Chromatin Organization in Giant Viruses: Revealing Non-Canonical Nucleosome Features

by

Chelsea M. Toner

B.S., Texas State University, 2018

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

2024

Doctoral Committee Members:

Karolin Luger, Advisor

Jen Kugel

Amy Palmer

Jenny Knight

John Rinn

Toner, Chelsea M. (Ph.D., Biochemistry)

Chromatin Organization in Giant Viruses: Revealing Non-Canonical Nucleosome Features

Thesis directed by Professor Karolin Luger

Abstract

The packaging of genomic DNA into defined nucleosomes has long been viewed as a universal and ancestral feature of the eukaryotic domain of life. However, the identification of "minimalist" histones in archaea in the nineties, and the recent discovery of putative histones in some giant viruses has challenged this dogma. These viruses harbor giant genomes, ranging from 2.8 Mb to 300 kb of double-stranded DNA, and encode a range of proteins involved in replication, transcription, and translation; machinery that most viruses typically "hijack" from the host during infection. Most notably, some giant virus families encode distant homologs of the four eukaryotic histones, in many cases as fused polypeptides. These unique virally encoded histones invoke curiosity into their role of viral genome regulation and question the current theories of eukaryogenesis. Here, I contributed to determining the cryo-EM structure of Melbournevirus histone-encoded histone doublets (H2B-H2A and H4-H3) assembled into nucleosome-like particles on DNA. I also independently solved the structure of the Medusavirus nucleosome-like particle and demonstrated the ability of these viral histones to assemble into higher order chromatin. Lastly, with Medusavirus relying solely on the host nucleus for DNA replication, I probed the potential for the formation of hybrid viral and eukaryotic nucleosomes. These results show that viral nucleosomes contribute to our understanding of their potential proto-eukaryotic provenance and raise questions regarding the 'repurposing' of nucleosomes to promote viral fitness.

Dedication

To my beloved mother, Lynne Clopton, whose radiant presence left us too early.

Acknowledgements

My time in the Biochemistry department at CU Boulder has given me many experiences, from remembering to remove the gosh-darn gel tape when running pre-cast gels to learning the intricacies of processing cryo-EM datasets. Never in my life did I think I would have had the wonderful opportunities presented to me here and I am forever thankful that I was able to achieve the dreams of my childhood. Endless thanks to Karolin, who was the facilitator of all my experiences. Without Karolin, I would not think about science as critically as I do or have had the chance to study giant viral histones. Such a unique little project she trusted me with, guided me through, and helped me complete.

While I knew that my journey through graduate school would be faced with difficulty, I did not realize how many outstanding colleagues would mentor and inspire me along the way. Thanks to Dr. Samuel Bowerman, who really saw just how hard I worked and did everything to support me. Your words of wisdom radiate with me always and you will forever be a true friend. Thanks to Briana Aboulache, who continuously encouraged me to follow my dreams of teaching. I couldn't have made it as S.C.O.P.E. president without you and I can't wait to watch you flourish into an amazing educator. And lastly, thanks to Dr. Nicole Hoitsma, who literally dropped into my life at just the right time. From burning phylogenetic trees to inspirational calendar quotes, you carried me through to the finish line with endless support and love.

Even when I wasn't working, I was fortunate to be given continual support in the past six years from my family. Thank you to Will, my life partner, for everything you have done and will do. You picked me up and put me back together through each and every difficulty I faced during this journey. I could not have done this without you, and I can't wait to see where life takes us and the pups next. Thank you to Dr. Judy Easton, my adopted mother, who inspired me to go down this path. You are the shining example of the educator I only hope to be in the future. From leaving me daily morning voice messages to gifting me the best books, you've filled the hole in my heart from my late mother with abundant love and support through it all.

Lastly, I must thank all of my friends from all around the world. I am so fortunate to be surrounded by so many wonderful individuals who have inspired me to be a well-rounded human and the caring color guard educator I am today. Whether our paths have crossed through color guard, science, or just life, each and every single one of you is special to me. The journey to get to this point was not easy and the endless laughter, 'you've got it', and ridiculous antics were what reminded me to let go so I could grab hold of tomorrow's possibility.

Contents

Chapter

Tables

Table

Figures

Figure

Chapter 1

Introduction

1.1 Who are 'giant viruses' and whom do they infect?

Viruses are commonly described as 'organisms' that maintain a mischievous purpose by hijacking their host's machinery for their own replication and continuation of their 'lives'. However, through years of research, it has been revealed that our understanding of viruses expands beyond what the initial broad definition entailed. Depending on their genomic makeup (DNA or RNA; single or double stranded) some viruses encode an enzymatic repertoire necessary for their own DNA replication and/or transcription, while others depend exclusively on the host's machinery. However, during a 1992 pneumonia outbreak, researchers noticed what they believed to be gram-positive bacterium within isolated Amoebae[1]. It was not until 2003 that researchers realized they had in fact isolated the Mimivirus, named for its ability to "mimic" a microbe in size, limited by their previous assumptions of virus particle sizes[2]. The Mimivirus maintained a uniquely large 1.2 Mb DNA genome and doubled all other known viruses in size with a particle diameter of 400 – 500 nm[3]. The discovery of Mimivirus introduced the new classification of 'giant DNA viruses' and sparked the identification of many giant viruses since then. These giant viruses have been found to encode most of their own proteins required for replication and transcription likely due to sheer

genome size[4]. With the broad coding potential of their exceptional genome size, including the presence of hallmark genes that define cellular life, the giant viruses bring a myriad of new and intriguing questions and hypotheses[5–7].

Giant viruses are classified as Nucleocytoplasmic large DNA viruses (NCLDV) belonging to the *Nucleocytoviricota* phylum of viruses. The International Committee on Taxonomy of Viruses (ICTV) currently classifies seven families within the NCLDV: *Ascoviridae, Asfarviridae, Iridoviridae, Marseilleviridae, Mimiviridae, Phycodnaviridae,* and *Poxviridae.* 21 proposed NCLDV remain unclassified and 32 new families proposed[8]. Phylogeny is complicated by the rapid discovery of new members of this phylum, either in the form of viral particles or as metagenomes[6,9]. However, various phylogenetic analyses have informally assigned these unclassified NCLDV to additional proposed families (Table 1.1). The universal taxonomic framework for viruses, including NCLDV, is based on the phylogenetic analysis of conserved genes across viral groups. These genes, referred to as virus hallmark genes (VHGs), are responsible for key functions in virus replication and virion morphogenesis. As the genomic diversity of NCDLV parallels their continuous discovery, and the number of conserved genes amongst them dwindles, the ICTV currently establishes the *Nucleocytoviricota* phylum exclusively upon the basis of the double jelly-roll major capsid proteins (DJR-MCPs). The main features of the NCLDV phylogeny have remained remarkably stable over several decades, even with varied choices of markers of phylogeny construction, suggesting that the major branches of the NCLDV are already firmly defined[9]. Furthermore, the observation that genome size is highly variable among families and does not appear to be correlated with host taxonomy highlights the complexity of both evolutionary and replication aspects of NCDLV survival.

Table 1.1 An adapted and updated list of ICTV classified and unclassified NCLDVs identified to date[10–19].

Virus Class	Virus Order	Virus Family	Virus Genus	Genome size (kb)	DNA replication (Location)	Host Taxonomy (Range)
Classified NCLDVs						
Megaviricetes	*Algavirales*	*Phycodnaviridae*	*Chlorovirus* *Coccolithovirus* *Phaeovirus* *Prymnesiovirus* *Raphidovirus*	155-474	Nucleus and Cytoplasm	Alveolata, Chlorophyta, Haptophyta, Stramenopiles
	Imitervirales	*Mimiviridae*	*Cafeteriavirus* *Mimivirus*	617-1259	Cytoplasm	Amoeba, Stramenopiles
	Pimascovirale	*Ascoviridae*	*Ascovirus* *Toursvirus*	119-186	Nucleus and Cytoplasm	Lepidoptera
		Iridoviridae	*Lymphocystivirus* *Megalogcytivirus* *Ranavirus* *Chloriridovirus* *Daphniairidovirus* *Decapodiridovirus* *Iridovirus*	106-220	Nucleus and Cytoplasm	Arthropods, Fish, Amphilibia
		Marseilleviridae	*Marseillevirus*	347-403	Nucleus and Cytoplasm	Amoeba
		Mamonoviridae (Proposed)	*Medusavirus medusae* *Medusavirus sthenus*	363-381	Nucleus	Amoeba
Pokkesviricetes	*Asfuvirales*	*Asfarviridae*	*Asfivirus*	170	Cytoplasm	Swines, Humans, Arthopods
	Chitovirales	*Poxviridae*	*Avipoxvirus* *Capripoxvirus* *Centapoxvirus* *Cervidpoxvirus* *Leporipoxvirus* *Macropopxvirus* *Molluscipoxvirus* *Mustelpoxvirus* *Orthopoxvirus* *Oryzopoxvirus* *Parapoxvirus* *Salmonpoxvirus* *Sciuripoxvirus* *Suipoxvirus* *Vespertilionpoxvirus* *Yatapoxvirus* *Alphaentomopoxvirus* *Betaentomopoxvirus* *Deltaentomopoxvirus* *Gammaentomopoxvirus*	134-360	Cytoplasm	Arthropods, Vertebrates

Unclassified NCLDVs					
N/A	N/A	*Pandoraviridae*	1909-2474	Cytoplasm	Amoeba
N/A	N/A	*Pithovirius*	610	Cytoplasm	Amoeba
N/A	N/A	*Mininucleoviridae*	70-74	N/A	Crustacea
N/A	N/A	*Cedratvirus*	589	N/A	Amoeba
N/A	N/A	*Choanovirus*	370-670	N/A	Choanoflagellates
N/A	N/A	*Faustovirus*	460	Cytoplasm	Amoeba
N/A	N/A	*Kamoebavirus*	351	N/A	Amoeba
N/A	N/A	*Klosneuvirus*	1385-1570	N/A	Amoeba
N/A	N/A	*Meelsvirus*	N/A	N/A	*Chaetognatha*
N/A	N/A	*Mollivirus*	652	Nucleus and Cytoplasm	Amoeba
N/A	N/A	*Namao virus*	109-196	N/A	Lake Strugeon
N/A	N/A	*Orpheovirus*	1473		*Vermamoeba vermiformis*
N/A	N/A	*Pacmanvirus*	395	N/A	Amoeba
N/A	N/A	*Plantanovirus*	N/A	N/A	*Saccamoeba lacustris*
N/A	N/A	*Sissivirus*	N/A	N/A	Amoeba
N/A	N/A	*Tupanvirus*	1440-1510	Cytoplasm	Amoeba
N/A	N/A	*Urceolovirus*	N/A	N/A	Amoeba
N/A	N/A	*Usurpativirus*	N/A	N/A	Amoeba
N/A	N/A	*Clandestinovirus*	582	N/A	*Vermamoeba vermiformis*
N/A	N/A	*Yasminevirus*	2100	N/A	*Vermamoeba vermiformis*

NCLDV infect a broad range of mostly lower eukaryotes, contributing to their fascinating diversity. The most common host are amoebae, with other unicellular hosts being dinoflagellates (algae) and choanoflagellates. Within amoeba, giant viruses infect the genus *Acanthamoeba spp.* and to date, this is the main genus used for culture support in the isolation of NCLDV from amoebae. In fact, Mimivirus was isolated from *Acanthamoeba polyphaga* while *Acanthamoeba castellanii* enabled the isolation of several other NCDLV, including Pandoraviruses, Pithovirus sibericum, Mollivirus sibericum, and Medusavirus[20,21]. It has been hypothesized that *Acanthamoeba spp.* serve as the natural hosts for NCLDV due to their ability to ingest any particle with a size of 0.5 µm or greater and mistaking NCLDV as food; perhaps explaining the prevalence of NCLDV isolation within these organisms[7,9]. Intriguingly, a few specific NCLDV have been shown

to infect arrow worms, lake sturgeon (marine fish), crustacea, vertebrates, and arthropods; which implies they have a broader variety of natural hosts than initially thought (Table 1.1). The Asfivirus and Poxvirius, part of the Pokkeviricete class of NCLDV, are two of the few known giant viruses specifically known to infect vertebrates such as pigs and humans (Table 1.1). This highlights the unique infectious potential of giant viruses and raises several questions such as: are unicellular hosts preferred; did host range expand upon viral genome expansion (or vice versa); and how do NCLDV adopt new hosts? Perhaps it is not so surprising that NCLDV infects higher order, multicellular eukaryotes given their relatively large particle sizes, however, the broad infectivity of NCLDV seems to elicit more curiosity than clarity.

The prevailing unifying features of the NCLDVs are their large pseudo-icosahedral capsids, greater than 0.2 μM, and large double-stranded DNA genomes, ranging from approximately 100 kb to more than 2.5 Mb (Table 1.1). With the largest of these viruses exceeding the size of many bacteria and archaea in both particle and genome size, the NCLDV shifts our notion of typical viral size and complexity. One of the most striking findings of NCLDV was that their proteomes contained various homologs of eukaryotic hallmarks that had not previously been detected in viruses. Among these were proteins involved in translation, RNA maturation, proteostasis, and metabolism. Although some giant viruses maintain many of these eukaryotic homologs, each NCLDV has its own unique functional composite of surprising eukaryotic homologs[22]. For example, a handful of NLDCV families encode DNA-packaging proteins such as histones, some as fused polypeptides[23]. With our understanding of viruses constantly evolving, giant viruses represent a new chapter in our working definition of the capabilities of a virus, by opening a door to understanding potential regulation of the viral genome by a virus itself.

1.2 Eukaryotic chromatin organization

Chromatin organization is a fundamental and essential component of the eukaryotic nucleus. Eukaryotic DNA must be highly compacted to fit into the nucleus and this is accomplished by wrapping DNA around histone proteins. Histones are small, basic, highly abundant proteins that help organize and protect the genome[24]. There are four canonical histones (H2A, H2B, H3, and H4) that come together as two dimers (H2A/H2B) and one tetramer $(H3\text{-}H4)_2$ to form a octamer, which wraps DNA into discrete ~150 bp disks (**Figure 1.1**)[25,26]. These disks, termed nucleosomes, are quite dynamic to allow for varying accessibility, as access to the wrapped DNA is necessary for essential cellular processes, such as replication and transcription. Additionally, there are numerous histone variants that can swapped out in and out of nucleosomes in different contexts to facilitate different functions[27]. Histone regulatory functions can be changed through post-translational modifications of their histone tails. By modulating tail modifications, histones can impact gene accessibility for cellular machinery[28,29]. As such, eukaryotes have evolved a suite of enzymes that can modify, remodel, and chaperone histones.

