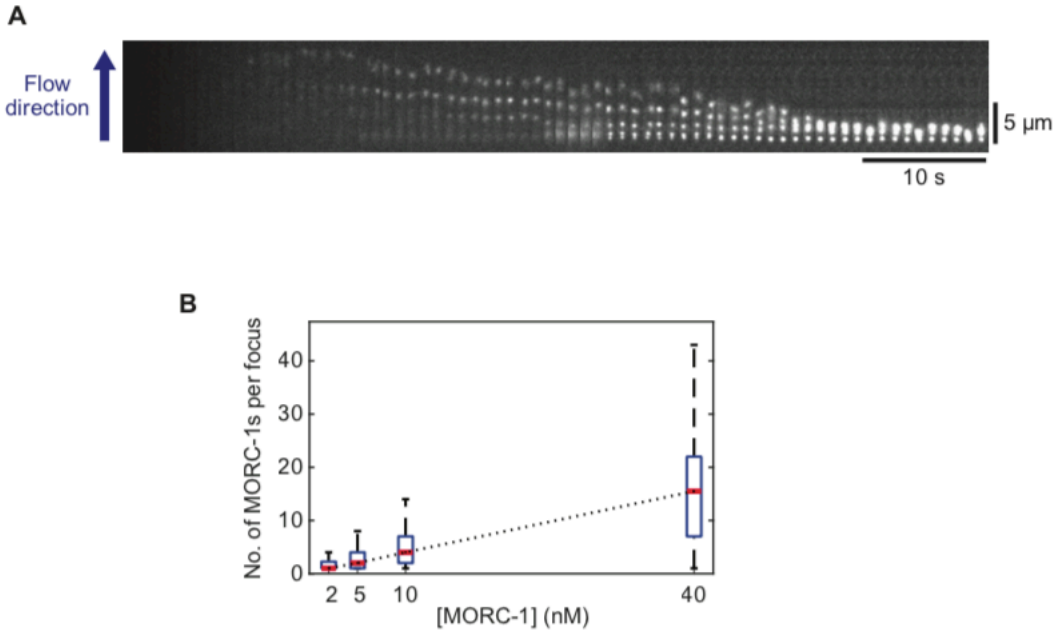


Figure 5-8: MORC-1 can diffuse along DNA and can form foci.

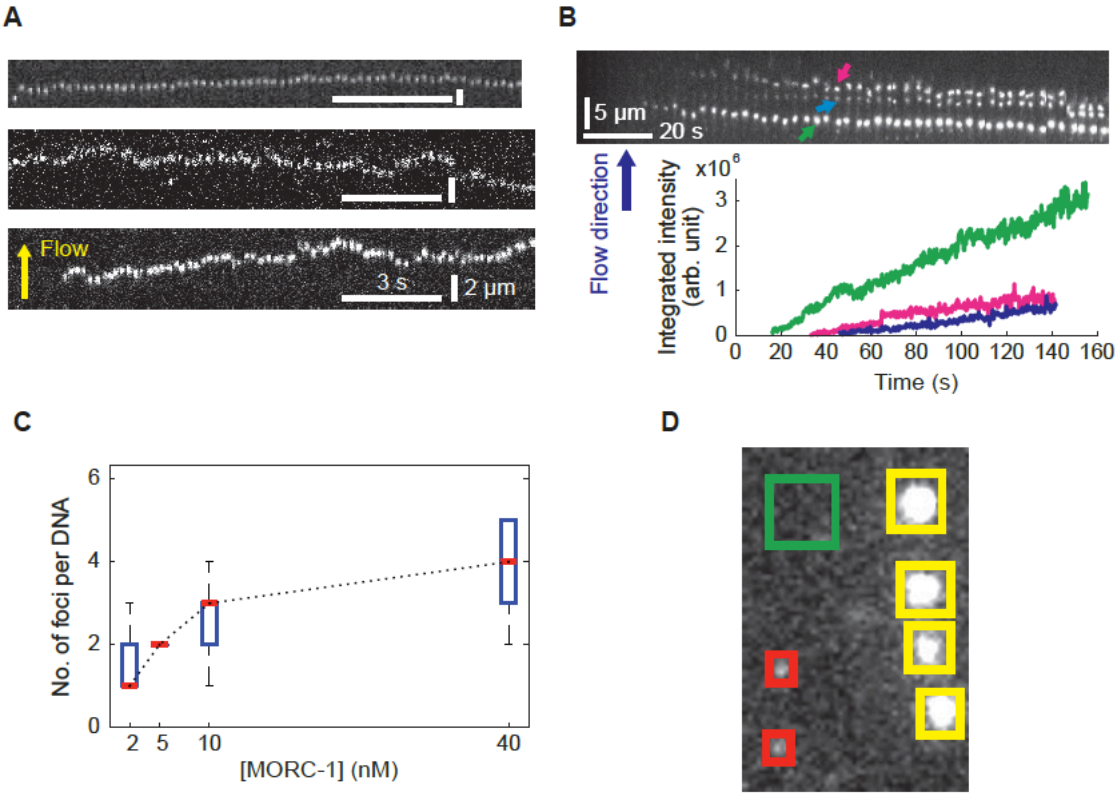


Figure 5-9: MORC-1 is localized to bodies *in vivo* and forms condensates *in vitro*.

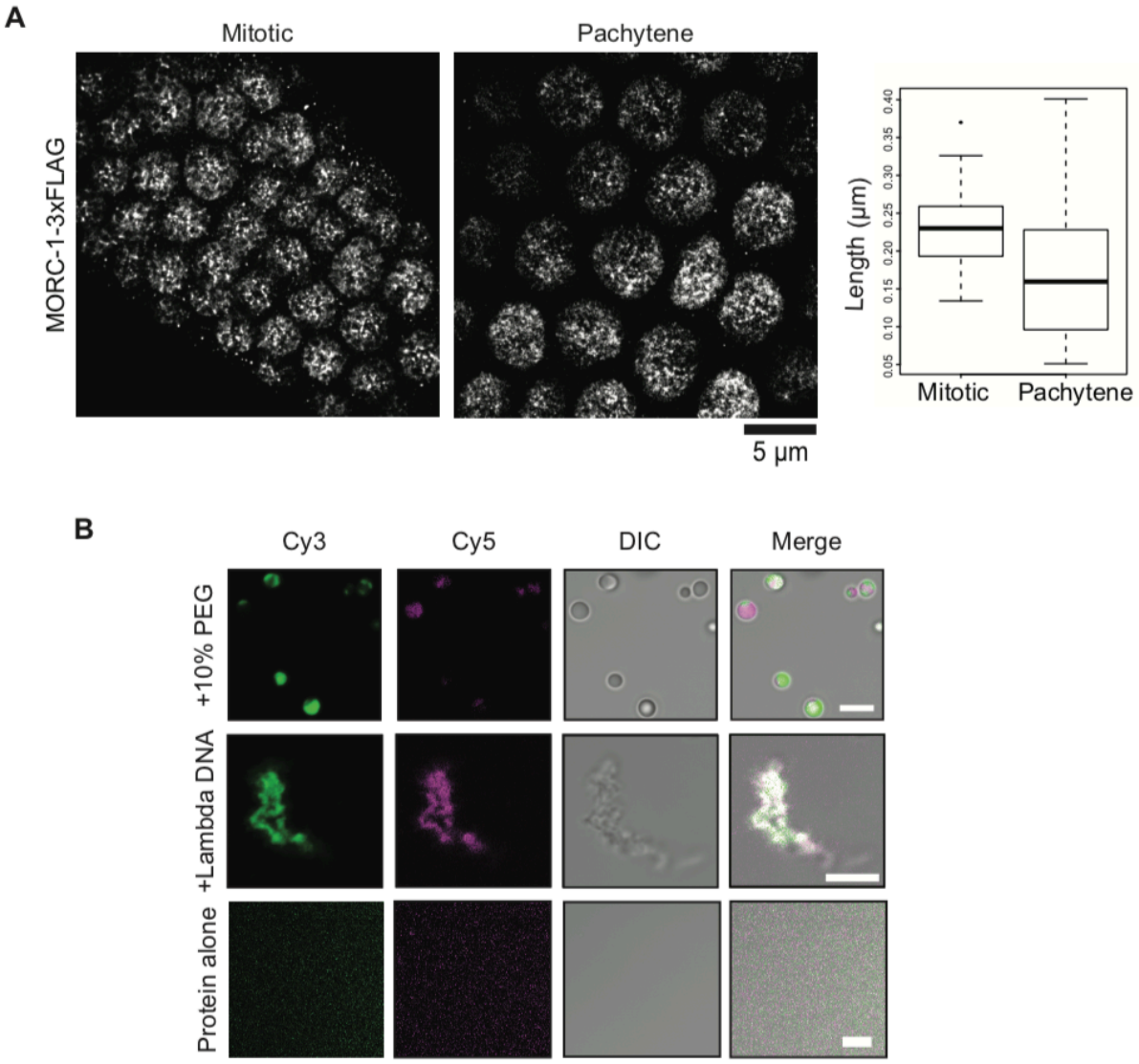


Figure 5-10: MORC-1 puncta are resistant to dissolution by 1,6 hexanediol.