Histones availability and fate are determined by their association with histone chaperone complexes which assist in histone folding, nuclear import, and genomic localization. Histone chaperones can even act as reservoirs for histones and assist in the assembly of the nucleosome[30,31]. Once assembled into nucleosomes, chromatin remodelers use the energy from ATP hydrolysis to exchange histones for their variants or relocate them entirely to different positions on DNA. As such, chromatin remodelers affect many essential chromosomal processes such as gene expression, replication, and DNA damage repair[32]. In addition, cellular processes are

signaled predominately through acetyl and methyl modifications on histones[33]. Together, this is a tightly woven regulatory system which allows histones to assemble into nucleosomes that package our genomes in cells, thus controlling gene expression with high precision.

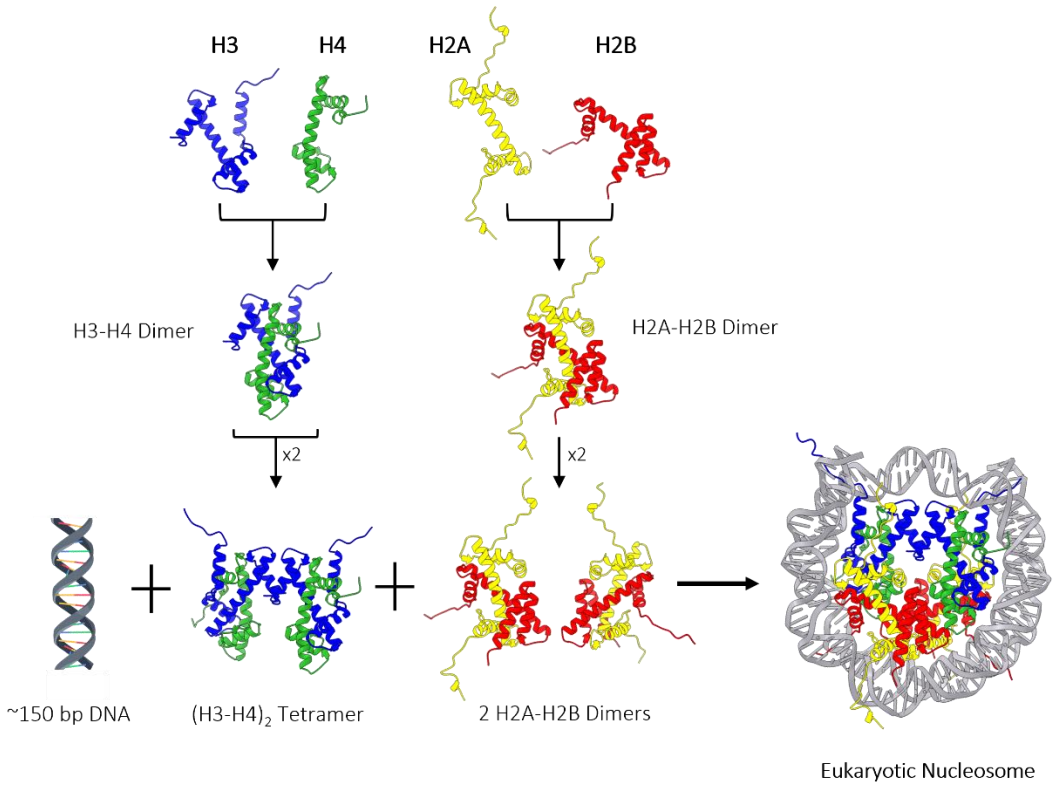


Figure 1.1 Histones assemble into octamers that wrap DNA.

Schematic representation of eukaryotic core histone assembly into the nucleosome. Histones and nucleosome structure are represented by PBD ID: 1AOI.

Nucleosomes are further compacted by H1 linker histones which associate with DNA between adjacent nucleosomes and promote inter-nucleosome interactions[34]. This stabilization by H1 contributes to chromatin compaction (heterochromatin), transcription regulation, and spacing of nucleosomes. Unlike other histones, H1 does not maintain the characteristic histone fold domain (α1–L1–α2–L2–α3) but instead maintains a winged helix DNA binding domain and an essential C-terminal unstructured tail that interacts with linker DNA between nucleosomes[35]. The

H1 linker histones are more variable that the four core histones, but their function in DNA regulation and resulting chromatin structure is well established[36].

1.3 Histones go viral

Unique amongst all viruses, homologs of histone genes have been discovered in a handful of NCLDV. Some of these giant viruses, such as the Bracovirus, only encode one histone gene (H4) while others within *Marseilleviridae*, encode two fused histone doublets (H2B-H2A and H4-H3)[37–40]. The diversity of these virally encoded histones is constantly expanding, and is also characterized by increasing instances of fused histones as doublets, triplets and even quadruplets[41,42]. Uniquely, the Medusavirus represents an instance of an NCLDV encoding each of the four quintessential core histones (H2A, H2B, H3, and H4) along with the H1 linker histone[21]. Currently there are 72 different viral histones across 40 different viral genomes[23]. The evolving phylogenetics consistently places these viral histones between eukaryotic histones and rooted archaeal histones of each clade[17,21,23]. This suggests that these histones represent an intermediate branching point and may have had a role in the evolution of eukaryotic histones.

1.4 Giant viruses and nucleosome evolution

One defining feature of the eukaryotic cell is the presence of the nucleus, however its evolutionary origin is still debated. The commonly accepted hypothesis is that the first ancient eukaryotic nucleus emerged from the fusion of an archaeal cell and bacterium, while the less supported viral eukaryogenesis hypothesis initially suggested the eukaryotic nucleus originated

from the virions of NCDLV (**Figure 1.2**)[43,44]. However, upon the discovery that the Mimivirus viral factories were similar in sizing to the nucleus and encoded a 7-methyl guanylate (m7G) capping apparatus along with eIF4E; the viral eukaryogenesis hypothesis transformed to propose the origin

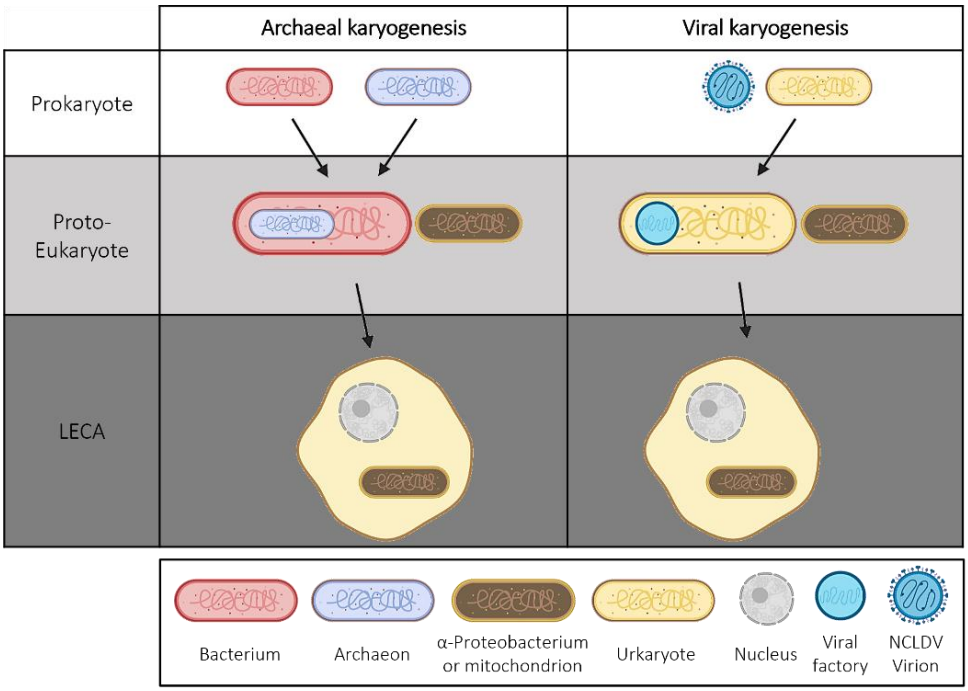


Figure 1.2 Models of eukaryogenesis.

Schematic representation of the development of LECA from a bacterial host (Archaeal karyogenesis) or a Urkaryotic host infected with a giant virus (Viral Karyogenesis). Figure adapted from Box 1: Models of eukaryogenesis; Heinkoff, et al., 2019[45].

of the nucleus originating from the viral factory as opposed to the virion itself[45,46]. As a priming player in mRNA splicing, nuclear export, and cytoplasmic translation, it is surprising that the m7G is generally absent from archaea but present in Mimivirus. This suggests a variety of scenarios, with the viral eukaryogenesis hypothesis suggesting that the eukaryotic nucleus descended from an ancient NCLDV that infected an archaeon with an enslaved alpha-proteobacteria (the precursor of the mitochondrion)[6]. One scenario specifically suggests that the Medusavirus, which encodes

eukaryotic homologs such as encodes Ran, DNA polymerase, and histones; represents a hypothetical ancestral NCLDV whose infection promoted the proto-eukaryotic host to develop a nucleus as a defense mechanism[47].

With the presence of histones in giant viruses and the suggestion that NCLDV may provide insight into the origin of the nucleus, the role of these viral histones in evolution comes into focus. Viral histones demonstrate a range in sequence similarly to their eukaryotic host histones, from looking remarkably like eukaryotic histones to demonstrating fused doublets, triplets, or quadruplets[23]. Phylogenetics analyses of individual histone moieties of histone doublets suggest that these viral histones may have diverged prior to the last eukaryotic common ancestor (LECA)[21,48–50]. This would imply that viral histone fusions existed during the time of proto-eukaryotic organisms and may have contributed to the strong pairing of histones seen in eukaryotes (H2A – H2B and H4 – H3)[23,42]. However, the histone encoding NCLDVs have divergent sets of histones and suggests that some viral histones may have evolved to have independent viral functions. It is not clear whether this divergence in viral functions of NCLDV histones preceded LECA[41]. Within the viral eukaryogenesis hypothesis, it is assumed that viral histones were acquired from their proto-eukaryotic hosts and that selection pressures may have led to the specialization of compacting the viral genome utilizing their own histone fusions. However it is entirely possible that eukaryotic histones evolved from proto-eukaryotic histones that were obtained and diversified by the NCLDV that infected them[23]. Although still debated, many seem to favor the scenario that NCLDV histones were acquired from their host and diversified into doublets post LECA[10,23,41,42]. With their highly diverse sets and varied fusions of histones, NCLDVs reveal their own viral

chromatin organization and bring about endless unanswered questions about what role they played in the evolution of the eukaryotic cell.

1.5 Melbournevirus and Medusavirus Maturation

Melbournevirus and Medusavirus are both isolated through co-culture with *Acanthamoeba spp*. and have particle sizes ranging from 180 to 250 nm in size[39,47]. While not the largest of the NCLDV in physical size, as the Pithovirus is 1500 nm, they are still comparable in size to most bacteria[1]. The Melbournevirus proteome is predicted to contain eukaryotic homologs associated with DNA replication and transcription, with extensive phylogenetic analysis suggesting that most of the genome was acquired through horizontal gene transfer. In fact, the Melbournevirus is part of the *Marseilleviridae* family of NCLDV which were the first giant viruses known to encode histone-like proteins, distinctively as fusions. However, while the Medusavirus proteome encodes each of the core histone-like proteins, it lacks many of the eukaryotic homologs seen in Melbournevirus[47]. Phylogenetic analysis of core histones in the Medusavirus places each at the root of their respective clades, suggesting that Medusavirus histones may provide insight into the origin of the nucleus[21]. With Melbournevirus and Medusavirus possessing their unique set of histones, it is currently not known what role, if any, viral histones play in the infectivity of each virus.

The Melbournevirus maintains a different replication process than other NCLDV due to its smaller particle size. While most NCLDV are taken up by phagocytosis of a single particle, the Melbournevirus can enter as a group of particles that act together as giant vesicles to enhance the efficiency of infection[1]. Only a few hours post-infection (PI), the Melbournevirus begins to generate

viral factories and to restructure the nuclear membrane. This allows the Melbournevirus to transiently recruit nuclear machinery into the cytoplasm to complete DNA replication, transcription, and translation. After 24 PI, the Melbournevirus releases its virions filled with the replicated viral genome (**Figure 1.3**)[39]. The Medusavirus, however, manipulates the host nucleus differently by employing it to complete viral DNA replication. After only 1 hr PI, Medusavirus DNA is found in the host nucleus, where viral DNA replication continues for 14 hr. At that point, viral factories in the cytoplasm locate themselves at the periphery of the nuclear membrane to uptake the viral DNA from the nucleus. The Medusavirus virions have been shown to release from the host as early as 14 hr PI, with particles containing various amounts of viral DNA likely due to the incomplete transfer from the host nucleus (**Figure 1.3**)[21].

As Melbournevirus replication takes place in the cytoplasm and viral histones have been identified in the viral particle, it is assumed that the Melbournevirus histone-like fusions regulate only the viral genome[40]. However, this does not negate the fact that these histone-like proteins are essential for viral infectivity, which raises questions regarding chromatin organization with histone fusions. Additionally, the intimate association of Medusavirus with the host nucleus provokes questions regarding the utilization of these histones during infection. Once these histone-like genes are translated, do they get transported exclusively to the viral factories? Are they shuttled to the nucleus, and can they maintain a role in the regulation of both the host genome and their own? Do the viral histone-like proteins associate with host histone proteins? Understanding the essential nature and role of viral histone-like proteins during infection would greatly contribute insight to the unique life cycle of NCLDV.

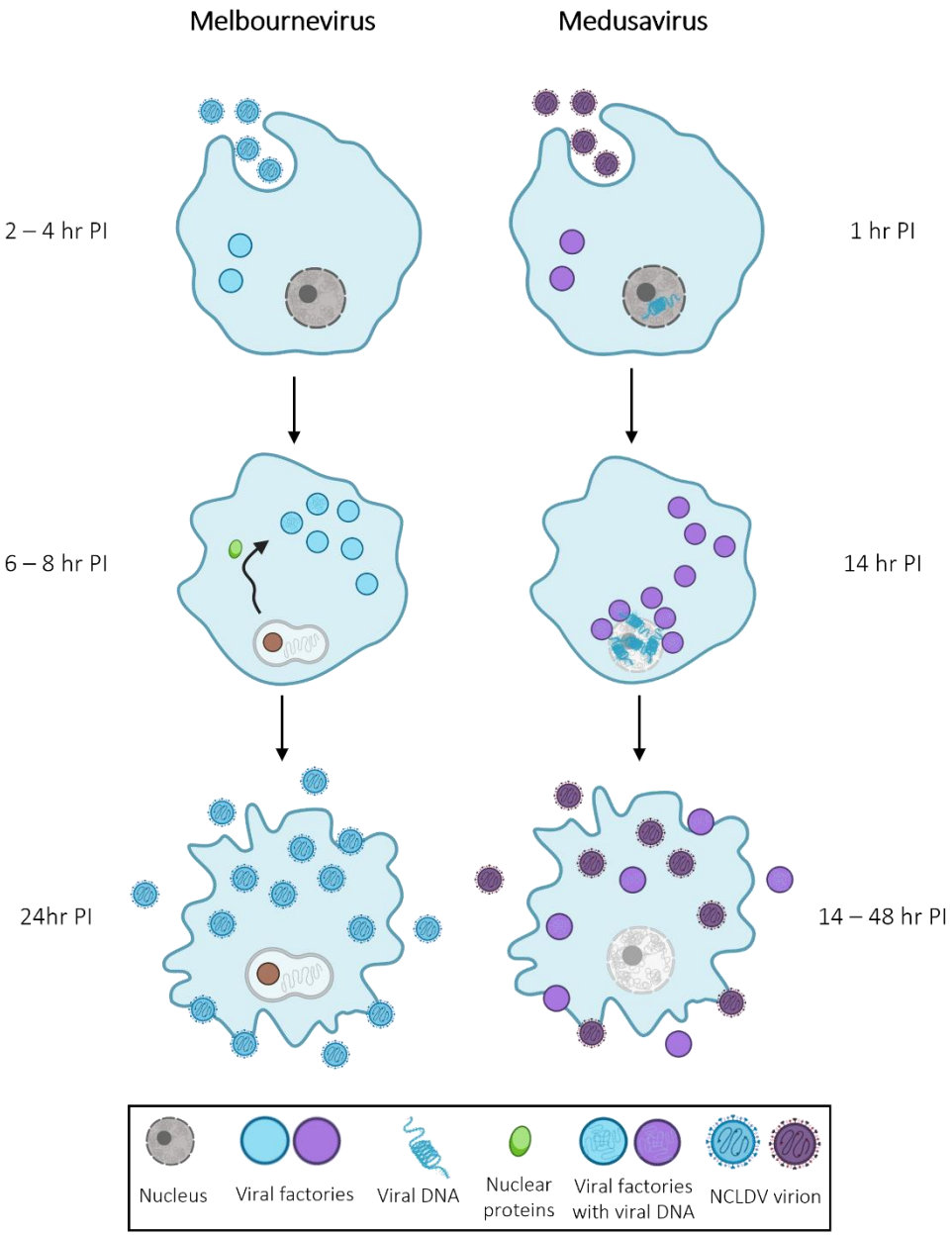


Figure 1.3 Host dependencies during Melbournevirus and Medusavirus maturation.

Schematic representation of the infection and maturation of both Melbournevirus and Medusavirus over their various infection times. The nuclear dependencies of each host are highlighted by the transient recruitment of host nuclear machinery for DNA replication by the Melbournevirus at 6 – 8 hr PI and the utilization of the nucleus for viral DNA replication by the Medusavirus through 14 hr PI[21,40,51].

Chapter 2

Virus-encoded histone doublets are essential and form nucleosome-like structures

Often seen as a hallmark of eukaryotes, the organization of genomic DNA through histones has been challenged by the presence of histones in archaea and more recently, in bacteria and in giant viruses. Here, I contributed to the first demonstration that Melbournevirus histone doublets form nucleosomes that are able to accommodate their linker region between histone folds. I helped Dr. Yang Liu purify histones from Melbournevirus, performed sequence analysis comparing other *Marseilleviridae* histones by considering key conserved eukaryotic residues, modeled potential linkers for each histone doublet, and performed molecular dynamic flexible during refinement of Dr. Yang Liu's Melbournevirus nucleosome-like particle structure (Figures 3.1A, S3.1A, S3.2, 3.4, S3.4, S3.5, Movies S3.1, S3.2, S3.3 and Tables S3.3, S3.4). Through cellular work performed by our collaborator, Dr. Chantal Abergel, it was shown that these viral histone doublets are essential for infectivity, co-localize to the viral factory, and are in relative abundancy similar to the Melbournevirus major capsid protein (Figures S3.1B-H, 3.2 and Table 3.1, S3.1). This work has been peer-reviewed and published in its present form in Cell under the title 'Virus-encoded histone doublets are essential and form nucleosome-like structures' (https://doi.org/10.1016/j.cell.2021.06.032). Drs. Karolin Luger, Yang Liu, Chantal Abergel, Keda Zhou, Samuel Bowerman, and I contributed to the writing and editing.

2.1 Summary

The organization of genomic DNA into defined nucleosomes has long been viewed as a hallmark of eukaryotes. This paradigm has been challenged by the identification of "minimalist" histones in archaea and more recently by the discovery of genes that encode fused remote

homologs of the four eukaryotic histones in *Marseilleviridae*, a subfamily of giant viruses that infect amoebae. We demonstrate that viral doublet histones are essential for viral infectivity, localize to cytoplasmic viral factories after virus infection, and ultimately are found in the mature virions. Cryogenic electron microscopy (cryo-EM) structures of viral nucleosome-like particles show strong similarities to eukaryotic nucleosomes despite the limited sequence identify. The unique connectors that link the histone chains contribute to the observed instability of viral nucleosomes, and some histone tails assume structural roles. Our results further expand the range of "organisms" that require nucleosomes and suggest a specialized function of histones in the biology of these unusual viruses.

2.2 Introduction

The organization of genomic DNA with histones into distinct complexes known as nucleosomes is a universal and highly conserved feature of all eukaryotes. The eukaryotic nucleosome core invariably comprises two copies each of the four unique histones H2A, H2B, H3, and H4. Each protein has a histone fold (HF) region that is structurally conserved between the four histones, as well as additional HF extensions and highly charged cationic tails that are unique to each[52]. H2B-H2A and H4-H3 form obligate heterodimers that assemble into an octamer that wraps 147 bp DNA to form nucleosomes[25].

The evolutionary origin of eukaryotic chromatin organization is widely thought to lie in the archaeal domain of life[45]. Archaeal histones are limited to the HF and are encoded by only one gene or a few closely related genes. They bind and bend DNA as homodimers or quasi-symmetric heterodimers using architectural principles also seen in eukaryotic histones but lack the ability to

form defined particles and instead exist in “slinky-like” assemblies that organize between 90 and ~600 bp DNA[53,54].

The genomes of some viruses (in particular nuclear DNA viruses and retroviruses that need to evade the host DNA damage recognition machinery) are organized into nucleosomes by appropriating eukaryotic histones and the host nucleosome assembly machinery during the latent and early lytic phase[55,56]. With the exception of SV40, viral genomes are not organized with nucleosomes in the capsid, and they do not encode viral histone homologs[56,57]. In contrast, distinct histone-like proteins with homology to eukaryotic H2A, H2B, H3, and H4 have been identified in the genomes of some nucleocytoplasmic large DNA viruses (NCLDVs). Several members of the *Marseilleviridae* isolated from the Amoeba *Acanthamoeba castellanii* encode homologs of the four histone proteins in native doublet form, where H4 is fused to H3 and H2B is fused to H2A (**Figure 2.1A**)[49]. These proteins are present in Marseillevirus virions, where they might participate in the organization of the large (>300 kb) viral genomes[39,58,59]. The arrangement of histones in doublets, the low level (<30%) of sequence similarity with eukaryotic histones, and their high degree of conservation within the family of *Marseilleviridae* (**Figures 2.1A and S2.1A**) suggest that they have evolved to fulfill viral functions other than (or in addition to) DNA compaction. Here, we show that viral histones are essential for viral infectivity, co-localize in the viral factory, and are in the same order of abundance as the major capsid protein in the viral capsids. Using cryogenic electron microscopy (cryo-EM), we demonstrate that viral histone doublets indeed interact with DNA to form nucleosome-like particles with distinct structural properties not seen in eukaryotes.

2.3 Results

2.3.1 MV histone doublets are essential for viral propagation and accumulate in viral factories and mature virions

Melbournevirus (MV) encodes three putative histone doublets (MEL_369, MEL_368, and MEL_149, here named MV-H2BH2A, MV-H4-H3, and MV-miniH2B-H2A). Putative histone domains in each doublet are linked by a ~20-amino-acid connector (**Figure 2.1A**). Although these doublet histone proteins share less than 30% amino acid sequence identity with eukaryotic histones, they are highly conserved within the group of *Marseilleviridae* (**Figure S2.1A**).

Secondary structure prediction suggests that viral histones form HFs (α1-L1-α2-L2-α3) but have either longer (MV-H2B-H2A) or shorter (MV-miniH2B-H2A) H2A C-terminal tails than their eukaryotic counterparts (**Figure 2.1A**). The H3 αN helix (which organizes the terminal turn of eukaryotic nucleosomal DNA) is predicted to be present in the viral H4-H3 doublets, while the sequence of the H2A docking domain (which tethers the H2A-H2B dimers to the $(H3\text{-}H4)_2$ tetramer in eukaryotic nucleosomes) diverges from eukaryotic H2A (**Figure S2.1A**). Many of the ‘‘signature’’ amino acids that are important for DNA binding in eukaryotic histones are conserved in viral histones. These include arginine side chains that reach into the DNA minor groove and, in MV-H4-H3, a threonine from the paired L1 loop (R-T pair), as well as many other basic amino acids that are also interacting with the DNA (**Figures 2.1A and S2.1A**).

To identify the localization of MV histone doublets during viral assembly, we imaged transfected cells expressing fluorescently (GFP) tagged MV-histone doublets as well as tagged amoeba histones at different time points post-infection (PI) and compared their localization with non-infected cells. Amoeba GFP-H2A concentrated only in nuclei, where it presumably associates

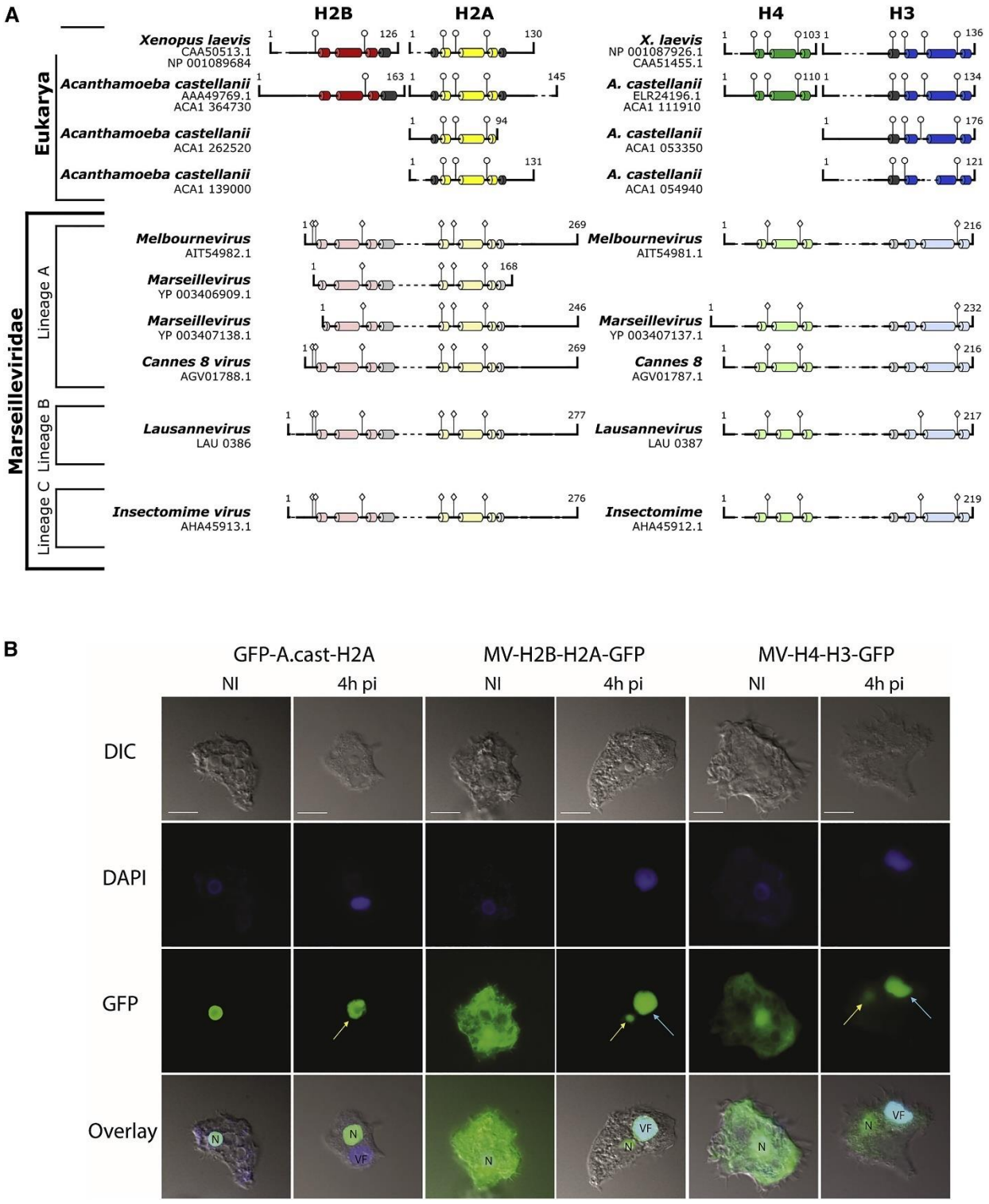


Figure 2.1 MV histone doublets are re-localized to viral factories (VFs).