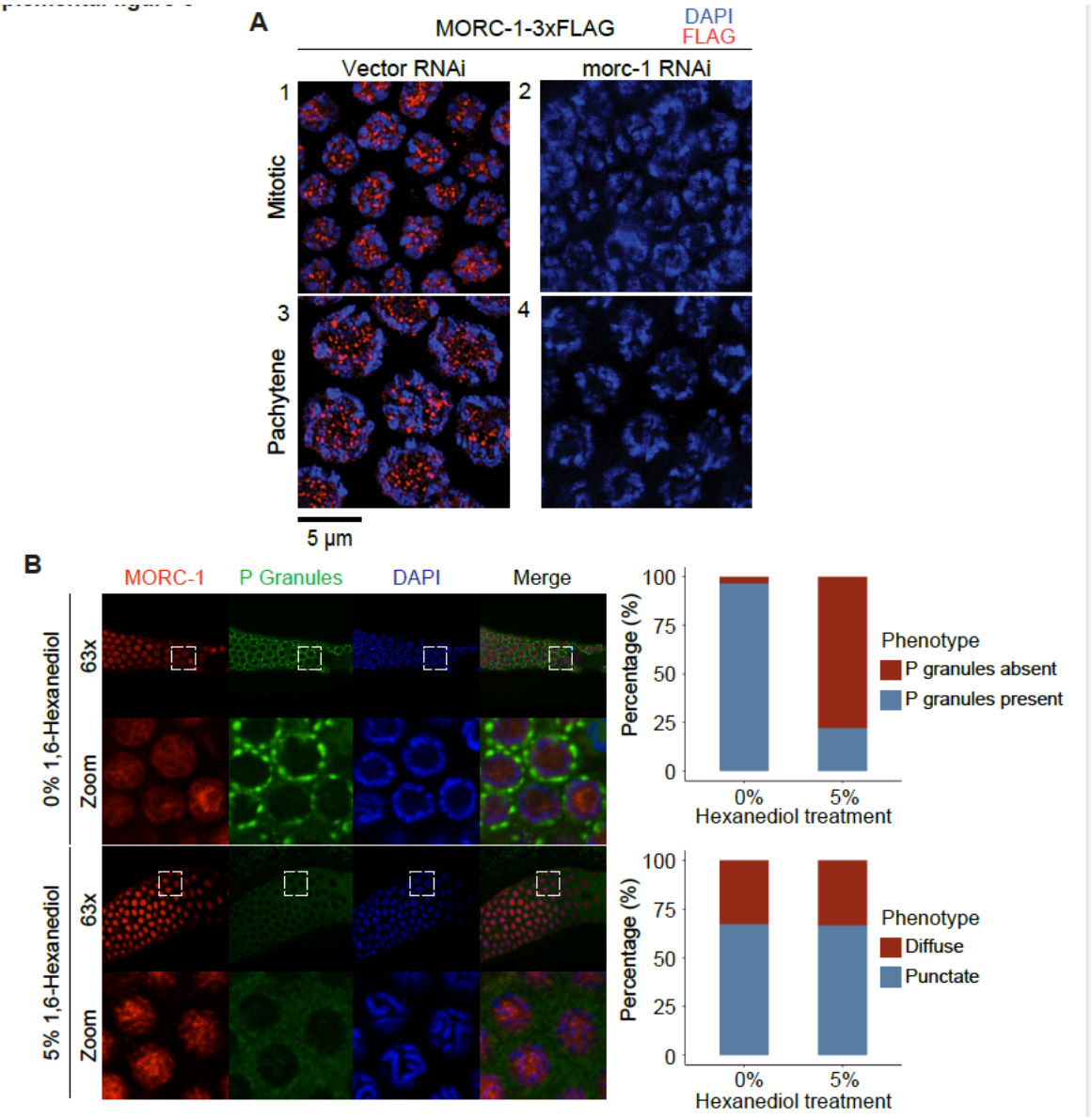


Figure 5-11: MORC-1 foci topologically entrap DNA.

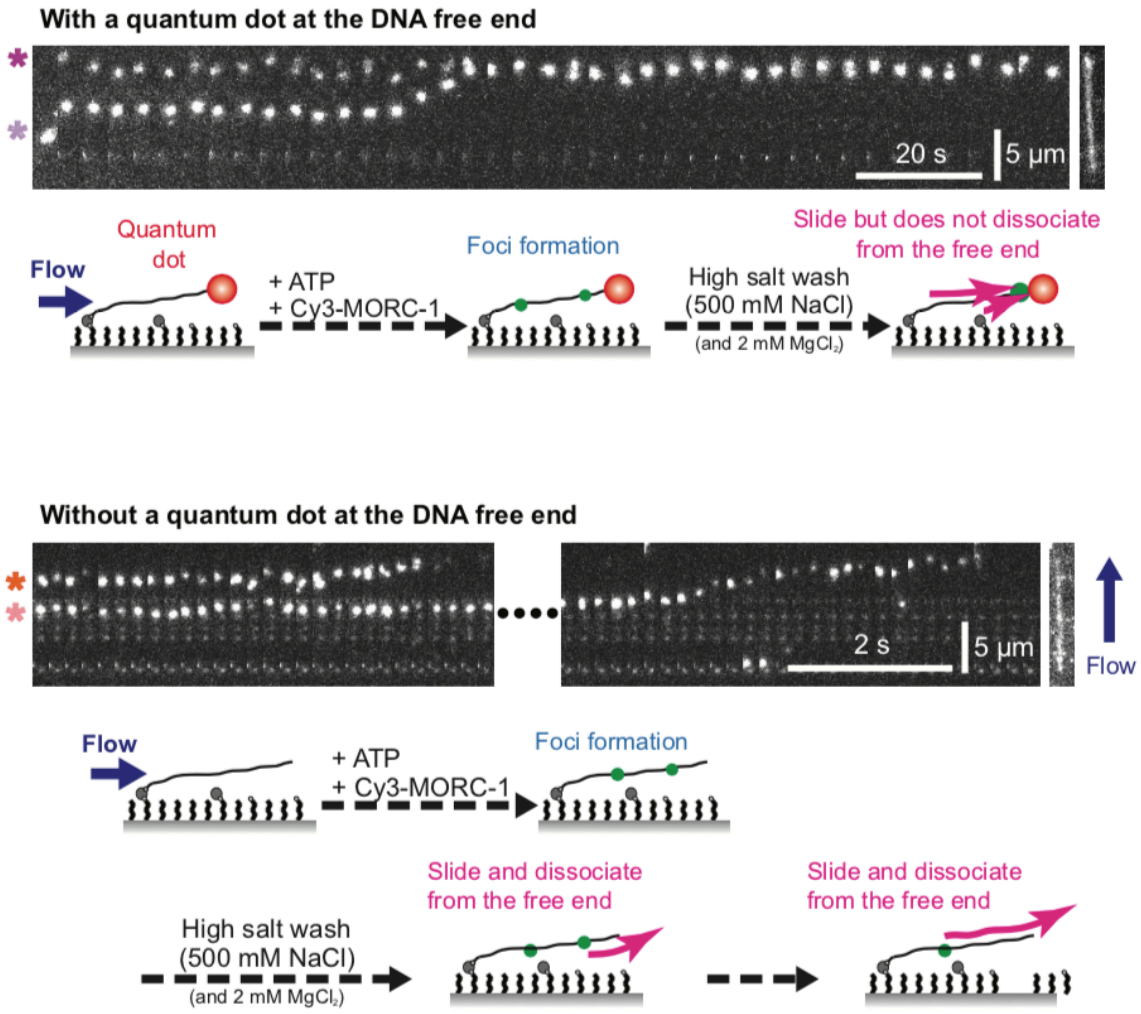


Figure 5-12: MORC-1 selectively retains circular DNA after high salt washes.