Histone dimer pairs (H2B-H2A and H4-H3) within Eukarya were aligned against the doublet Marseilleviridae histones using HHpred's multiple sequence alignment tool (ClustalΩ). Known α

helices from the histone fold (HF) domain in Eukarya are dark-colored tubes (H2B, red; H2A, yellow; H4, green; H3, blue; and additional helices, gray). Predicted α helices in MV histones were generated using HHrped's Quick 2D prediction web server (shown in lighter coloration) within the Marseilleviridae histone doublets. Known R-T pairs and DNA binding residues are shown in Eukarya histones along with their conservation within Marseilleviridae histones; additional predicted DNA-binding residues are shown (positions demonstrated by lollipops). (B) Light microscopy fluorescence images (scale bar, 10 μm) of A. castellanii cells transfected with GFP-A. castellanii-H2A, MV-H2B-H2A-GFP, and MV-H3-H4-GFP, non-infected and infected with MV at 4 h PI. While GFP-A. castellanii-H2A concentrates only in the nucleus (N) of the non-infected cells, MV-H2B-H2A-GFP and MV-H3-H4-GFP are scattered in the entire cell (including the nucleus). Upon virus infection, GFP-A. castellanii-H2A remains in the nucleus (yellow arrows), while MV-H2B-H2A-GFP and MV-H3-H4-GFP re-localize to the VF (cyan arrows). DAPI staining remains in the nucleus all along the infection, but the intense fluorescence in the late VF hides the staining of the nucleus at 4 h PI.

with genomic DNA to form nucleosomes, regardless of MV infection (**Figures 2.1B and S2.1B**). In contrast, unique and distinct patterns were observed for the transfected MV-histone doublets upon virus infection, particularly in the viral factory (VF; foci in the amoeba cytoplasm, where viral transcription, replication, and assembly take place). MV-H2B-H2A-GFP and MV-H4-H3-GFP were initially dispersed throughout the whole cell, including the nucleus, but their localization started to change between 1 and 2 h PI (**Figures S2.1C and S2.1D**). Eventually, MV-H2B-H2A-GFP and MV-H4-H3-GFP accumulated and concentrated predominantly as foci in the viral factory (**Figures 2.1B, S2.1C, and S2.1D, cyan arrows**), although they were still present in nuclei at much lower concentrations (**Figure 2.1B, yellow arrows**). Similar localization patterns were obtained with MV-miniH2B-H2A-GFP (**Figure S2.1E**). In addition, co-localization of MVH2B-H2A-GFP and MV-H4-H3-mRFP along with viral DNA was observed in the viral factory, indicating that MV histones are recruited for viral DNA packaging and virion production (**Figure S2.1F**). Thousands of copies of the AT-rich 360-kb viral genome are produced in the viral factory to produce approximately thousands of mature virions, compared to the 24 copies of a 46.7-Mb GC-rich amoeba genome. As such, DAPI

staining for the viral factory overpowers the weaker staining of the nucleus, which is not a sign of host genomic DNA disappearance during the infectious cycle (Figure S2.1G).

Mass spectrometry proteomic analysis of the purified MV virions was performed (Table 1) [58]. MV-histone doublets are among the most abundant proteins in the MV viral proteome, with the number of copies per capsid in the range of the major capsid protein. Importantly, the amount of histone doublets (even when using the most conservative estimate), is sufficient to package the entire viral genome with nucleosomal particles with a repeat length of ~180 bp (~4,000 histone doublets per 360 kilo base pairs (kbp) viral genome).

Table 2.1 Mass spectrometry proteomics of the purified MV virions. Histone doublets are indicated in red.

Protein ID	Annotation	Final rank	iBAQ R1	Copy R1	Rank R1	iBAQ R2	Copy R2	Rank R2
MEL_236	hypothetical protein	1	1,7E+09	14087	1	3E+09	15107	1
MEL_342b	hypothetical protein	2	1,2E+09	9916	2	2E+09	12017	2
MEL_368	**histone H4-H3 doublet**	3	9,8E+08	**8362**	4	2E+09	**10344**	3
MEL_305	**major capsid protein**	4	1,1E+09	**9240**	3	2E+09	**9240**	4
MEL_274	hypothetical protein	5	5,7E+08	4835	5	1E+09	5385	5
MEL_369	**histone H2B-H2A doublet**	6	5E+08	**4223**	6	8E+08	**3895**	6
MEL_247	Papain-like cysteine protease	7	3,5E+08	2935	7	8E+08	3812	7
MEL_280	hypothetical protein	8	2,2E+08	1862	8	5E+08	2429	8
MEL_234	hypothetical protein	9	2,1E+08	1776	9	5E+08	2354	9
MEL_241	hypothetical protein	10	1,8E+08	1569	10	4E+08	2208	10
MEL_358	hypothetical protein	11	1,4E+08	1211	11	3E+08	1454	11
MEL_207	thioredoxin	12	1,3E+08	1083	12	3E+08	1424	12
MEL_020	hypothetical protein	13	1,2E+08	1053	13	2E+08	1252	13

MEL_202	hypothetical protein	14	9,7E+07	828	15	2E+08	1072	14
MEL_260	hypothetical protein	15	9,8E+07	835	14	2E+08	1004	15
MEL_200	Transmembrane domain containing protein	16	7,9E+07	676	16	2E+08	851	16
MEL_065	hypothetical protein	17	7E+07	596	17	2E+08	775	17
MEL_025	peptidase	18	6,2E+07	531	19	1E+08	640	19
MEL_097	hypothetical protein	19	6,3E+07	534	18	1E+08	613	21
MEL_282b	hypothetical protein	20	5,4E+07	463	24	1E+08	657	18
MEL_181	ATP-dependent helicase	21	5,7E+07	482	21	1E+08	634	20
MEL_354	Transmembrane domain containing protein	22	5,8E+07	493	20	1E+08	604	22
MEL_365b	hypothetical protein	23	5,6E+07	477	22	1E+08	601	24
MEL_196	hypothetical protein	24	5,5E+07	468	23	1E+08	601	23
MEL_294	thioredoxin	25	5,4E+07	463	25	1E+08	583	25
MEL_360	AAA-family ATPase	26	5,1E+07	435	27	1E+08	543	27
MEL_244	hypothetical protein	27	5,1E+07	437	26	1E+08	532	28
MEL_185	hypothetical protein	28	4,6E+07	389	29	1E+08	572	26
MEL_275	hypothetical protein	29	4,9E+07	420	28	1E+08	505	29
MEL_269	Transmembrane domain containing protein	30	4,2E+07	354	30	9E+07	460	32
MEL_228	Transmembrane domain containing protein	31	3,9E+07	333	32	9E+07	476	30
MEL_168	lipase	32	3,9E+07	328	33	9E+07	476	31
MEL_041	Peptidase	33	3,9E+07	334	31	8E+07	381	36
MEL_331	hypothetical protein	34	3,7E+07	312	34	8E+07	396	34
MEL_352	lectin domain containing protein	35	3,1E+07	263	38	9E+07	442	33
MEL_211	Transmembrane domain containing protein	36	3,5E+07	297	35	7E+07	379	37
MEL_380	serine/threonine protein kinase	37	3,3E+07	279	37	8E+07	393	35

MEL_149	**mini H2B-H2A histone**	38	3,4E+07	**289**	36	7E+07	**362**	39
MEL_089	hypothetical protein	39	2,9E+07	249	39	7E+07	368	38
MEL_216	helicase	40	2,8E+07	241	40	6E+07	303	40
MEL_235	serine/threonine protein kinase	41	2,7E+07	233	41	6E+07	282	44
MEL_338	hypothetical protein	42	2,3E+07	199	44	6E+07	296	41
MEL_301	Transmembrane domain containing protein	43	2,5E+07	209	43	6E+07	282	43
MEL_231	hypothetical protein	44	2,5E+07	216	42	5E+07	266	45
MEL_223b	membrane protein	45	2,1E+07	180	45	6E+07	288	42
MEL_297	hypothetical protein	46	2E+07	167	49	5E+07	265	46
MEL_213b	hypothetical protein	47	2,1E+07	178	47	5E+07	247	47
MEL_215	hypothetical protein	48	2,1E+07	177	48	5E+07	237	48
MEL_255	Papain-like cysteine protease	49	2,1E+07	179	46	5E+07	234	49
MEL_387	hypothetical protein	50	1,9E+07	159	52	4E+07	212	50
MEL_219	Ribonuclease III	51	1,9E+07	163	50	4E+07	207	53
MEL_278	disulfideoxidoreductase	52	1,9E+07	160	51	4E+07	208	51

Analysis of published transcriptomic data obtained for a member of the *Marseilleviridae* during its infectious cycle in *A. castellanii* sheds light on the possible contribution of the host to viral nucleosome assembly (**Figure 2.2A**)[40]. As expected, the viral DNA polymerase is expressed before the viral histones, between 1 and 2 h PI. Expression of viral histones begins 2–4 h PI and continues along the infectious cycle. Host histone expression, however, decreases after 2 h PI, as is the case for most of the *A. castellanii* genes (see, for reference, the actin gene in gray). The cellular histone chaperones are either never expressed (FUN_000176, FUN_002445, FUN_006481, FUN_014954, FUN_006327 lanes in gray at the bottom of the figure) or have their expression

decreased after 2 h PI. No obvious homologs to the cellular histone chaperones were detected in the *Marseilleviridae* genome.

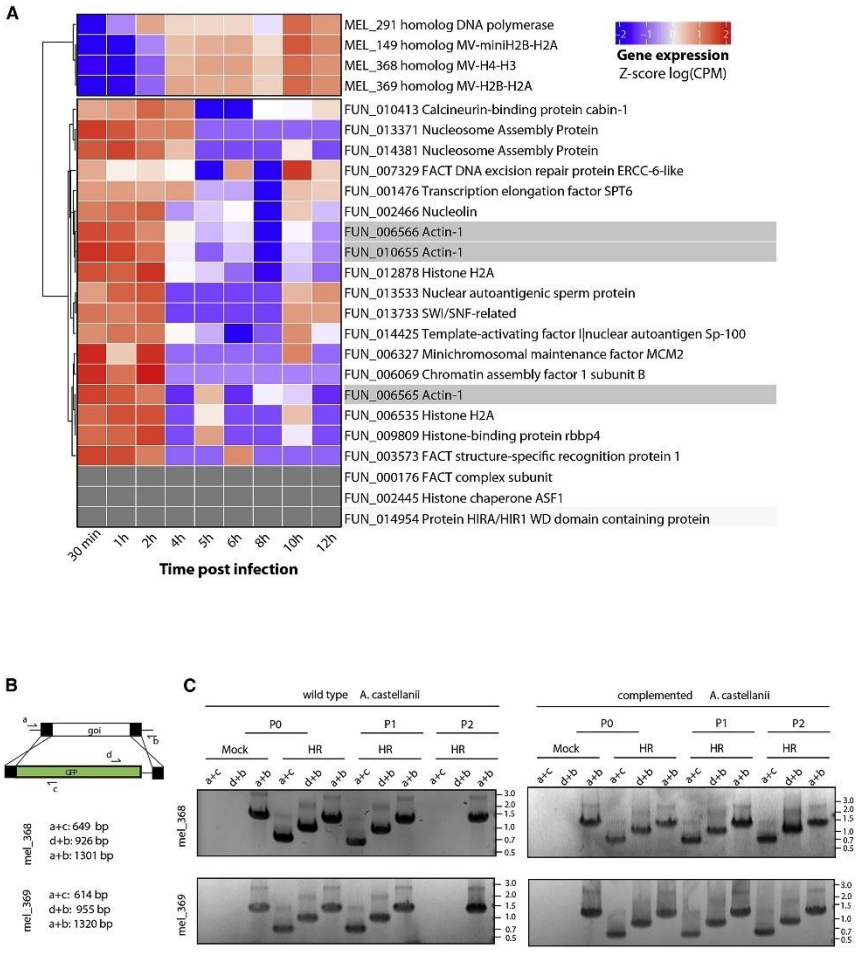


Figure 2.2 MV histone doublets are essential for virus fitness.

(A) Histone (MV and host) expression and host histone chaperones expression profile during the MV infectious cycle. Host actin genes expression were used for reference. (B) Schematic representation of the strategy used to generate the recombinant mel_368 or mel_369 KO. Expected PCR product size are also shown. Primer a: HB146 or HB147 and primer b: HB145 or HB148 for mel_368 and mel_369 respectively; primer c, GFP reverse; primer d, GFP forward (Table S1). goi, gene of interest. (C) PCR demonstrates correct integration of vectors in the locus of mel_368 or mel_369. Analysis of fitness changes associated with histone KO was analyzed by virus competition assay in wild-type or complemented amoebas. Mock cells were incubated with Superfect in the absence of plasmid. a+c, 5′ integration; d+b, 3′ integration; a+b, wild-type locus. HR, homologous recombination. See also Figure S2.1 and Table S2.1.

In order to assess the fitness cost associated with the impairment of potential nucleosome formation in MV capsids, we designed a strategy to knock out either *mel_369* or *mel_368* (Figures 2.2B and S2.1H). *A. castellanii* cells were transfected with linearized plasmid prior to infection with MV particles. Recombinant viruses were quickly detected by PCR upon 3–5 h PI (primers listed in Table S2.1). Expression of GFP could not be detected at 3h PI, likely due to the late expression profile of the histone promoters. The presence of mutant viral particles was followed during two generations by infecting wild-type amoeba compared to amoeba expressing an ectopic copy of the histones for trans-complementation (Figure 2.2C). Importantly, mutant viral particles were quickly outcompeted by wild-type MV in absence of trans-complementation, while growth of the mutant virus was sustained when the missing copy of the histone gene was stably expressed in the amoeba (Figure 2.2C). This demonstrates that both *mel_369* and *mel_368* are essential, and their knockout (KO) is associated with a strong fitness cost to the virus. Altogether, the distinct localization and expression patterns reveal specific targeting of MV histone doublets to the viral factory and the viral particle.

2.3.2 MV doublet histones form defined, unstable nucleosome-like particles

To determine what types of complexes MV-H2B-H2A, MV-miniH2B-H2A, and MV-H4-H3 doublets form with DNA, we expressed, purified, and refolded the proteins from *E. coli* (Figure S2.2A). When combined with DNA (‘‘601’’ nucleosome positioning sequence) of varying lengths, and using the classic salt-gradient nucleosome reconstitution protocol, MV-H2B-H2A and MV-H4-H3 form defined nucleosome-like particles (MV-NLPs) that are stable at 37C, whereas individual histone doublets fail to form defined bands on DNA (Figure 2.3A)[60]. The composition of the MV-

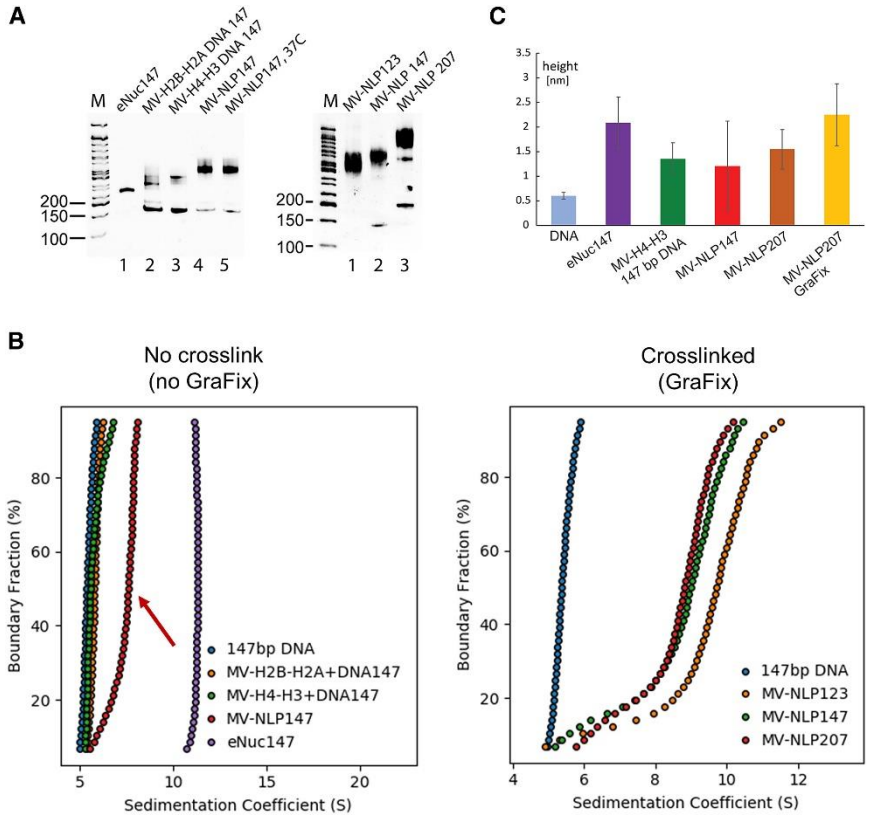


Figure 2.3 Histone doublets in *Marseilleviridae* form nucleosome-like particles (MV-NLPs).

(A) Native PAGE of reconstituted MV-NLPs with "601" DNA of various lengths. Left panel: individual and combinations of histone doublets reconstituted onto 147 bp DNA; right panel: MV-H2B-H2B and MV-H4-H3 reconstituted onto 601 DNA of varying lengths. (B) Sedimentation velocity analytical ultracentrifugation (SV-AUC) of MV-NLPs. Left: van Holde-Weischet plot of eukaryotic nucleosomes (eNuc), histone-DNA complexes with individual MV histone doublets, and native MV-NLPs with 147 bp DNA (no GraFix); right: van Holde-Weischet plot of crosslinked (GraFix-ed) MV-NLPs with 123, 147, or 207 bp DNA. (C) Height profile of MV-NLPs with 147 or 207 bp DNA, obtained by atomic force microscopy (AFM). The average height profiles with standard deviations of the particles from each sample are shown; representative, original images are shown in Figure S2.2, and statistics are shown in Table S2.2.

NLP was confirmed by sucrose gradient fractionation (**Figure S2.2B**). MV histones also formed defined particles on a native 181 bp DNA fragment derived from the MV genome (GC content = 45%). Irrespective of DNA sequence, and unlike *X. laevis* nucleosomes (a representative of eNuc), MV-NLPs dissociate upon heat treatment at 55C (**Figure S2.2C**). MV-miniH2B-H2A alone binds DNA

poorly, and no homogeneous MV-NLPs were formed with MV-H4-H3 and DNA (**Figure S2.2D**). We therefore focused on MV-NLPs with full-length MV-H2B-H2A and MV-H4-H3.

Sedimentation velocity analytical ultracentrifugation (SV-AUC) yields information on the size and shape distribution of macromolecular assemblies in solution. The diffusion-corrected sedimentation value of a particle is proportional to its mass, and inversely proportional to its viscous drag. The sedimentation behavior of MV-histone doublets assembled onto 147 bp DNA was compared to that of *X. laevis* histones reconstituted on the same DNA ($eNuc_{147}$). The $eNuc_{147}$ particle sediments at 11 S, while MV-NLP_{147} sediments at ~8 S, with a ''tail'' toward lower S values, indicating dissociation of DNA at the relatively low (300 nM) concentrations used for AUC experiments (**Figure 2.3B**)[61]. Particles reconstituted with individual MV histone doublets sediment at ~5.5 S (**Table 2.2**). The AUC-derived molecular mass calculated for MV-NLP_{147} (**Figure 2.3B, red arrow**) suggests that at least one copy of the H2B-H2A doublet has been lost due to the sample dilution in AUC (**Table 2.2**). To counteract the apparent dissociation of MV-NLPs, we employed gradient fixation (GraFix) with glutaraldehyde[62]. Histone doublets were efficiently crosslinked, and the position of the main peak fraction in the gradient as well as migration on a native gel was unchanged upon crosslinking, suggesting little to no structural changes (**Figure S2.2E**). Particles reconstituted on 123, 147, and 207 bp DNA fragments sedimented between 8.8 and 9.9 S after crosslinking (**Figure 2.3B, Table 2.2**). The frictional ratio (f/f_0) (a measure of particle ''extension,'' where larger values indicate increased drag, slowing sedimentation) of all MV-NLPs was higher than that of $eNuc_{147}$, and f/f_0 values increase in correlation with DNA length (**Table 2.2**). $MVNLP_{123}$ has the highest S-value among MV-NLP

Table 2.2 S values (S(20,W)), frictional ratios (f/f0) and calculated molecular weights (including confidence intervals) of MV histone-DNA complexes derived from SV-AUC.

Sample		S(20,W)	f/f0	Molecular weight (kDa) Exper./Theoretical
No GraFix	147bp DNA	5.10 (5.09, 5.10)	3.09 (3.09, 3.09)	117.13 (116.9, 117.3) / 97
	MV-H2B-H2A 147	5.68 (4.61, 6.75)	2.66 (2.51, 2.81)	134.61 (89.83, 179.38) / 150
	MV-H4-H3 147	5.28 (4.59, 5.96)	2.79 (2.70, 2.87)	120.93 (104.14, 137.71) / 146
	MV-NLP147	7.78 (7.74, 7.82)	2.0 (1.96, 2.04)	179.17 (173.55, 185.80) / 206
	eNuc147	11.16 (11.15, 11.16)	1.54 (1.52, 1.56)	218.5 (214.4, 222.5) / 210
GraFix	MV-NLP123 (nucleosome-like)	9.86 (7.99, 11.72)	1.77 (1.52, 2.02)	221.46 (170.93, 271.99) / 190
	MV-NLP147 (nucleosome-like)	8.79 (7.30, 10.29)	2.01 (1.86, 2.14)	217.79 (140.57, 295.00) / 206
	MV-NLP207 (nucleosome-like)	9.01 (7.48, 10.55)	2.26 (2.00, 2.51)	256.58 (194.54, 318.62) / 245

constructs, despite having the smallest mass, because of its significantly lower f/f_0 value. This suggests that the additional DNA in $MVNLP_{147}$ and $MV\text{-}NLP_{207}$ is extended.

We visualized MV-NLPs by atomic force microscopy (AFM). Particles reconstituted with either MV-H4-H3 alone or the full complement of histone doublets were deposited onto mica surfaces, and their height profiles were compared to $eNuc_{147}$ (Figures 2.3C and S2.2F; Table S2.2).

The majority of $eNuc_{147}$ survived the low concentration (2 nM) required for AFM imaging and presented the characteristic height profile of ~2.1 nm[34]. In contrast, over 30% of the observed particles heights indicative of free DNA, indicating complete MV-NLP_{207} disassembly, and the remaining particles had lower height profiles (~1.3 nm) than the $eNuc_{147}$ complex (**Figures 2.3C and S2.2F; Table S2.2**). This reduced height suggests that MV-NLP_{207} transiently disassemble at the low concentration required for AFM. Crosslinking MV-NLP prevented disassembly and resulted in particles comparable in height to $eNuc_{147}$.

2.3.3 MV-NLPs resemble eukaryotic nucleosomes

Single-particle cryogenic electron microscopy (cryo-EM) was used to visualize crosslinked MV-NLP_{207}. GraFix was required as the majority of untreated MV-NLPs dissociate during plunge freezing. Raw images and 2D classes show that MV-NLPs share many characteristic features of eukaryotic nucleosomes (**Figure S2.3A**), with defined segments of extending linker DNA visible in many classes. After 3D classifications and refinement, we obtained a structure of the $MVNLP_{207}$ at 6.1 Å (**Figure S2.3A; Table S2.3**). A second dataset obtained with GraFix-treated particles reconstituted onto 147 bp DNA ($MVNLP_{147}$) yielded an improved overall resolution of ~4.0 Å (**Figure S2.3B; Table S2.3**). In both structures, electron density for the DNA and histone helices are clearly distinguishable (**Videos S2.1 and S2.2**), which allowed the assignment of ~120 bp of bound DNA and all histone chains (**Figure 2.4A; Video S2.3**).

The overall dimensions of the DNA superhelix, the path described by the DNA around the histone core, and the overall layout of HF helices are similar between MV-NLPs and eukaryotic nucleosomes (**Figures 2.4A and 2.4B; Table S2.4**). However, DNA wrapping is incomplete and

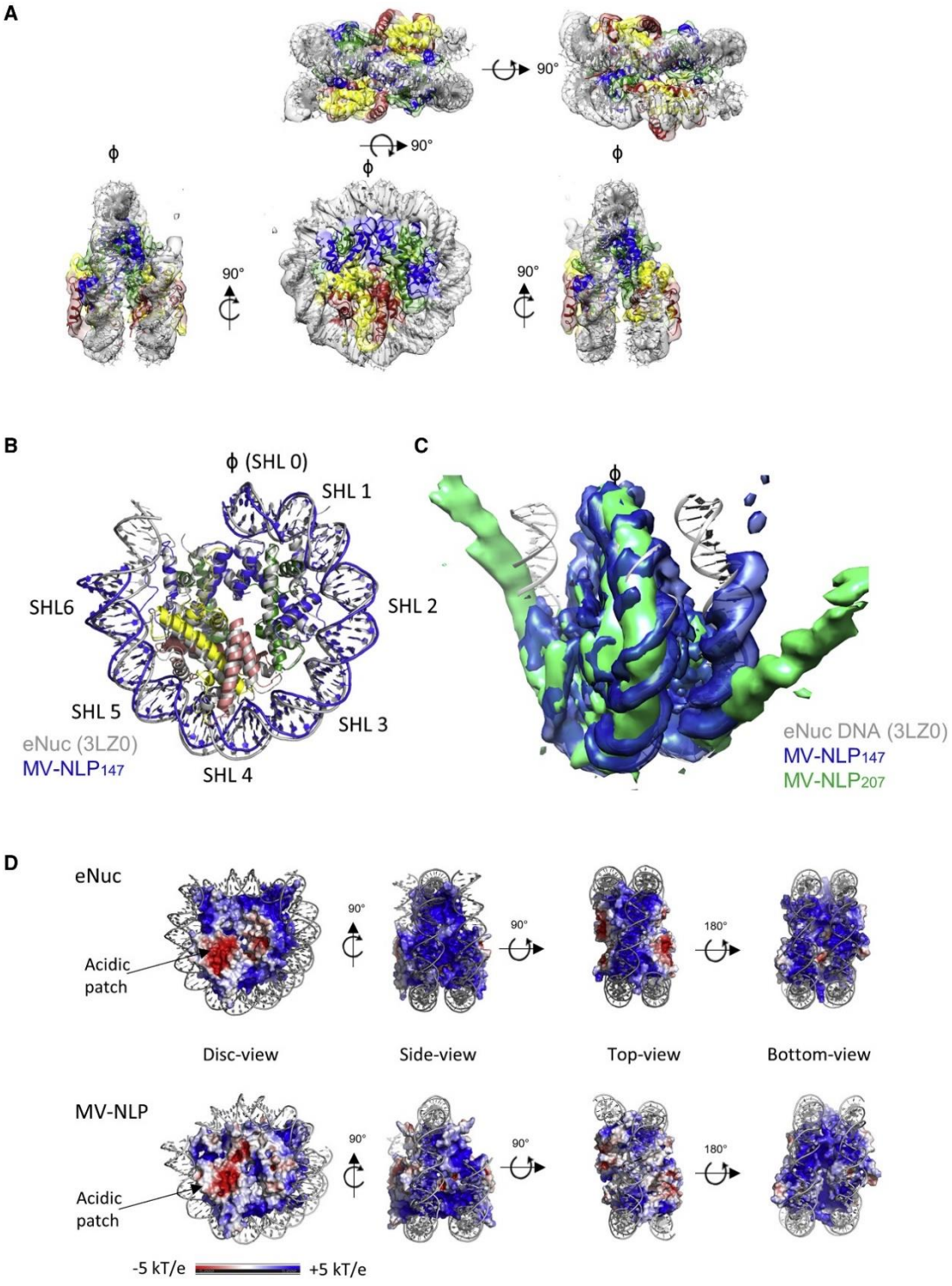


Figure 2.4 Cryo-EM reveals that MV-histone doublets form nucleosome-like structures with asymmetrically extending DNA.

(A) Overview of MV-NLP$_{147}$ and electron density. The equivalent regions of MV H3, H4, H2A, and H2B are shown in blue, green, yellow, and red, respectively. (B) Overlay of MV-NLP$_{147}$ (blue) with

eNuc (gray). Only 80 bp of DNA with associated histones are shown for clarity. Superhelix locations (SHLs) are numbered from 0 to 6 starting from the nucleosome dyad (ϕ). (C) Comparison of the DNA path of $eNuc_{147}$ (gray ribbon diagram), MV-NLP_{147} (blue electron density), and MV-NLP_{207} (green electron density). (D) Charged surface representation of the histones for MV-NLP_{147} and $eNuc_{147}$. Coordinates for $eNuc_{147}$ were taken from 3LZ0. See also Figures S2.3A and S2.3B and Table S2.3.

asymmetric in both MV-NLPs, as only ~120 bp are wrapped in most particles. This is more pronounced in MV-NLP_{207}, where well-defined electron density describes a straight path for DNA extending away from the histone core at super helical location (SHL) 4.5, whereas the DNA extending at the other side (from SHL 5.5) appears to be more disordered (**Figure 2.4C; Video S2.1**). In MV-NLP_{147} complex slightly more DNA is bound (up to SHL 5.5 on both sides), and only one arm of extending DNA has defined density (**Figure 2.4C; Video S2.2**). This asymmetry is likely a consequence of the asymmetric 601 DNA sequence[63,64].

Initial models of MV-histone doublets were generated though homology modeling, where the ~28- and ~20-amino-acid connectors in H4-H3 and H2B-H2A were constructed through de novo methods[65,66]. These were docked into the MV-NLP_{147} density, with good agreement with nucleosome-like configurations (**Figures2. 4A and 2.4B**; correlation coefficient between model and experimental density 0.766).