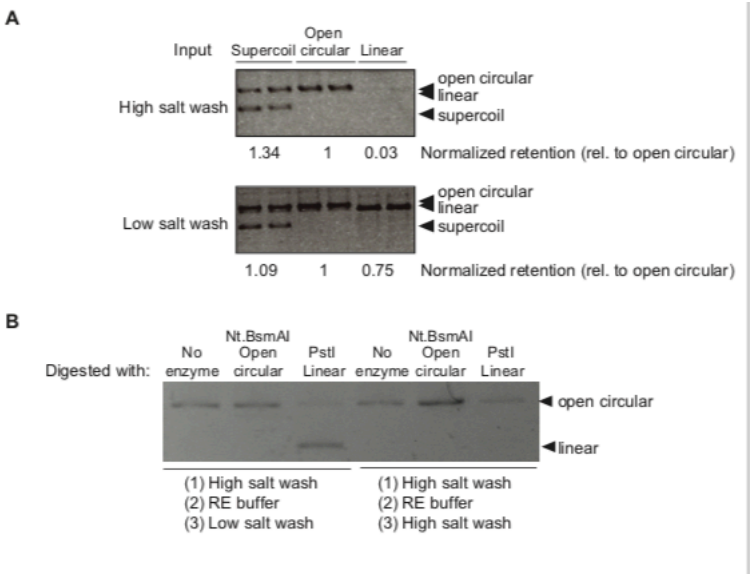


Figure 5-13: Static MORC-1 foci can block other MORC-1 foci from leaving DNA; entrapment is not nucleotide dependent

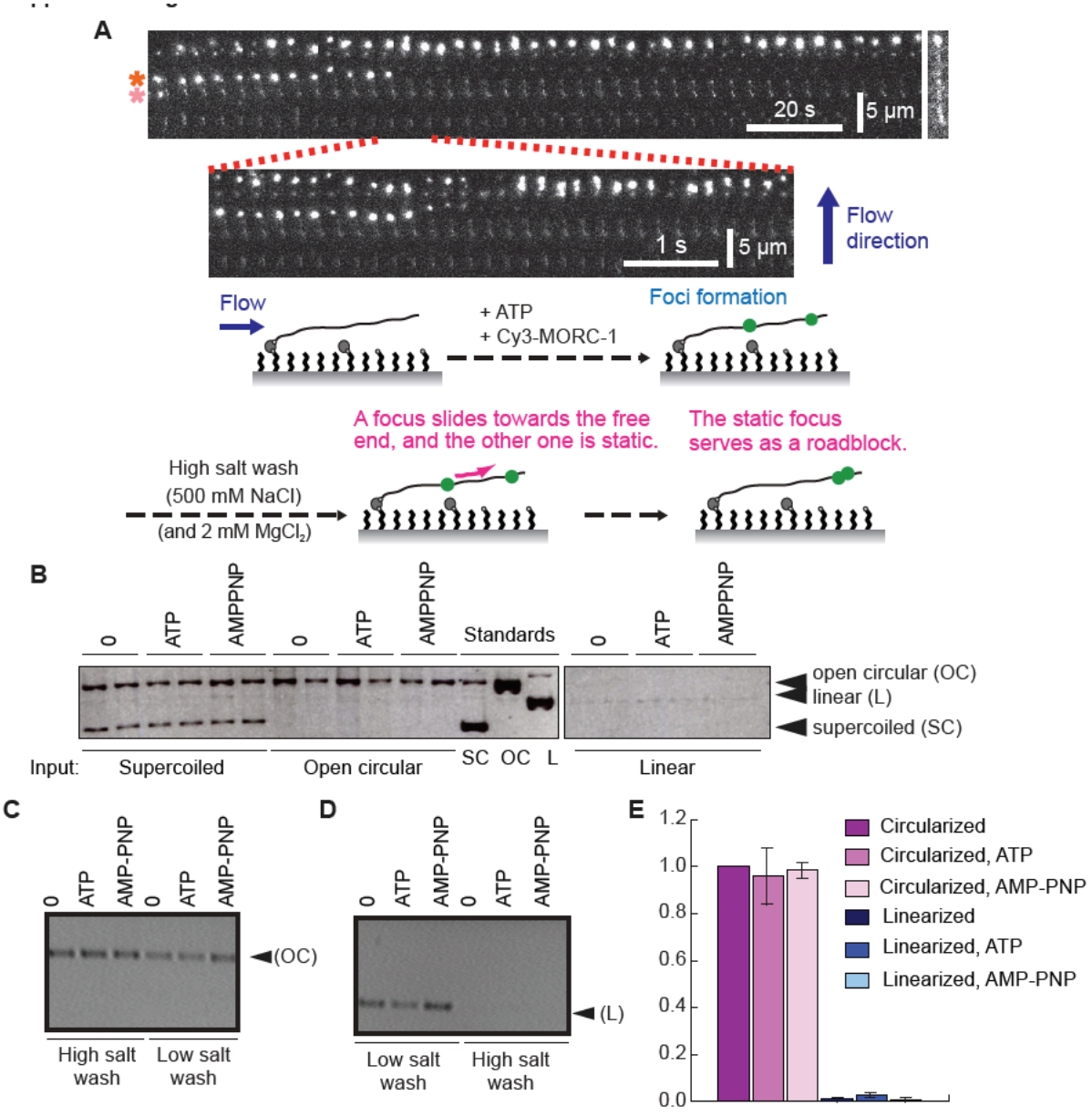


Figure 5-14: MORC-1 compacts chromatin.