Overall, the MV-histone core is less positively charged than the eukaryotic histone octamer (pI of ~9.5 versus ~11.0; **Figure 2.4D**). While basic amino acids describe a distinct path for the DNA in both histone cores, and many amino acid side chains that engage DNA are conserved between viral and eukaryotic histones (**Figures 2.1A and S2.1A**), positive charges along the helical path are less pronounced in MV histones, in particular in the region formed by the H3-H3' four-helix bundle

(‘‘dyad’’). This contributes to the reduced stability of MV-NLPs compared to eNuc. Additionally, the acidic patch, the primary point of interactions between eukaryotic nucleosomes and nuclear proteins[67], differs in size and charge, due to fewer acidic residues in MV-H2A (**Figure 2.4D**).

The two MV-H4-H3 doublets superimpose onto the eukaryotic $(H4\text{-}H3)_2$ tetramer with a backbone root-mean-square deviation (RMSD) of ~2 Å (**Figure 2.5A; Table S2.4**). The main-chain arrangements of the MV-H3-H3’ four-helix bundles at the nucleosomal dyad are similar as in $eNuc_{147}$, but the interface lacks the hallmark histidine-cysteine configuration. Each MV-H2B-H2A doublet interfaces with the H4 portion of a MV-H4-H3 doublet through a four-helix bundle of similar architecture as the H4-H2B interface in $eNuc_{147}$. However, the MV-H2B α2 helix is one turn shorter than the eukaryotic sequence (due to a conserved proline at position 83), and the tyrosines that form p-stacking interactions in eNuc systems are consistently absent in both MV-H4 and MV-H2B (**Figure 2.5B**). Additionally, while the main chains of the two H2A L1 loops (^{38}NYAE^{41}) are close to one another in $eNuc_{147}$ and form direct interactions through their side chains, the distance between L1 loops in $MV\text{-}NLP_{147}$ (^{153}GGCS^{156}; positions 214–217 in **Figure S2.1A**) is increased and shows no direct L1-L1 interactions. L1-L1 separation is accompanied by a similar increase in the distance between H2B-H2A centers (36.3 Å in $eNuc_{147}$ versus 38.6 Å in $MV\text{-}NLP_{147}$; **Table S2.4**). In addition to being further apart in the $MV\text{-}NLP_{147}$ complex, the H2B-H2A HFs reorient away from the nucleosome core (**Figures 2.5C and S2.4A; Table S2.4**). Conversely, the H2A docking domains of MV-H2A are angled inward, toward the dyad axis, in an altogether different arrangement than what is seen in $eNuc_{147}$ (**Figure 2.5C; Table S2.4**; 6–8 Å backbone RMSD for chains C and G, respectively). Together, these structural changes may contribute to the reduced stability and

increased propensity to form sub-NLPs (hexasomes and tetrasomes) of MV-NLPs in comparison to the eNuc complex.

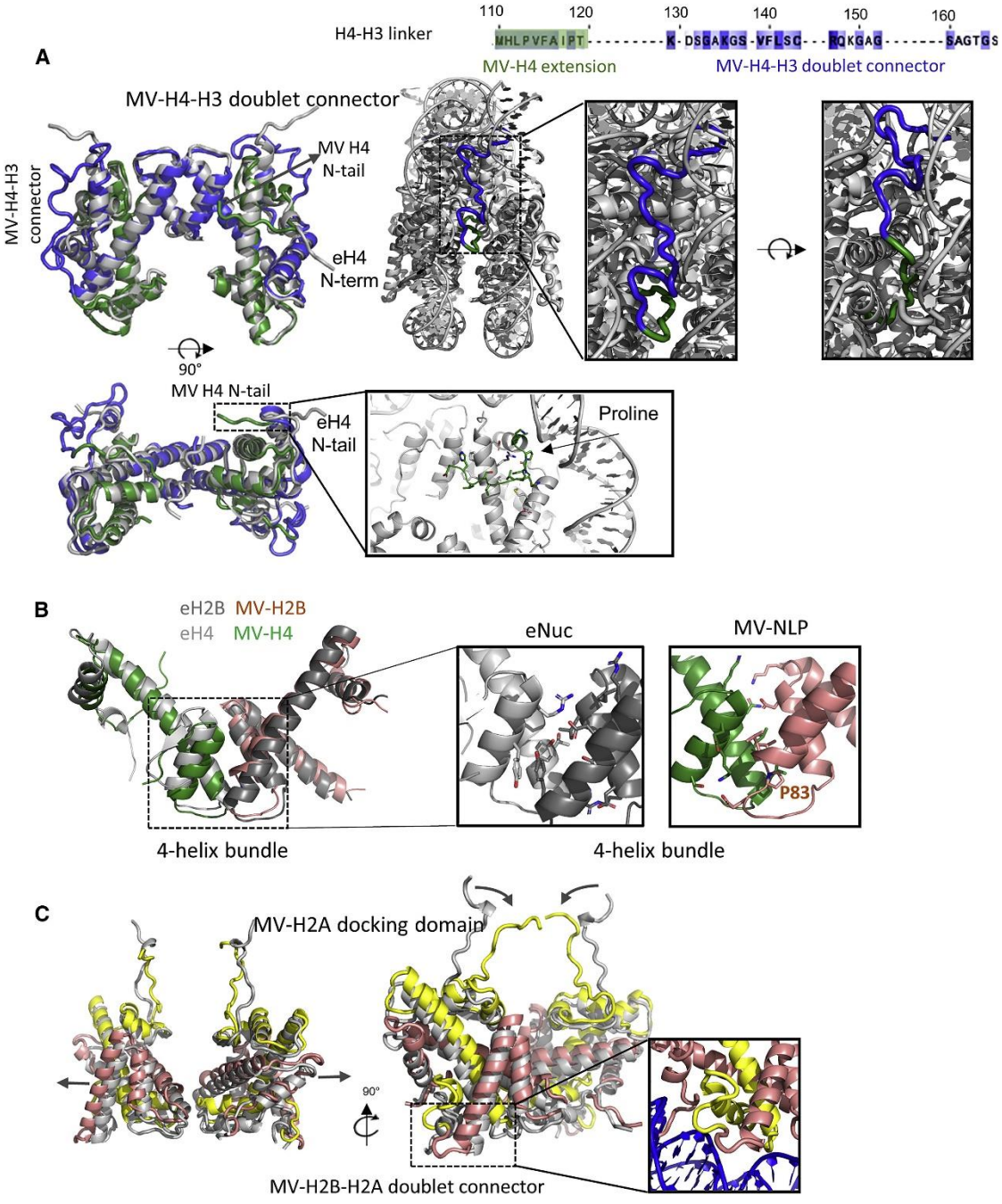


Figure 2.5 Comparison of MV-NLP and eNuc histone structures.

(A) Superposition of two MV-H4-H3 doublets (green and blue) with the eukaryotic (H3-H4)2 tetramer in gray (left), and a close-up of MV-H4-H3 doublet connector (right). Interactions of the

MV-H4 N-terminal tail with the nucleosome are also shown. (B) A comparison of the four-helix bundle structure formed by H4 and H2B. (C) Superposition of two MV-H2B-H2A doublets (in red and yellow) with eukaryotic H2A-H2B dimers in gray (left), and a close-up of the MV-H2B-H2A doublet connector and docking domain. See Table S2.4 for detailed information. Additional possible configurations for both connectors are shown in Figure S2.4.

The connector linking the C terminus of MV-H2B with the N terminus of MV-H2A is easily accommodated, as these are near each other. A variety of favorable conformations are predicted through de novo modeling, with several different arrangements in agreement with observed weak density in MV-NLP$_{147}$ (**Figure 2.5C, inset, and Figure S2.4B**). A conserved arginine is pointing into the minor groove, and main-chain phosphate interactions further hold the loop in place, echoing the contributions of the H2A N-terminal domain in eNuc. Connecting the MV-H4 C terminus with the N-terminal tail of MV-H3 is conceptually more difficult. In eukaryotic H4, the eight most C-terminal amino acids are engaged in contacts with H2A and H2B in the interior of the histone core. Our models show that the equivalent region in MV-H4 maintains docking domain interactions with MV-H2A. Physical constraints dictate that the missing linker segment must continue toward the DNA superhelix and project upward in the direction of MV-H3 αN, with which it connects (Figure 5A). Electron density attributed to the connector is visible between the DNA gyres on both sides and in both structures (MVNLP$_{207}$ and MV-NLP$_{147}$). Although not sufficiently defined to allow modeling through real-space refinement, de novo generation of linker structures provided multiple feasible conformations (**Figure S2.4C**), and molecular dynamics flexible-fitting simulations confirmed that these orientations are physically relevant (**Figure 2.5A**). A stretch of 10 predominantly hydrophobic amino acids packs against the region connecting MV-H3αN with α3 to form a hydrophobic core (**Figures 2.5A and S2.4C**), and the connector is also positioned to engage in main-chain contacts with the DNA backbone at SHL ± 1.5. Mostly flexible and small amino acids

(GAGSAGTGS) form the turn toward MV-H3 αN (Figure 2.5A). As such, the MV-H4-H3 connector (which is highly conserved among *Marseilleviridae* doublets) can be accommodated within the structural framework of a canonical nucleosome, although its presence likely contributes to the unwrapping of the terminal 10 bp of DNA observed in both MV-NLP densities.

The other MV-histone tails that are not engaged in doublet formation also differ from the histone tails in eNuc. The C-terminal tail of MV-H2A is significantly longer than that of eukaryotic H2A (Figure S2.1A). The MV-H4 N-terminal domain (equivalent to the first 26 amino acids in eH4) does not extend over the DNA but takes the opposite direction to interact with MV-H4 α2 and α3 (Figure 2.5A). This density is as well defined as that of the main HF helices, suggesting that this domain may consistently interact with its own histone core rather than with DNA or the acidic patch in neighboring particles, as observed for the eukaryotic H4 tail[68,69]. The amino acid residues in this region are highly conserved among *Marseilleviridae* and differ from the eH4 tail (Figure S2.1A). A conserved proline (P34) at the tip of MV-H4 α1 might be responsible for the redirection (Figure 2.5A).

As is the case with the eH4 tail, the first 20 amino acids of the MV-H4 tail are too disordered to be observed in the electron density map.

2.4 Discussion

The discovery that the genomes of some giant viruses encode histone-like proteins in which the dimerization partners are fused into a single chain was surprising[49]. Our results that these MV-histone doublets associate with the viral factory and indeed with the mature virus and that they form nucleosome-like particles with unique properties support the idea that they are involved in

viral genome organization into the capsids. Instead of organizing DNA into nucleoprotein ‘‘core complexes’’ inside the viral capsid, *Marseilleviridae* utilize their own specialized histone doublets to form nucleosomes and possibly more compact chromatin within the viral particle. In the context of the ongoing debate regarding the role of viruses in the emergence of the eukaryotic nucleus, our findings also provide a new dimension to the diverse role of histones in genome organization of non-eukaryotic entities[44,50].

Marseilleviridae exhibit intermediate dependency on their host, as they develop their viral factory in the cytoplasm but transiently recruit host nuclear proteins to the viral factory to transcribe early genes until the virally encoded RNA polymerase has been synthesized[58]. This is achieved through a reorganized leaky nucleus during the early phase. Amoeba histones are exclusively localized in the host nucleus whether the cells are infected or not, whereas transfected viral histones are present in the nucleus and cytoplasm in uninfected cells but move into viral factories upon infection to integrate into mature virions. This suggests that amoeba histones remain bound to cellular DNA, while the viral histones that made it into the nucleus apparently do not interact with amoeba genomic DNA and can leave the nucleus to associate with the viral factory. As such, amoeba histones are likely targeted to the nucleus and assembled into chromatin by histone chaperones and transporters that mostly avoid viral histones. As is the case for the other nuclear proteins transiently leaving the cell nucleus upon *Marseilleviridae* infection, it remains to be seen whether transfected viral histones are actively or passively recruited to the viral factory.

Many ‘‘nucleosome signature features’’ are maintained among archaeal, eukaryotic, and now viral nucleosomes, such as the overall geometry and arrangement of the DNA superhelix, achieved by interactions between the DNA minor groove backbone and the main chain of

antiparallel L1L2' loops (from H4-H3 and H2B-H2A pairs) and the utilization of positive a helix dipole moments to bind DNA[52]. Several of the ''sprocket arginines'' and their pairing with a threonine from the dimerization partner, as well as the highly conserved R-D clamps stabilizing the HF, are conserved among eukaryotic, archaeal, and viral histones[70]. However, MV-doublet histones stably bind only ~120 bp DNA as opposed to the 147 bp in eukaryotic nucleosomes. In our cryo-EM structures, the unbound DNA is characterized by distinct density indicating that it is in a defined orientation, and in this, it differs from the structures of eukaryotic variant nucleosomes that are also characterized by organizing only ~120 bp DNA[71,72]. Unique to MV-NLPs is also the packing of the H4 N-terminal tail onto the surface of the histone core, the overall less positive charge of the histone core surface including the acidic patch, and, most notably, the linking of MV-H4-H3 and MV-H2B-H2A to form doublet histones. Fused histone genes are also observed in some archaea but to date not in any eukaryotic genome[45]. The amino acid linkers connecting the histone moieties to form the doublets, which are conserved in sequence across *Marseilleviridae*, can be accommodated in the structure. However, the MV-H4-H3 connector (which is conserved in length and amino acid sequence between members of the *Marseilleviridae*) likely contributes to destabilizing the last turn of DNA. Overall, MV-NLPs are significantly destabilized compared to eNuc, and this seems to be an intrinsic property related to their virus-specific function. While this article was in revision, the structure of MV nucleosomes was published, with identical results and conclusions regarding the structural composition of the particles[48].

A common trait of DNA viruses is that they must actively package their DNA into capsids to protect it outside the host cell environment. On the other hand, they also need to make their genomes accessible to the transcription machinery immediately upon infection. Previous analyses

of the infectious cycle of MVs revealed that the capsids appear to dissolve ~2 h PI, leaving a spherical electron dense core in the cytoplasm[58]. Our study suggests that *Marseilleviridae* assemble virally encoded histones into nucleosomes to condense and protect their genome, to allow it to fit into the ~200-nm-diameter icosahedron, also explaining the higher than expected electron density of the spherical core[58]. This process is essential for MV propagation, since mutant viral particles are rapidly lost in competition with wild-type viruses, strongly suggesting that they are noninfectious. Arguably, viral nucleosomes must be metastable to make the genome (once transferred into the cytoplasm) accessible to transcription by the host RNA polymerase recruited to the viral factory. This might explain the requirement for virus-encoded histones, as ATP-dependent remodelers utilized by eukaryotes to facilitate transcription through chromatin are localized mostly in the nucleus. The regions in MV histones that convey this metastability (such as the H4-H3 connector and the distinct makeup of the four-helix bundle regions) are highly conserved among *Marseilleviridae*. Why these histones exist as doublets, what determines their specificity for the viral genome, and which (if any) assembly factors they rely on for their association with the viral DNA are intriguing open questions that warrant further research.

Genetic manipulation of giant viruses has been a neglected topic of research since their discovery despite the immense richness in genes with unknown functions encoded by their genome. Homology recombination has been used as a powerful tool to modify the genome of other viruses, and we show here that it can be efficiently utilized for genetic manipulation of MV[73]. Moreover, the essentiality of genes that encode MV histones, like *mel_369* and *mel_368*, can be overcome by trans-complementation, introducing ectopic copies of the gene in the amoeba

genome. Future efforts will be directed to identify selection systems to isolate mutant viruses for phenotypic analysis.

2.4.1 Limitations of Study

Our study demonstrates that virally encoded histone doublets assemble into nucleosome-like particles *in vitro* and that they are present in the virus at amounts sufficient to package the entire viral genome into nucleosomes. However, we have no direct proof that the histones are indeed bound to viral DNA in the capsid. This is because any attempt to open the extremely stable virus capsid also destroys chromatin structure. However, the presence of strong electron density in the capsid core suggests that the genome is highly compacted. Methods such as cryo-electron tomography or correlative light electron microscopy (that do not require opening of the capsid) might be needed to visualize the organization of chromatin in the capsid. Although co-localization of the two histone doublets suggests that they are forming nucleosomes in the viral factory to which they are recruited, we have no direct proof of this.

The timing of the association of viral histones with DNA in the VF (in particular when nucleosomes are formed and which chaperones, if any, assist in the process) has not yet been investigated, as it requires an infectious virus whose genome encodes tagged histones. These exciting experiments have only now become possible due to our breakthrough in being able to genetically manipulate the virus.

Finally, visualizing nucleosome-like particles on the cryo-EM grid required crosslinking (as did a related study describing the structure of the MV nucleosome that was published while this article was under review[48]). It is possible that some aspects of the structure are more dynamic

than suggested in these two structures. The high mobility of the connectors, together with the overall resolution reported here (~4 Å), precluded a detailed description of the main chain in the connecting regions. It is therefore gratifying that the two independently determined structures are very similar, even in the ill-defined regions.

2.5 Materials and Methods

2.5.1 Histone sequence alignment and secondary structure prediction

Predicted *Marseilleviridae* histone-like proteins were aligned with eukaryotic histone proteins with HHpred's Multiple Alignment using Fast Fourier Transform (MAFFT) with a 1.53 gap open penalty. Lineages A, B, and C of the *Marseilleviridae* family were included, lineage D was excluded due to sequence similarity with lineage C (~90%) and lineage E was excluded due to the lack of annotated histone-like proteins. Using the MAFFT alignment, protein secondary structures were predicted using HHpred's Quick 2D structural prediction webserver to demonstrate structural conservation of the histone fold domain between Eukarya and Marseilleviridae[74].

2.5.2 Fluorescence localization of Melbournevirus histones in infected *Acanthamoeba castellanii* cells

The three Melbournevirus (AIT54904) histones encoding genes Mel-368 (MV-H4-H3), Mel-369 (MV-H2B-H2A) and Mel-149 (MV-miniH2B-H2A) were amplified by PCR from genomic DNA and cloned into an in-house modified plasmid derived from the pGAPDH-GFP amoebal expression plasmid[75]. Briefly, the GAPDH promoter was replaced by the stronger EF1 gene promoter and the

genes were inserted using the NdeI restriction site to yield C-terminally GFP-tagged proteins. The *A. castellanii* histone H2A (2) gene (ACA1_364730A) was amplified from Acanthamoeba castellanii (Douglas) Neff (ATCC 30010TM) genomic DNA and cloned into the same plasmid engineered to yield a N-terminally GFP-tagged protein.

A. castellanii cells were transfected with 6 mg of each plasmid using Superfect (QIAGEN). Selection of transformed cells was initially performed at 30 µg/mL Neomycin and increased up to 100 µg/mL within a couple of weeks. Transfected *A. castellanii* cells were grown on poly-L-lysine coating coverslips in a 12-well plate and infected with Melbournevirus at a MOI of 50 except for the negative control. At 30 min, 1h, 2h, 3h and 4h post-infection (pi), cells were fixed with PBS containing 3.7% formaldehyde for 20 min at room temperature. After one wash with PBS buffer, coverslips were mounted on a glass slide with 4 ml of VECTASHIELD mounting medium with DAPI and the fluorescence was observed using a Zeiss Axio Observer Z1 inverted microscope using a 63x objective lens associated with a 1.6x Optovar for DIC, DAPI or GFP fluorescence recording.

To perform the co-localization experiment, the previous plasmid was modified to replace GFP with mRFP. We screened several selection markers to allow the selection of the co-transfected cells and Nourseothricin was the only antibiotics highly efficient on Acanthamoeba cells. The Neomycin-resistance gene (neo) was thus replaced by the Nourseothricin-resistance gene (nat1) in the previous plasmid. Briefly, the plasmid was digested using NdeI and XbaI to remove the GFP and the linearized plasmid was gel purified. The mRFP gene was amplified and inserted by homologous recombination using the In-Fusion kit (Takara Bio) according to the manufacturer protocol. The subsequent plasmid was amplified by PCR to get rid of the Neomycin-resistance gene and Nourseothricin-resistance gene was inserted by homologous recombination. The

Melbournevirus gene encoding the H3-H4 histone was then cloned using the NdeI site to yield a C-terminally mRFP-tagged protein. The structure of the construct was confirmed by sequencing. *A. castellanii* cells expressing the MV-H2A-H2B-GFP proteins were transfected with 6µg of the MV-H3-H4-mRFP plasmid and selection of the co-transfected cells was performed using 100 µg/mL Neomycin and 30 to 50 µg/mL Nourseothricin. Cells were grown in a 96-well glass bottom plate and infected at a MOI of 50. At 4h pi, the cells were washed with PBS containing 10 µg/mL of DAPI and observed for DIC, DAPI, GFP or mRFP fluorescence recording.

2.5.3 Mass spectroscopy proteomic analysis

The mass spectrometry proteomic of the purified virions was performed as previously described[58]. Data are available at the ProteomeXchange Consortium (proteomecentral.proteomexchange.org) via the PRIDE partner repository with the accession code PXD003910. Based on the structure of Melbournevirus exhibiting a triangulation number T = 309 corresponding to 9240 Major Capsid Proteins, the copy numbers for each protein in the viral proteome was estimated using the computed intensity-based absolute quantification (iBAQ) values[59].

2.5.4 *Marselleivirus-Acanthamoeba castellanii* transcriptome

The time-course transcriptome analysis of *A. castellanii* cells infected by Marseillevirus was performed using the RNA-seq data from[40]. We first annotated the Marseillevirus genome (accession GU071086) as described, as well as the A. castellanii Neff host genome (assembly GCA_000193105.1) using the Funannotate pipeline (version 1.7; https://github.

com/nextgenusfs/funannotate)[76]. We then mapped the RNA-seq reads corresponding to 30 min pi (ERR3528397), 1 h pi (ERR3528398), 2 h pi (ERR3528399), 4 h pi (ERR3528400), 5 h pi (ERR3528401), 6 h pi (ERR3528402), 8 h pi (ERR3528403), 10 h pi (ERR3528404) and 12 h pi (ERR3528405) to the host and viral transcripts using Bowtie (version 2.3.4.1)[77]. Transcript quantification was done using RSEM (version 1.3.0) and data normalization using the edgeR R package[78,79]. Host and viral mapped reads were normalized separately with the ''TMMwsp'' method and then combined to a single matrix. Expression values were then expressed in log transformed counts per million of mapped reads (CPM). Finally, we scaled each transcript expression values to a Z-score and produced a heatmap for the genes of interest. A hierarchical clustering of the heatmap was obtained using a Pearson correlation distance and the ''Wards.D2'' method.

2.5.5 Knockout generation cloning strategies and analysis

Vectors for knockout of MEL_369 or MEL_368 were generated by inserting 500 base pairs homology arms into the pGAPDH-EGFP vector[75]. Homology arms were amplified from Melbournevirus genomic DNA produced by Wizard genomic DNA purification kit (PROMEGA) according to manufacturer's specifications. Primers used are presented in Table S2.1 (HB112-115 and HB120-123 for amplification of homology arms of MEL_368 or MEL_369, respectively), and inserted into the NdeI and XbaI sites by infusion (Takara). The vectors were digested with NdeI and XbaI (NEB) prior to transfection into *A. castellanii*. *A. castellanii* cells were transfected with 8 µg of each linearized plasmid using Superfect (QIAGEN) and infection by Melbournevirus was performed 1- hour post transfection. One well was used to analyze the genomic integration using primers

presented in Table S1.1 after 5-hours postinfection. Another well was kept to analyze the infectiousness of neo-synthesized viral particles. To remove the excess of viral particle used for infection, the growth medium was removed 1-hour post-infection and fresh growth medium was added to the culture. Washing of the cells was repeated 5 times. After lysis, the viral progeny was recovered and kept for additional round of infection.

Vector plasmids for trans-complementation were generated by directed mutagenesis using infusion (Takara) utilizing the primers HB155-HB160 to insert untagged genes into the pEF1-mRFP-NAT target vector. 6 mg of circular vectors were transfected into *A. castellanii* using Superfect (QIAGEN). Selection of transformed cells was initially performed at 30 µg/mL Nourseothricin and increased up to 100 µg/mL within a week post transfection. Competition assay were performed by passaging the recovered transfected viruses subsequently. Again, the presence of recombinant DNA was performed by PCR using infected cells 5-hours post infection. Extracellular virions were removed by washing 5 times. Genomic DNA was obtained by Wizard genomic DNA purification kit (PROMEGA) according to manufacturer's specifications.

2.5.6 MV histone doublet purification

Melbournevirus ORF 369 (H2B-H2A), ORF 368 (H4-H3) and ORF 149 (miniH2B-H2A) were cloned into pET-28 plasmid for expression and purification from E.coli, using protocols adapted from eukaryotic histones[60]. 6L of E. coli cells were re-suspended and lysed with 80 mL of guanidinium lysis buffer (6 M Guanidinium HCl, 20 mM sodium phosphate pH 6.8 and 2 M NaCl) and sonicated for 30 s at 50% strength for 4 times. Lysate was spun at 16,000 rpm for 25 minutes, and the supernatant was incubated with Nickel NTA Agarose Beads (GoldBio) for approx. 45

minutes at RT. The nickel beads were pelleted, and the supernatant was discarded. The beads were resuspended in 8 M Urea lysis buffer (8 M Urea, 20 mM sodium phosphate pH 6.8, and 2 M NaCl) and incubated for approximately 1 hour at room temperature.

Proteins were eluted using histone elution buffer (8 M Urea, 20 mM sodium phosphate pH 6.8, 2 M NaCl, and 500 mM Imidazole). For protein refolding, histones were dialyzed against 1 M Urea Buffer (1 M urea, 1 mM EDTA, 20 mM Tris-Cl pH 7.5 and 2 M NaCl) for 6 hr, followed by overnight dialysis against the same buffer without urea (1 mM EDTA, 20 mM Tris-Cl pH 7.5 and 2 M NaCl). Precipitated protein was removed by centrifugation, and the supernatant was concentrated and run over a size-exclusion column S200 equilibrated in 1 mM EDTA, 20 mM Tris-Cl pH 7.5 and 2 M NaCl. Samples were stored in 20% glycerol at –80°C.

2.5.7 MV histone-DNA complex (nucleosome) reconstitution

Melbournevirus H2B-H2A, H4-H3 and DNA of the specified lengths were mixed at a ratio of DNA to MV-H4-H3 to MV-H2B-H2A = 1.0: 2.4: 2.4. Widom 601 DNA at 147 bp and 207bp, and 181 bp of a conserved Melbournevirus native DNA fragment selected from its genome (GC content = 45%) were used. Viral NLPs and X.laevis nucleosomes were both reconstituted by gradient dialysis[60]. Reconstituted products were analyzed by 5% native-PAGE.

2.5.8 Sucrose gradient sedimentation and gradient fixation (GraFix) crosslinking

To form the sucrose gradient, 6 ml top solution (50 mM NaCl, 20 mM HEPES, 1 mM EDTA, pH 7.5 and 5% sucrose) was added to a tube (Beckman, 331372). Then 6 mL bottom solution (50 mM NaCl, 20 mM HEPES, 1 mM EDTA, pH 7.5 and 40% sucrose) was slowly added to the bottom

until it reached the halfway mark. The tubes were placed on a gradient maker (BioComp Gradient Master) to form a continuous gradient. 200 ml of sample at 4 mM was loaded on top and spun at 4 C for 18 h at 30,000 rpm (Beckman, Rotor SW41Ti). Fractioned samples were dialyzed against 50 mM NaCl, 20 mM Tris-Cl, 1 mM EDTA, pH 7.5 and 1 mM DTT to remove the sucrose.

For GraFix, the same procedure was applied, except that 10% to 30% of glycerol was used instead of sucrose. In addition, 0.15% of glutaraldehyde was added to the bottom solution to form the continuous density and crosslinker gradient. After GRAFIX, the samples were eluted and dialyzed against 50 mM NaCl, 20 mM Tris-Cl, 1 mM EDTA, pH 7.5 and 1 mM DTT to quench the crosslinking reaction and remove glycerol.

2.5.9 Sedimentation velocity analytical ultracentrifugation (SV-AUC)

SV-AUC with absorbance optics (λ = 260 nm) was used to evaluate the homogeneity of the complexes in solution and to infer their molecular size and shape. Samples (composition as indicated) at 300 nM were spun at 30–35,000 rpm at 20°C in a Beckman XL-A ultracentrifuge in 50 mM NaCl, 20 mM Tris-Cl, 1 mM EDTA, pH 7.5 and 1 mM DTT, using an An60Ti rotor. Partial specific volumes of samples were estimated using UltraScan III version 4.0; this program was also used for all data analysis[80,81].

Time-invariant and radial-invariant noise contributions were subtracted from the experimental sedimentation velocity data by 2- dimensional Spectrum Analysis (2DSA), followed by refinement using the Genetic Algorithm-Monte Carlo (GA-MC) approach[82]. Sedimentation coefficients and molecular weights in Table 2.1 were extracted from resulting GA-MC models. Modeling calculations were performed on the UltraScan LIMS clusters at the Bioinformatics Core

Facility at University of Texas Health Science Center at San Antonio and Lonestar cluster at Texas Advanced Computing Center. Integral sedimentation coefficient distributions (G(s)) were obtained from noise-corrected experimental data with the Enhanced van Holde-Weischet Analysis method. G(s) plot figures were created using python (https://github.com/Luger-Lab/AUC-analysis)[83].