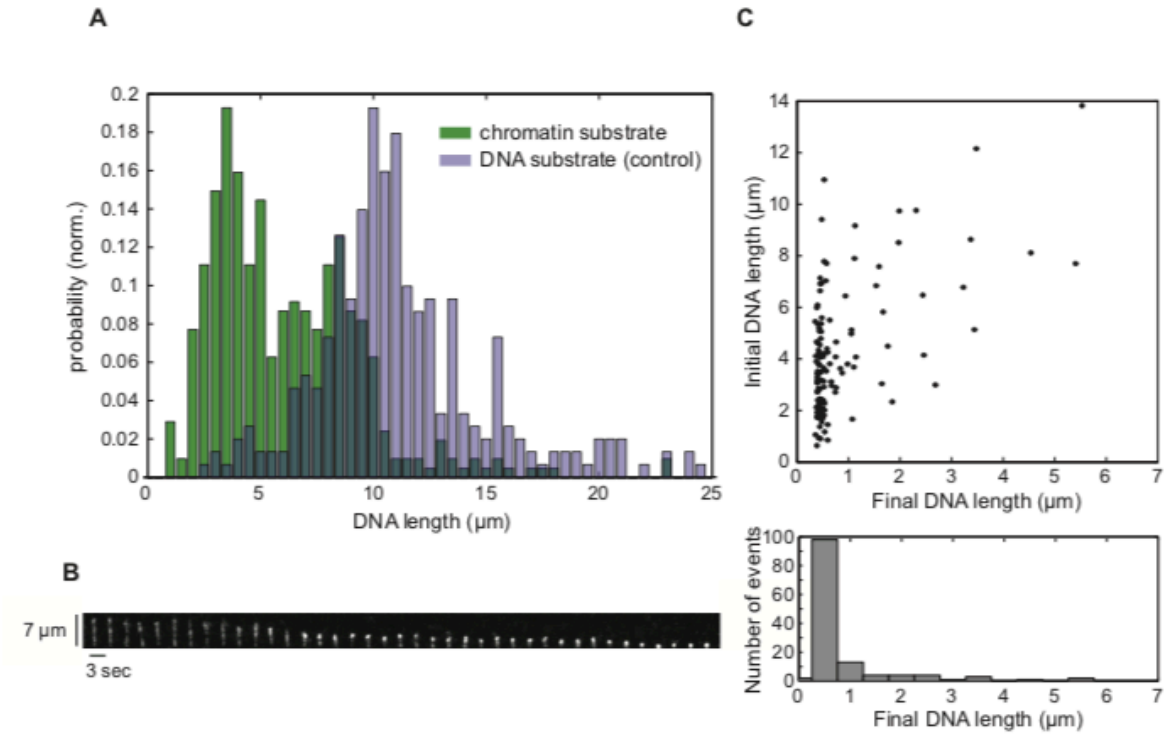


Figure 5-15: Model of MORC-1 compaction

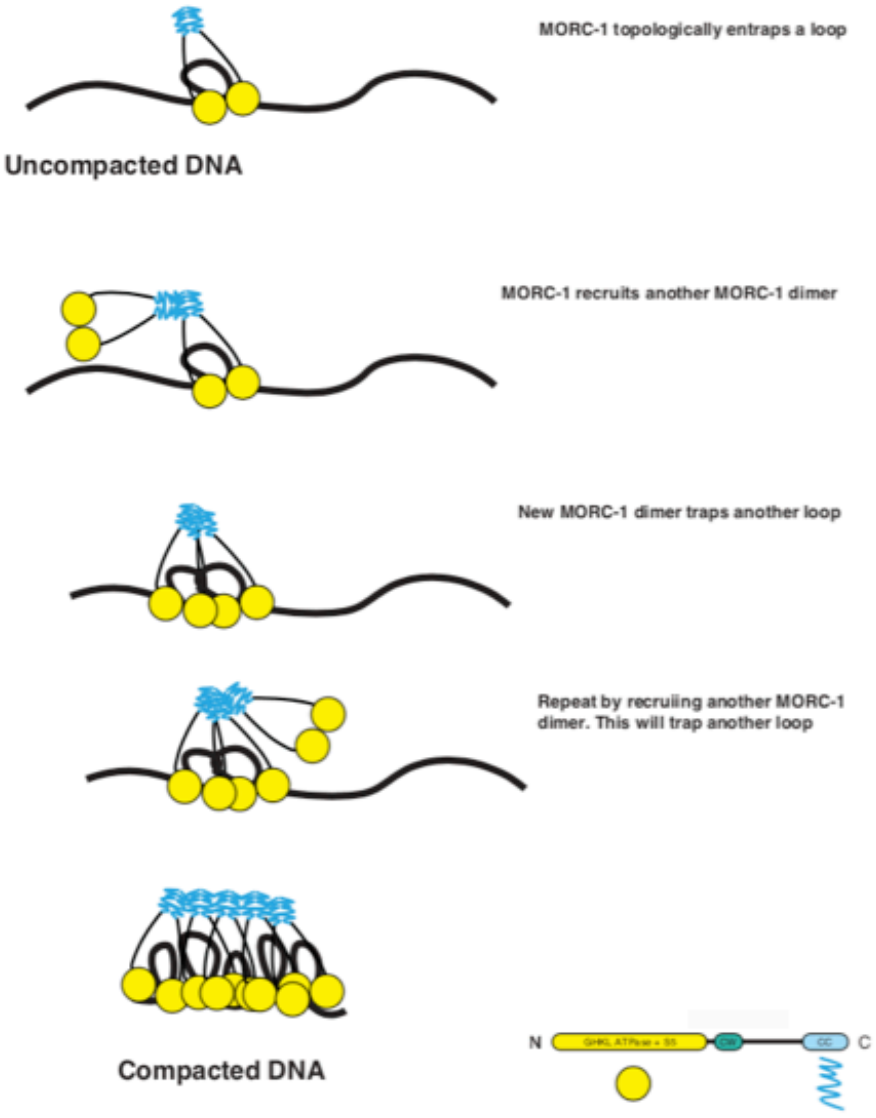


Figure legends

Figure 5-1. MORC-1 Native Mass Spectrometry.

(A) Cartoon schematic of the MORC-1 construct. Protein fusion contains an N terminal 6x Histidine tag and a maltose binding protein (MBP) solubility tag. The catalytic ATPase module is comprised of the GHKL ATPase domain and the S5 fold (S5). MORC-1 also contains a CW domain (CW) in addition to a C terminal coiled coil (CC) domain. IUPRED2A analysis indicates that MORC-1 possesses a disordered region. (B) Coomassie gel of purified 8xHis-MBP-MORC-1 protein. "<" denotes protein. (C) MORC-1 can form multimers. Native mass spectrum of overlapping charge states that correspond to MORC-1 monomer, dimer, trimer, and tetramer. The charge envelope for the monomer is represented by purple triangles, dimer by green diamonds, trimer by yellow circles, and tetramer by red triangles. Accompanying table reports the measured and expected masses in kilodaltons (kDa), and relative abundance in the sample.

Figure 5-2. MORC-1 is a DNA binding protein.

(A) Protein ([MORC-1] =100 nM, 200 nM, 400 nM, 800 nM) was incubated with 32P labeled 250 bp DNA probe derived from λ-DNA and run on a 3% agarose gel, then dried before exposure to a phophor screen. (B) 200 nM MORC-1 was added to 0.5 nM 32P labeled 250bp λ-DNA that was premixed with cold 250 bp λ-DNA, 250 bp scrambled DNA, or 50 bp λ-DNA probe (1 nM, 10 nM, 20 nM, 50 nM) and processed as described above.

Figure 5-3. MORC-1 binds long DNAs over short DNAs.

Protein ([MORC-1]= 200 nM, 400 nM, 800 nM, 1.6 μM, 3.2 μM) was incubated with a double stranded DNA ladder. Complexes were resolved on a 5% native acrylamide gel and visualized using SyBr Gold staining. Asterisk(*) denotes MORC-1 DNA complexes

Figure 5-4. MORC-1 compacts DNA using a loop trapping mechanism.

(A) A schematic of the flow stretched DNA assay.(B) A representative kymograph demonstrating compaction of a quantum dot labelled λ-DNA over time. Scale bars represent 10 seconds and 5 μm. (C) Effect of ATP on MORC-1 compaction. Data was calculated from 48 different experiments, with a total of 1199 different trajectories. n=85, 48, 98, 122, 66, 84, 57, 97 (without ATP), and n=66, 94, 79, 52, 61, 76, 47, 67 (with 2 mM ATP). Error bars: SE. (D) AMP-PNP further stimulates DNA compaction. Compaction rates were from two independent experiments for each condition. Error bars represent the SE. n=45, 46, and 56 for experiments with no nucleotide, 2 mM ATP, and 2 mM AMP-PNP, respectively. (E) Left: schematic of DNA motion capture assay depicting the location of the five EcoR1 binding sites. Middle: Plotted trajectories of location of EcoR1-E111Q conjugated quantum dots over time. Below: representative kymograph. [MORC-1] = 40 nM, [ATP]=1 mM.

Figure 5-5| MORC-1 compaction is nucleotide responsive.

(A) MORC-1 compaction does not robustly occur in 300 mM NaCl or above. Histogram analyzing n=25 compaction events in 150 mM and 300 mM NaCl. (B) Kymograph showing that compaction is partially reversible. [MORC-1]=60 nM, [ATP]=2 mM. Pre-compacted DNA was subjected to a high salt wash (containing 500 mM NaCl, 2 mM MgCl2, 2 mM ATP). Scale bar represents 5 seconds and 5 μm. (C) MORC-1 compaction rate is responsive to ATP

concentration. Relative compaction rates were measured using 100 nM of MORC-1. n=86, 93, 55, 51 for [ATP]=0, 1, 2, 4 mM, respectively. The error bars (SE) were calculated using the error propagation formula. The observed drop in compaction rate is likely due to substrate inhibition. (D) ATP stimulates DNA compaction. The ratios of compaction rates in the presence of ATP to those in the absence of ATP were calculated. The error bars (SE) were calculated using the error propagation formula.

Figure 5-6. MORC-1 compacts doubly-tethered DNA without actively extruding a loop. (Related to Figure 5-4). Slack, doubly tethered DNA was subjected to orthogonal flow and stained with SYTOX Orange. Upon adding MORC-1, double tethered DNA became taut, indicating the presence of DNA compaction without large looping events. Single tethered DNA in the same field of view compacted down to the tether point. Experiment was conducted in the presence of 2 mM ATP.

Figure 5-7A. MORC-1 forms bodies that grow in size with concentration.
(A) A representative kymograph showing Cy3 labeled MORC-1 cluster formation. (B) At varying MORC-1 concentrations, foci intensities were quantified at a fixed time point (140 seconds after flowing in sample) and presented as box plots. The number of clusters was n=21, 34, 55, and 34 for [MORC-1] = 2, 5, 10, and 40 nM, respectively. The red line inside each box corresponds to the median value while the bottom and top edges of each box correspond to the 25th and 75th percentiles, respectively. A Black dotted line connecting nearby median values was added for visual guidance and showed displayed a strong linear correlation (Pearson correlation coefficient: r = 0.99996).

Figure 5-8. MORC-1 can diffuse along DNA and can form foci.

(A) Three examples showing MORC-1 diffusion along flow-stretched DNA at low MORC-1 concentration. The MORC-1 concentration for the top panel is 1 nM, and 2 nM for the middle and the bottom panels. All experiments were performed in the presence of 2 mM ATP. Scale bars correspond to 3 seconds and 2 μm. (B) MORC-1 forms foci on DNA and focus size increases over time. A representative kymograph and accompanying graph illustrating increases in fluorescent intensity of foci marked by magenta, blue, and green arrows over time. [Cy3-MORC-1] = 40 nM, [ATP] = 2 mM. There is no quantum dot on the DNA. (C) Box plots showing the number of foci per DNA. Foci were counted 140 seconds after MORC-1 addition. Red line inside each box: Median value. The bottom and top edges of each box: 25th and 75th percentiles. Black dotted lines connecting nearby median values were added for visual guidance. (D) Method of counting the number of MORC-1 molecules in foci. The net integrated signal intensities for foci (yellow box) were calculated by subtracting integrated background (green box). Likewise, the net integrated signal intensities of single Cy3 (red box) were calculated. The ratio of the net integrated focus signal intensity to that of the single Cy3 is an estimated number of MORC-1 dimers in each focus.

Figure 5-9. MORC-1 is localized to bodies in vivo and forms condensates *in vitro.*

(A) MORC-1 are found in nuclear bodies of varying sizes in vivo. MORC-1 mitotic bodies (n=96) are on average (mean) 229 nm; MORC-1 pachytene bodies (n=53) are on average (mean) 170 nm. Scale bar represents 5 μm. (B) MORC-1 undergoes liquid-liquid phase separation. Unlabeled protein was buffer exchanged into 75 mM NaCl and mixed with labeled protein to achieve a final population of 1-2% labeling. Single color droplets were pre-formed by adding a final concentration of 10% PEG 3350. To image dual color droplets, pre-formed droplets were

mixed together and incubated at room temperature before imaging. To image dual color MORC-1 on DNA, Cy5-MORC-1 was added to Cy3-MORC-1 pre-incubated with DNA. White represents areas of signal overlap. Scale bars represent 5 μm.

Figure 5-10: MORC-1 puncta are resistant to dissolution by 1,6 hexanediol.

(A) Maximum intensity projection images of immunostaining are shown for C. elegans mitotic germ cells (1) and pachytene cells (3) expressing MORC-1::3xFLAG. RNAi against morc-1 abrogates anti-FLAG signal in mitotic cells (2) and pachytene cells (4). (B) MORC-1 puncta are resistant to dissolution by 1,6-hexanediol. (Left) Extruded gonads from C. elegans expressing MORC-1:3xFlag worms are stained with anti-Flag (red), anti-CSR-1 (green), a P granule marker, and DAPI after incubation in buffer or 5% 1,6-hexanediol. (Upper right) Quantification of number of P granules per nucleus in the indicated concentrations of 1,6-hexanediol. n=144 cells for 0% 1,6-hexanediol and n=72 cells for 5% 1,6-hexanediol. p=2.2e-16 (two sample t test). (Lower left) Quantification of the percentage of cells having MORC-1 in a diffuse or punctate state. n=144 cells for 0% 1,6-hexanediol and n=72 cells for 5% 1,6-hexanediol. p=0.9192 (two sample t test). Though the P granules dissolve in 1,6-hexanediol, MORC-1::3xFlag puncta do not.

Figure 5-11. MORC-1 foci topologically entrap DNA.

Representative kymographs of Cy3-MORC-1 foci movement on quantum dot labeled DNA (top panel, [MORC-1]=5 nM, [ATP]=2 mM) and on bare DNA (bottom panel, [MORC-1]=10 nM, [ATP]=2 mM). Schematics below the kymographs illustrate the experiment. MORC-1 is applied to the flow cell and allowed to form foci, then subjected to a 500 mM NaCl and 2 mM MgCl2 wash in the presence of 2 mM ATP. Scale bars represent 20 seconds and 5 μm. SYTOX Orange

stained DNA is shown to the right of the kymographs to indicate approximate location of foci on the DNA. Note that the quantum dot is not visible any of the displayed spectral channels. In the top panel, the top focus (magenta asterisk) contains approximately 2 MORC-1s and the bottom focus (lavender asterisk) contains approximately 7 MORC-1s. In the bottom panel, the time gap between the two events as marked by the dashed line is 7.2 seconds. The top focus (orange asterisk) has approximately 7 MORC-1s and the bottom (peach asterisk) has approximately 13 MORC-1s.

Figure 5-12. MORC-1 selectively retains circular DNA after high salt washes.

(A) MORC-1 was incubated with DNA of varying topologies and precipitated with magnetic resin, then washed twice before proteinase K elution. Retained DNA is visualized by resolving samples on an 1% 1x TAE agarose gel and post-stained with ethidium bromide. (B) 400 nM MORC-1 was incubated with 50 ng open circular DNA that was made by Nt.BsmAI digestion and precipitated with magnetic resin, then washed twice with a high salt buffer and once with Restriction Enzyme (RE) buffer before digesting with restriction enzymes to alter the topology. The sample was then washed using low or high salt, then eluted using proteinase K. Retained DNA is visualized by resolving samples on an 1% 1x TAE agarose gel and post-stained with ethidium bromide. Note that PstI digestion was not complete and there were trace amounts of open circular plasmid remaining. However, MORC-1 washed with high salt only retains the open circular form and not the linearized material.

Figure 5-13. Static MORC-1 foci can block other MORC-1 foci from leaving DNA; entrapment is not nucleotide dependent. (Related to Figures 5 and 6)

(A) Static foci can serve as roadblocks for mobile MORC-1 foci. Representative kymograph

demonstrating that MORC-1 focus movement is blocked by a static MORC-1 focus. Foci were pre-formed on DNA by flowing in 10 nM MORC-1 in the presence of 2 mM ATP, then washed with a high salt buffer (500 mM NaCl, 2 mM MgCl2, 2 mM ATP). The top focus (orange asterisk) contains 10 MORC-1 molecules while the bottom focus (peach asterisk) contains 7. Scale bars are 20 seconds and 10 μm (top), and 1 second and 5 μm (bottom). (B) MORC-1 topological preferences are not nucleotide dependent. 6.4 μM MORC-1 was incubated with 200 ng of various forms of DNA in the presence of 2 mM nucleotide and 4 mM MgCl2, then subjected to a high salt wash (500 mM NaCl). MORC-1 enrichment of open circular or supercoiled forms was similar despite additional nucleotide. Loss of the linear form also occurred regardless of additional nucleotide. Note: the supercoiled stock contained presence of some open circular plasmid. For (C) circular DNA and (D) linear DNA, addition of nucleotide does not significantly influence MORC-1 binding preferences under low or high salt washes. [MORC-1] = 400 nM. (E) Quantitation of retention preferences after high salt washes from panel b, c, and d; densitometry was performed using Fiji, then normalized to circular DNA retention without any additional nucleotide. All circular (supercoiled and open circular) samples were considered together. Error bars represent standard error of the mean. Circular n=3; linear n=2.

Figure 5-14. MORC-1 compacts chromatin.

(A) Distribution of substrate lengths under flow. Mode of DNA substrates (n=301) = 10.03 μm; Mode of chromatin substrates (n=415) = 2.84 μm. Substrates were flowed into the flow cell and subsequently visualized with 3 μM SYTOX Orange. (B) Kymograph depicting MORC-1 mediated compaction of a chromatinized template. Scale bars represent 7 μm and 3 seconds, respectively. [MORC-1] = 75 nM. (C) MORC-1 compacts chromatinized substrates of varying lengths. [MORC-1] = 75 nM was flowed into the flow cell and compaction was allowed to

proceed. Upper panel; scatter plot of individual trajectories plotted by the initial and final lengths. Bottom panel; projection of the scatter plot, binning the trajectories by final DNA length.

Figure 5-15. Model of MORC-1 compaction

Supplemental Video and File guide

Video S1 | MORC-1 compacts DNA.

Upon addition of MORC-1 to the flow cell, quantum dot end labeled lambda DNA travels towards the DNA tether point. [MORC-1] = 60 nM, [ATP] = 2 mM, Flow direction: upward, Movie speed: 4x real-time.

Video S2 | MORC-1 does not extrude DNA loops.

Both ends of bacteriophage lambda DNA were tethered on the microfluidic flow cell with slack. Subsequent addition of 80 nM MORC-1 and 2 mM ATP led to DNA compaction. While the two DNAs became taut, no extruded loops were observed. Dimension: 18.7 μm x 8.3 μm. Movie speed: 6x real-time.

Video S3 | MORC-1 forms foci on DNA.

Cy3-labeled MORC-1 foci appear on a flow stretched DNA and grow in size over time. As DNA compaction proceeds, they eventually merge into one large focus. [Cy3-MORC-1] = 40 nM, [ATP] = 2 mM, Flow direction: upward, Movie speed: 4x real-time.

Video S4 | Formation and growth of MORC-1 foci.

Diffusing MORC-1 molecules can form static foci. Collisions between sliding molecules and static foci lead to apparent growth of foci, although free MORC-1 molecules are also likely recruited from solution. Note that collisions among foci are also involved in the growth of foci. [Cy3-MORC-1] = 10 nM, [ATP] = 2 mM, Flow direction: upward, Speed: 20 frames per second = 4x-real time.

Video S5 | Static MORC-1 foci are dislodged upon high salt wash.

Cy3-labeled MORC-1 foci were pre-formed on flow stretched DNA, then subjected to a high salt wash (500 mM NaCl, 2 mM $MgCl_2$) in the presence of 2 mM ATP. The high salt wash causes static MORC-1 foci to quickly move along the DNA in the flow direction and appears to cause them to slide off the DNA. Flow stretched DNAs are not quantum dot labeled in this example. [Cy3-MORC-1] = 10 nM, [ATP] = 2 mM, Flow direction: upward, Movie speed: 3x real-time.

Video S6 | MORC-1 foci are retained on DNA end labeled with a QDot. As in Video S5 MORC-1 foci were pre-formed on DNA and then subjected to high salt wash**.** The presence of a quantum dot at the DNA ends blocks dissociation of MORC-1 foci. Note that the quantum dot is not visible in this spectral channel. Movie speed: Flow stretched DNAs are quantum dot labeled in this example. [Cy3-MORC-1] = 5 nM, [ATP] = 2 mM, Flow direction: upward, 20 frames per second = 4x real-time.

Video S7 | MORC-1 foci are not retained on DNA without QDots. Pre-formed foci slide along and dissociate from DNA upon high salt wash. [Cy3-MORC-1] = 10 nM, [ATP] = 2 mM, Flow direction: upward, Flow stretched DNAs are not quantum dot labeled in this example. 20 frames per second = 4x-real time.

Video S8 | Static MORC-1 foci act as a roadblock to diffusing foci.

In this movie, a static MORC-1 focus is capable of preventing a mobile focus from dissociating from the free end of DNA. Flow stretched DNAs are not quantum dot labeled in this example.

[Cy3-MORC-1] = 10 nM, [ATP] = 2 mM, Flow direction: upward, 20 frames per second = 4x-real time.

Supplemental Table | Raw analysis data for MORC-1 compaction rates

The Excel document contains raw analysis data for MORC-1 compaction rates. Each tab represents the MORC-1 concentration, and it contains results both in the absence and presence of ATP. In the Summary tab, the information on the compaction rate fold increases in the presence of ATP is included as well.

Methods:

Cloning and Purification of 8x-His-MBP-CeMORC

C. elegans MORC-1 coding sequence (CDS) was cloned into the pRSF vector with a 8xHis MBP tag at the N terminus using In-Fusion cloning (Clontech 638920). Plasmid was transformed into BL-21 (DE3) Rosetta2 cells (Novagen) and grown at 37°C, induced with 0.1 mM isopropyl β-D-1-thiogalactopyranoside and moved to 16°C for overnight expression. Cells were harvested then lysed by sonication in 20 mM Tris-Cl pH 7.5, 500 mM NaCl, 10 mM imidazole, 5 mM DTT, and protease inhibitors. Lysate was clarified by centrifugation and supernatant was applied to Ni-NTA agarose (Qiagen). After binding, resin was washed with 20 mM Tris-Cl pH 7.5, 500 mM NaCl, 30 mM imidazole, 5 mM DTT, then treated with at least 2 units of apyrase (NEB M0398L) prior to elution except for the experiment shown in Supplementary Fig. 2. Protein was eluted from the resin then dialyzed against 20 mM Tris pH 7.5, 150 mM NaCl, 5mM DTT, 10% glycerol. Sample was then subjected to size exclusion chromatography in dialysis buffer. Fractions were pooled and concentrated in an Amicon Ultra-15 (Millipore) for storage.

Gel shift assays

Probe was 32P labeled with T4 PNK (NEB) and used at a final concentration of 0.45 nM unless otherwise indicated. Protein and probe was incubated at 23°C for 30 minutes in 20 mM Tris pH 7.5, 50 mM NaCl, 1 mM DTT, 1 mM MgCl2, .01 mg/ml BSA, then resolved on a 3% 0.5X TBE agarose gel and visualized by exposure to a phosphorscreen unless otherwise indicated. All images were collected with a Typhoon Imager 9000 (GE Amersham). For the ladder gel shift, a 100 bp double stranded DNA ladder (NEB 3231S) was used as substrate, and visualized by SyBR gold staining (Invitrogen S11494).

Native mass spectrometry

Sample was processed and analyzed as previously described (Li et al., 2016 [See Chapter 4]).

DNA substrate preparation for single-molecule experiments

DNA substrates were prepared as previously described13. Briefly, to prepare a DNA substrate with a biotin at one end and a quantum dot at the other end, we annealed a biotinylated BL2-oligo (5'-ggg cgg cga cct/BioTEG/-3') and an oligo with digoxigenin (5'-agg tcg ccg ccc aaa aaa aaa aaa /Digoxigenin/-3') to each cos site of genomic bacteriophage λ-DNA (NEB). The biotin at one end of the substrate was used for surface-tethering of the DNA to a neutravidin coated coverslip while digoxigenin was used for labeling the DNA with anti-digoxigenin antibody-conjugated quantum dots (Thermo Fisher). The DNA substrate used in DNA motion capture experiments was constructed by annealing the BL1 oligo (5'-agg tcg ccg ccc/BiotinTEG/-3') to λ-DNA.

Single-molecule flow-stretching assays

Coverglass passivation and flow cell construction and were performed as previously described(Kim and Loparo, 2016). For this study coverslips were functionalized with 4% biotinylated PEG. A home-built total internal reflection fluorescence (TIRF) microscope was used as described in Graham et al., (Graham et al., 2017). DNA flow stretching experiments were performed by first preincubating the λ-DNA substrate with a biotin and a digoxigenin at each end with anti-digoxigenin antibody-conjugated quantum dot at room temperature. 0.25 mg/mL neutravidin was applied to the flow cell and unbound neutravidin was washed from the flow cell after 5~10 minutes by EcoRI binding buffer (EBB buffer: 10 mM Tris, pH 8.0, 150

mM NaCl, 10 mM MgCl2) that contains 0.2 mg/mL BSA. Quantum-dot labeled biotinylated λ-DNA was flowed into the flow cell at a rate of 20-30 μL/min, and unbound excess DNA was washed away using MORC-1 sample buffer (20 mM Tris, pH 7.5, 150 mM NaCl, 2 mM MgCl2). MORC-1 stock was diluted to the desired concentration by adding MORC-1 sample buffer either in the absence or presence of 2 mM ATP, and then flowed into the flow cell at a rate of 50 μL/min. The movies of flow-stretching assays were recorded using MicroManager (Edelstein et al., 2014). DNA compaction trajectories were determined by fitting the positions of quantum dots using Gaussian fitting with custom-written MATLAB (MathWorks) code that has provided in the previous publication (Kim and Loparo, 2016).

DNA motion capture assay

The assay was performed as previously described (Kim and Loparo, 2016; Kim and Loparo, 2018).

Double-tethered DNA assay

A flow cell was constructed by sandwiching a cross-shaped (1.8 mm and 2.5 mm widths, respectively) channel made out of double-sided tape (0.12 mm thickness) between a passivated coverglass and a quartz slide. PE60 tubes were used as inlet and outlet tubing for each channel. 0.25 mg/ml of neutravidin (Thermo Scientific) was added to the flow cell using a gel loading pipet tip. After 6-7 minutes of incubation, EBB buffer (10 mM Tris pH 8.0, 150 mM NaCl, 10 mM MgCl2, 0.2 mg/ml bovine serum albumin) was flowed in using a syringe pump to wash away excess neutravidin. 180 μL of the 400-fold diluted BL1-DNA substrate was flowed in at a rate of 30 μL/min and allowed to incubate for 2 minutes in the absence of flow. Unbound DNAs were then washed away by flowing in 140 μL of MORC sample buffer at a rate of 100 μL/min.

In order to tether the other end of DNA to the flow cell, 380 µL of 100 nM BL2-oligo was flowed in at 30 µL/min, and 120 µL of MORC-1 sample buffer was immediately flowed in at 100 µL/min to remove unbound BL2-oligo. DNA compaction was observed by adding 80 nM MORC-1 and 6 nM SYTOX-Orange to the flow cell using the channel perpendicular to DNA tethering in the presence of 2 mM ATP.

Chromatin assembly and compaction assay

Chromatin compaction assays were imaged on [Robocop]. Biotinylated DNA was chromatinized using the Active Motif chromatin assembly kit (Active Motif 53500). Assembly was checked by MNAse digestion. Substrate was diluted into EBB+BSA and introduced into the flow cell using the parameters described above. 75 nM MORC-1 was diluted into MORC sample buffer as described above supplemented with 2mM ATP, and introduced into the flow cell also using parameters described above, and visualized by 3 nM SyTOX Orange staining. Chromatin compaction movies were recorded and MORC-1 mediated compaction was determined by measuring the change in the flow stretched DNA length using custom-written MATLAB (MathWorks) code (Song et al., 2016).

Circular DNA binding assay

Unless otherwise noted, 2.4 µM protein was incubated for 30 minutes at room temperature with a pBR322 derivative of varying topologies in 20 mM Tris, pH 7.5, 150 mM NaCl, 1mM DTT). Topology was altered by restriction enzyme digest. Protein was precipitated using NiNTA magnetic agarose (Qiagen), then washed 2x with either a high salt (600 mM NaCl) or a low salt buffer (150 mM NaCl). Reactions were quenched with a stop buffer (20 mM Tris pH 8, 2% SDS, 1 mM EDTA) and samples were eluted by proteinase K digestion. For the re-linearization assay,

400 nM MORC-1 was bound to 50 ng total pBR322 pre-modified with Nt.BsmAI. Protein was precipitated as described above, then washed with 20 mM Tris pH 7.5, 500 mM NaCl, 1mM DTT, then buffer exchanged into a Restriction Enzyme (RE) buffer (New England Biosciences Cutsmart Buffer, B7204) and digested for 30 minutes room temperature, then washed again with either a low salt buffer (150 mM NaCl) or a high salt buffer (500 mM NaCl.) For assays with nucleotide, all buffers were supplemented with 4 mM MgCl2. Reactions were loaded on a 1% 1xTAE gel and run overnight at room temperature, then post stained using ethidium bromide, and imaged using a Gel Logic 212 Pro (Carestream). Quantification was performed using Fiji (Schindelin et al., 2012). Intensities were averaged if there were replicates within the experiment, and normalized to the value obtained for open circular.

Protein labeling with fluorescent dyes

Proteins were incubated with NHS-Cy3 or NHS-Cy5 dye (Lumiprobe 11020, 13020) at 4℃. Labeled protein was separated from free dye using Biospin 6 columns (Biorad). Cy labeling efficiency was quantified using a Nanodrop; final concentrations were determined by averaging three readings.

Liquid droplet sample preparation and imaging

Samples were buffer exchanged using Micro Bio-spin P-30 gel columns (Biorad 732-6223) into 20 mM Tris pH 7.5, 150 mM NaCl, 5mM DTT. Concentration was measured three times using a NanoDrop spectrophotometer (Thermo Fisher). Protein was diluted to the 75 mM NaCl. Final amounts: 10% PEG or 500 ng total, or 0.8 nM final, λ-DNA (NEB N3011S) with a final unlabeled protein concentration of 4.3 uM with 1-2% Cy-labeled protein acting as tracer. Final

reaction volume was 20 µL. Samples were imaged using a Zeiss LSM 880 with spectral imaging at 63X magnification. Signals from Cy3 and Cy5 were linear unmixed in Zen 9.0 (Zeiss).

Gonad dissection and immunostaining

Gravid adult C. elegans were dissected in egg buffer (25 mM HEPES pH 7.4, 118 mM NaCl, 48 mM KCl, 2 mM EDTA, 0.5 mM EGTA), containing 15 mM sodium azide and 0.1% Tween-20. Samples were fixed in 1% formaldeyhyde in egg buffer for 10 seconds followed by freeze-cracking in liquid nitrogen, 1 min methanol fixation at -20°C, and three washes in PBST. Slides were then incubated with primary mouse anti-FLAG antibody (1:100, Sigma F1804) in a humid chamber overnight at 4°C or 2 h at room temperature. Slides were washed three times in PBST and then incubated with secondary antibody (1:300, Invitrogen AlexaFluor 555 goat anti-mouse) in a humid chamber for 2 h at room temperature. Slides were washed three times in PBST, stained with 0.5 µg/mL DAPI in PBST for 15 min, washed in PBST, and mounted with Vectashield (Vectorlabs H-1000).

Structured illumination microscopy (SIM) and image processing

SIM was performed with Zeiss ELYRA PS.1 system (Carl Zeiss, Germany) equipped with a 100x/1.46 oil immersion objective and a PCO Edge 4.2 camera. Excitation wavelengths used were 405 nm and 561 nm. Images were acquired after 5 rotations and 5 standard shifts of the grating pattern. Image processing was carried out using Zen 2.1 SP3 software. For wild type, n (mitotic)=11, n (pachytene)=12; For RNAi, n (mitotic)= 3, n (pachytene)=4.

STORM imaging

Immunofluorescence of extruded gonads was performed as previously described with minor modifications (Köhler et al., 2017). Briefly, germline extrusion was performed in egg buffer (25

mM HEPES pH 7.4, 118 mM NaCl, 48 mM KCl, 2 mM EDTA, 0.5 mM EGTA), containing 30 mM sodium azide and 0.1% Tween-20. Samples were fixed in 1% formaldeyhyde in egg buffer for 10 seconds followed by a 1 min methanol fixation at -20°C. Primary mouse anti-FLAG antibody used was 1:100 (Sigma F1804) in normal goat serum and PBST and the secondary antibody was used at 1:100 (Invitrogen AlexaFluor 647 goat anti-mouse) in PBST. All washes and staining were performed in suspension. Germlines were mounted in imaging buffer (Tris•HCl, pH 7.5, containing 100 mM cysteamine, 5% glucose, 0.8 mg/mL glucose oxidase, and 40 μg/mL catalase) and imaged immediately. At least n (mitotic) =8 and n (pachytene)=11 were imaged. Scale bar represents 5 μm.

STORM image processing

STORM was performed with a Zeiss ELYRA PS.1 system (Carl Zeiss, Germany) equipped with a 100x/1.46 oil immersion objective and a PCO Edge 4.2 camera. The 647-nm laser was used to excite fluorescence from antibody-labeled proteins after photobleaching them into the dark state. The 405-nm laser was adjusted during image acquisition. Images were acquired with 43.55 ms exposure time per frame for >13,000 frames. Events that emit fewer than 500 photons or that were not resolved to within 50 nm were discarded. Image processing was performed using Zen 2.1 SP3 software.

Hexanediol treatment and imaging

Gravid adult C. elegans were dissected in egg buffer (118 mM NaCl, 48 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 25 mM HEPES [pH 7.4]), containing 15 mM sodium azide and 0.1% Tween-20. Samples were incubated briefly in either 1x egg buffer or 5% 1,6-hexanediol in egg buffer then fixed in 1% formaldeyhyde in egg buffer for 10 sec followed by a 1 min methanol fixation at -

20°C. Primary mouse anti-FLAG antibody (Sigma F1804) was used at 1:100 and rabbit anti-CSR-1 was used at 1:200 in normal goat serum and PBST. The secondary antibody was used at 1:300 (Invitrogen AlexaFluor 555 goat anti-mouse and 488 goat anti-rabbit) in PBST. All washes and staining were performed in suspension. Germlines were stained with DAPI (0.5 μg/mL) then mounted with Vectashield (Vectorlabs H-1000). Images were acquired on a Zeiss LSM700 confocal microscope at 63x magnification. Image processing was performed using Zen SP5 software. Experiments were performed in biological triplicate. A total of n=18 gonads were imaged in 0% hexanediol, and n=19 gonads were imaged in 5% hexanediol.

References

Alberti, S., Gladfelter, A., and Mittag, T. (2019). Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. Cell 176, 419-434.

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al., (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell 175, 1842-1855.e1816.

Brower-Toland, B.D., Smith, C.L., Yeh, R.C., Lis, J.T., Peterson, C.L., and Wang, M.D. (2002). Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA. Proceedings of the National Academy of Sciences 99, 1960-1965.

Corbett, K.D., and Berger, J.M. (2005). Structural Dissection of ATP Turnover in the Prototypical GHL ATPase TopoVI. Structure 13, 873-882.

Cuylen, S., Metz, J., and Haering, C.H. (2011). Condensin structures chromosomal DNA through topological links. Nature Structural & Molecular Biology 18, 894-901.

Dong, W., Vannozzi, A., Chen, F., Hu, Y., Chen, Z., Zhang, L., and Pritham, E. (2018). MORC Domain Definition and Evolutionary Analysis of the MORC Gene Family in Green Plants. Genome Biology and Evolution 10, 1730-1744.

Douse, C.H., Bloor, S., Liu, Y., Shamin, M., Tchasovnikarova, I.A., Timms, R.T., Lehner, P.J., and Modis, Y. (2018). Neuropathic MORC2 mutations perturb GHKL ATPase dimerization dynamics and epigenetic silencing by multiple structural mechanisms. Nature communications 9.

Dutta, R., and Inouye, M. (2000). GHKL, an emergent ATPase/kinase superfamily. Trends in Biochemical Sciences 25, 24-28.

Edelstein, A.D., Tsuchida, M.A., Amodaj, N., Pinkard, H., Vale, R.D., and Stuurman, N. (2014). Advanced methods of microscope control using μManager software. Journal of Biological Methods 1.

Ganji, M., Shaltiel, I.A., Bisht, S., Kim, E., Kalichava, A., Haering, C.H., and Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. Science 360, 102-105.

Graham, T.G., Wang, X., Song, D., Etson, C.M., van Oijen, A.M., Rudner, D.Z., and Loparo, J.J. (2014). ParB spreading requires DNA bridging. Genes & development 28, 1228-1238.

Graham, T.G.W., Walter, J.C., and Loparo, J.J. (2017). Ensemble and Single-Molecule Analysis of Non-Homologous End Joining in Frog Egg Extracts. In DNA Repair Enzymes: Cell, Molecular, and Chemical Biology, B.F. Eichman, ed. (Academic Press), pp. 233-270.

Harris, C.J., Husmann, D., Liu, W., Kasmi, F.E., Wang, H., Papikian, A., Pastor, W.A., Moissiard, G., Vashisht, A.A., Dangl, J.L., et al., (2016). Arabidopsis AtMORC4 and AtMORC7 Form Nuclear Bodies and Repress a Large Number of Protein-Coding Genes. PLOS Genetics 12, e1005998.

Inoue, N., Hess, K.D., Moreadith, R.W., Richardson, L.L., Handel, M.A., Watson, M.L., and Zinn, A.R. (1999). New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. Human Molecular Genetics 8, 1201-1207.

Iyer, L.M., Abhiman, S., and Aravind, L. (2008). MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. Biology direct 3, 8.

Kang, H.-G., Woo Choi, H., von Einem, S., Manosalva, P., Ehlers, K., Liu, P.-P., Buxa, S.V., Moreau, M., Mang, H.-G., Kachroo, P., et al., (2012). CRT1 is a nuclear-translocated MORC endonuclease that participates in multiple levels of plant immunity. Nature communications 3, 1297.

Kanno, T., Berta, Davide G., and Sjögren, C. (2015). The Smc5/6 Complex Is an ATP-Dependent Intermolecular DNA Linker. Cell Reports 12, 1471-1482.

Kim, H., and Loparo, J.J. (2016). Multistep assembly of DNA condensation clusters by SMC. Nature communications 7, 10200.

Kim, H., and Loparo, J.J. (2018). Observing Bacterial Chromatin Protein-DNA Interactions by Combining DNA Flow-Stretching with Single-Molecule Imaging. In Bacterial Chromatin, pp. 277-299.

Köhler, S., Wojcik, M., Xu, K., and Dernburg, A.F. (2017). Superresolution microscopy reveals the three-dimensional organization of meiotic chromosome axes in intact Caenorhabditis elegans tissue. Proc Natl Acad Sci USA 114, E4734-E4743.

Kroschwald, S., Maharana, S., Mateju, D., Malinovska, L., Nüske, E., Poser, I., Richter, D., and Alberti, S. (2015). Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules. eLife 4.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. Nature 547, 236-240.

Larson, A.G., and Narlikar, G.J. (2018). The Role of Phase Separation in Heterochromatin Formation, Function, and Regulation. Biochemistry 57, 2540-2548.

Li, S., Yen, L., Pastor, W.A., Johnston, J.B., Du, J., Shew, C.J., Liu, W., Ho, J., Stender, B., Clark, A.T., et al., (2016). Mouse MORC3 is a GHKL ATPase that localizes to H3K4me3 marked chromatin. Proceedings of the National Academy of Sciences of the United States of America 113, E5108-5116.

Lin, Y., Protter, D.S.W., Rosen, M.K., and Parker, R. (2015). Formation and Maturation of Phase Separated Liquid Droplets by RNA Binding Proteins. Molecular cell 60, 208-219.

Mimura, Y., Takahashi, K., Kawata, K., Akazawa, T., and Inoue, N. (2010). Two-step colocalization of MORC3 with PML nuclear bodies. Journal of cell science 123, 2014-2024.

Moissiard, G., Bischof, S., Husmann, D., Pastor, W.A., Hale, C.J., Yen, L., Stroud, H., Papikian, A., Vashisht, A.A., Wohlschlegel, J.A., et al., (2014). Transcriptional gene silencing by Arabidopsis microrchidia homologues involves the formation of heteromers. Proceedings of the National Academy of Sciences of the United States of America 111, 7474-7479.

Moissiard, G., Cokus, S.J., Cary, J., Feng, S., Billi, A.C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B.R., McCord, R.P., et al., (2012). MORC family ATPases required for heterochromatin condensation and gene silencing. Science 336, 1448-1451.

Murayama, Y., and Uhlmann, F. (2013). Biochemical reconstitution of topological DNA binding by the cohesin ring. Nature 505, 367-371.

Pastor, W.A., Stroud, H., Nee, K., Liu, W., Pezic, D., Manakov, S., Lee, S.A., Moissiard, G., Zamudio, N., Bourc'his, D., et al., (2014). MORC1 represses transposable elements in the mouse male germline. Nature communications 5, 5795.

Protter, D.S.W., Rao, B.S., Van Treeck, B., Lin, Y., Mizoue, L., Rosen, M.K., and Parker, R. (2018). Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly. Cell reports 22, 1401-1412.

Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., et al., (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science 361.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al., (2012). Fiji: an open-source platform for biological-image analysis. Nature Methods 9, 676-682.

Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. Science 357.

Shin, Y., Chang, Y.-C., Lee, D.S.W., Berry, J., Sanders, D.W., Ronceray, P., Wingreen, N.S., Haataja, M., and Brangwynne, C.P. (2018). Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. Cell 175, 1481-1491.e1413.

Song, D., Graham, T.G.W., and Loparo, J.J. (2016). A general approach to visualize protein binding and DNA conformation without protein labelling. Nature communications 7.

Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. Nature 547, 241-245.

Takahashi, K., Yoshida, N., Murakami, N., Kawata, K., Ishizaki, H., Tanaka-Okamoto, M., Miyoshi, J., Zinn, A.R., Shime, H., and Inoue, N. (2007). Dynamic regulation of p53 subnuclear localization and senescence by MORC3. Molecular biology of the cell 18, 1701-1709.

Tchasovnikarova, I.A., Timms, R.T., Douse, C.H., Roberts, R.C., Dougan, G., Kingston, R.E., Modis, Y., and Lehner, P.J. (2017). Hyper-activation of HUSH complex function by Charcot-Marie-Tooth disease mutation in MORC2. Nat Genet 49, 1035-1044.

Updike, D.L., Hachey, S.J., Kreher, J., and Strome, S. (2011). P granules extend the nuclear pore complex environment in theC. elegansgerm line. The Journal of Cell Biology 192, 939-948.

Weiser, N.E., Yang, D.X., Feng, S., Kalinava, N., Brown, K.C., Khanikar, J., Freeberg, M.A., Snyder, M.J., Csankovszki, G., Chan, R.C., et al., (2017). MORC-1 Integrates Nuclear RNAi and Transgenerational Chromatin Architecture to Promote Germline Immortality. Developmental cell 41, 408-423.e407.

Wheeler, J.R., Matheny, T., Jain, S., Abrisch, R., and Parker, R. (2016). Distinct stages in stress granule assembly and disassembly. eLife 5.