2.5.10 Atomic force microscopy (AFM)

All samples were imaged in air on JPK/Bruker NanoWizard Sa with TAP300-Gold (Ted Pella) cantilevers using fast scan glass block. Images were usually scanned at 1.5 3 1.5 mm at 1- 3 Hz. All buffers were filtered through 0.02 µM filters (Anotop) before use. Mica was freshly cleaved, treated with APTES for 30 min, rinsed with water and then dried with nitrogen gas utilizing a 0.22 µm PES filter. Samples were serially diluted in TE (20 mM Tris-HCl pH 7.5, 1 mM EDTA) to a final concentration of 2 nM. Within 2 minutes of final dilution, 30 µl sample was placed on the ATPES mica slide for 2 minutes, rinsed with 1.0 mL of water and dried with filtered N2 gas. Images were analyzed using JPK SPM software V 6.1.158. Each 1.5 3 1.5 mm image was divided into 4 quadrants and digitally zoomed to obtain clearer visuals of separated particles. Lines were drawn and dragged through each particle with the greatest height of each recorded. Analysis of heights were completed in either MS Excel and/or GraphPad Prism.

2.5.11 Single particle cryo-electron microscopy (cryo-EM) and data processing

Complexes of MV-NLPs from GraFix were concentrated to 2-3 µM using Amicon Ultra-4 centrifugal filters (Ultracel 50K, Millipore). CFlat 1.2/1.3 (Au) grids were glow discharged (EMItec, Lohmar, DE) at 40 mA for 30 s. 4 µl sample was applied onto the grid before manual plunge freezing

into ethane. Images for MV-NLP with 147 bp ‘‘601’’ DNA were acquired at nominal magnification of 64000x on a FEI Titan Krios (300 kV), equipped with a Gatan K3 Summit direct detector. Pixel size was 1.065 Å. The movies were captured in super resolution mode with electron dose rate at 10.5 electrons per pixel per second for 5.4 s and 0.108 s per frame. Defocus range was –0.8 to –2.0 mm. Images for MV-NLP with 207 bp ‘‘601’’ DNA were acquired at nominal magnification of 29000x on a FEI Tecani F20 (200 kV), equipped with a Gatan K3 Summit direct detector. Pixel size was 1.219 Å. The movies were captured in super resolution mode with electron dose rate at 10 electrons per pixel per second for 8 s and 0.2 s per frame. Defocus range was –1.0 to –2.5 mm.

Both datasets were processed (Motion correction and CTF estimation) by cryoSPARC (v2.12.4)[84,85]. Images were evaluated by inspecting the CTF fitting resolution. Roughly 1000 particles were then manually picked for generating autopicking templates through 2D classification in CryoSPARC. For MV-NLP with 147bp DNA, this yielded 2,648,493 particles across 6,552 micrographs, and these particles were subjected to four iterative rounds of 2D classification in order to discard bad particles. Two *ab initio* models were then generated, followed by heterogeneous refinement of the nucleosome-like class (34,418 particles). Non-uniform refinement, as well as global and local CTF refinements, were performed to improve the density to the final resolution, as estimated by GFSC. For MV-NLP with 207bp DNA, template-based particle picking yielded ~2,131,563 particles across 3,495 micrographs, and these particles were subjected to four iterative rounds of 2D classifications to get the best nucleosome particles. Two ab initio models were then generated, followed by heterogeneous refinement of the nucleosome like class. Non-uniform refinement with 377,702 particles, were performed to improve the density to the final resolution, as estimated by GFSC.

Initial models were built by fitting MV histone homology models (see below) into the final 3D electron maps in UCSF Chimera, yielding an initial correlation coefficient of 0.7086[86]. These models were then iteratively modified and locally refined in COOT[87].Molecular dynamics flexible-fitting (MDFF, details below) was then used utilized to simultaneously balance correlation with the EM density and atomic-level energy evaluation (electrostatic configurations and removal of steric overlaps).

2.5.12 Homology modeling

Initial homology models of Melbournevirus ORF 369 (H2B-H2A) and ORF 368 (H4-H3) doublets were constructed using SWISSMODEL, and the histone structures from the *X. laevis* nucleosome (PDB 1AOI) were used as a reference[65,88,89]. Using Modeler (v9.20), 10,000 loops of each doublet (H2B-H2A Y152 to S192 and H4-H3 F122 to T203, with sequence numbers identified as in Figure S1A) were generated[90]. To eliminate improbable loops within each doublet, clash identification was performed using CPPTRAJ of the Amber MD package (v18), where a cutoff distance of 0.8 Å between potentially over-lapping atoms was used[91]. Any loops identified by CPPTRAJ without a clash were manually parsed through to eliminate loops that were not sterically overlapping but also not physically relevant, such as those forming ‘‘knots’’ in the histone folds. While several configurations with reasonable qualitative agreement to the EM density were identified (**Figure S2.4**), the loop with the overall best correlation with the density was selected as the initial conformation for further refinement via COOT, PHENIX, and MDFF (see below)[92–94]. Figures were rendered using Chimera and VMD[95].

2.5.13 Molecular dynamic flexible fitting (MDFF) protocol

MDFF simulations were initialized from the PHENIX-refined structure based upon our initial homology modeling. This method improves fitting to EM data by introducing external forces, which bias the structure from its initial conformation into the empirical density. Simulations were conducted using the GPU-enhanced NAMD engine (v2.13) with the CHARMM36 forcefield in an implicit solvent environment, biasing forces were applied to the protein backbone and heavy atoms of the DNA, in accordance with the level of resolution provided by the density[96,97]. The simulation timestep was 2 fs, and a total simulation time of 5 ns was sufficient for convergence into the empirical density. Distance- and angle-based restraints of 20 kcal/mol/ $Å^2$ and 20 kcal/mol/degree^2 were applied to structured regions to prevent over-fitting of the density by breaking of domains with known secondary structure, and dihedral restraints of 50 kcal/mol/degree^2 were applied to maintain proper chirality during the biasing process. Over the course of the simulation, correlation between the model and experimental density only showed a modest improvement from 0.761 to 0.766, but significant benefits to residue conformations were observed, such as the removal of drastic Φ and Ψ orientations or the repositioning of sidechains that were originally used to explain main chain densities due to overfitting by real-space refinements in PHENIX.

2.5.14 Structural characterization of the MV-NLP$_{147}$ model

Comparisons between MV-NLP$_{147}$ and eNuc$_{147}$ (PDB 3LZ0); were conducted using VMD, and the protein backbone (N, C, and Ca) and DNA sugar ring (C1', C20 , C30 , C4', and C50) atoms were used for RMSD and center-of-geometry calculations (**Table S2.3**)[98]. To minimize discrepancies

between each system, the $eNuc_{147}$ molecule was least-squares fit to the MV-NLP_{147} using the core α helices and central 121 bp of DNA as a reference. In this way, RMSD values represent the changes in the global positioning of each constituent, rather than measuring the conformational rearrangement within each individual piece, which was altogether quite low (typically ~1-2 Å). MV-H2B-H2A dimer reorientation, relative to the eukaryotic H2A-H2B dimer, was determined by placing fictitious particles at the geometric centers of the first and last turns of the H2B α2 helix and calculating the dihedral angle describing this imaginary atom arrangement (Figure S2.4A). In the $eNuc_{147}$ system, the fictitious atoms describing the C-terminal end of each helix were not located at the final turn of the helix, but rather at the turn within the helix that would yield an analogous length of the H2B α2 helix in MV-NLP_{147}, to provide the best direct comparison of the two systems. Larger magnitude values in this measurement correspond to dimers that are angled further away from the dyad-axis, as well as from one another.

2.6 Supplemental Materials

Figure S2.1A

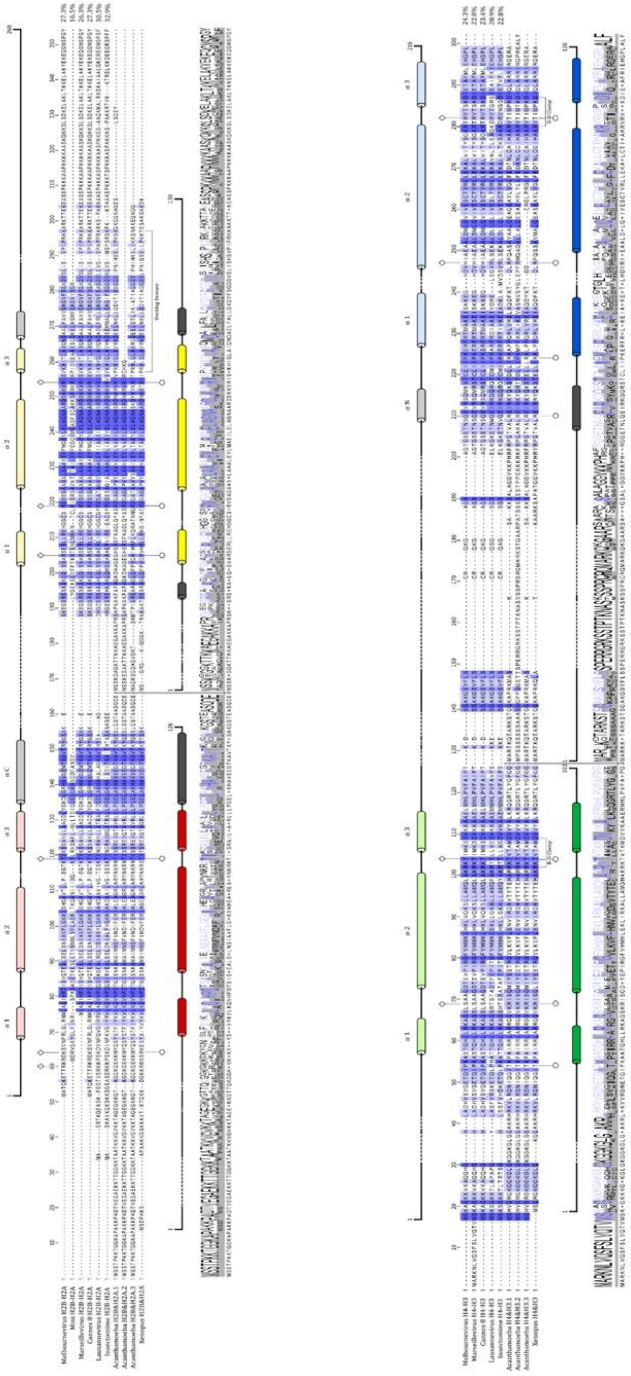


Figure S2.1B

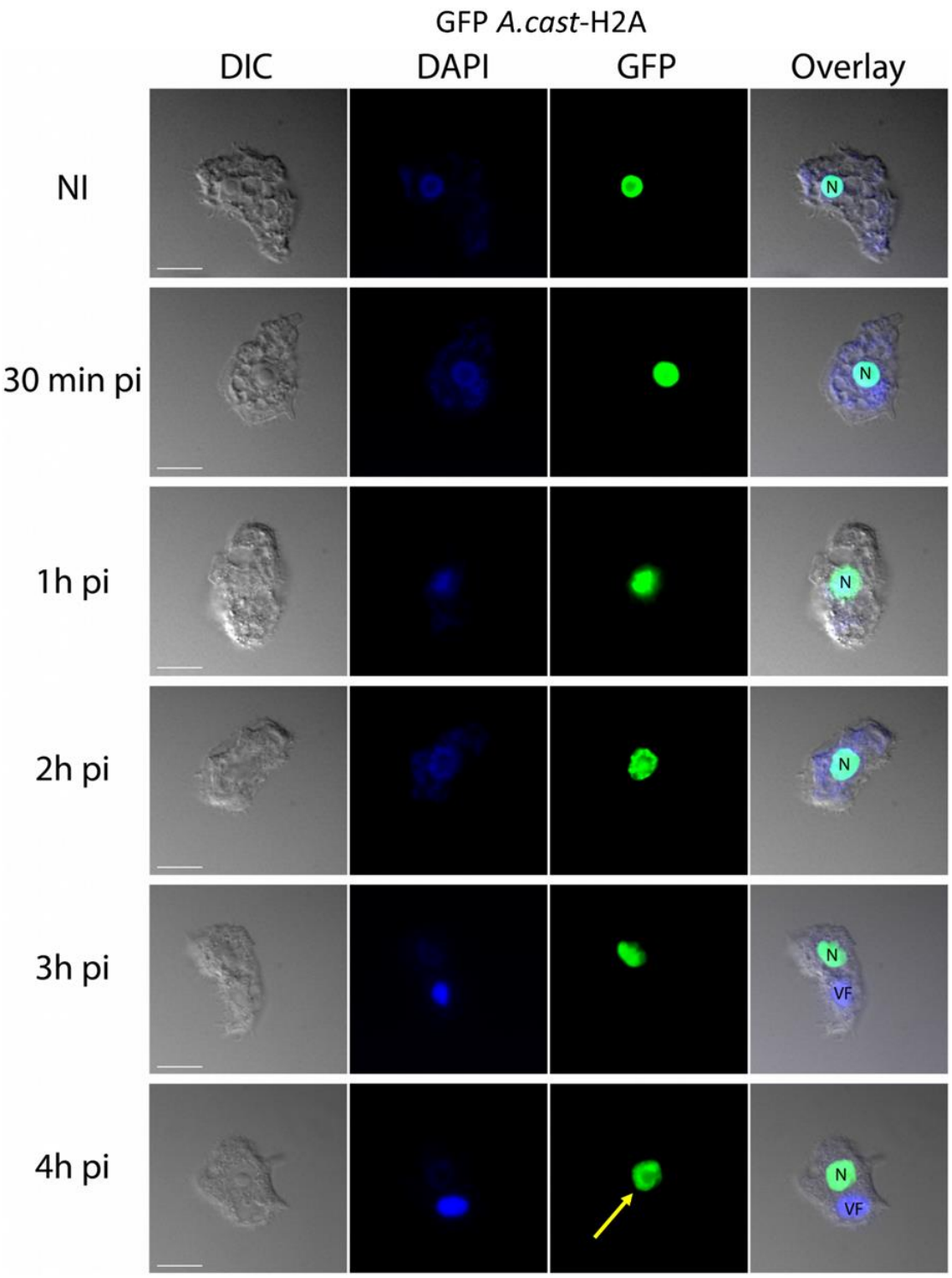

GFP A.cast-H2A
DIC
DAPI
GFP
Overlay
NI
30 min pi
1h pi
2h pi
3h pi
4h pi
N
VF

Figure S2.1C

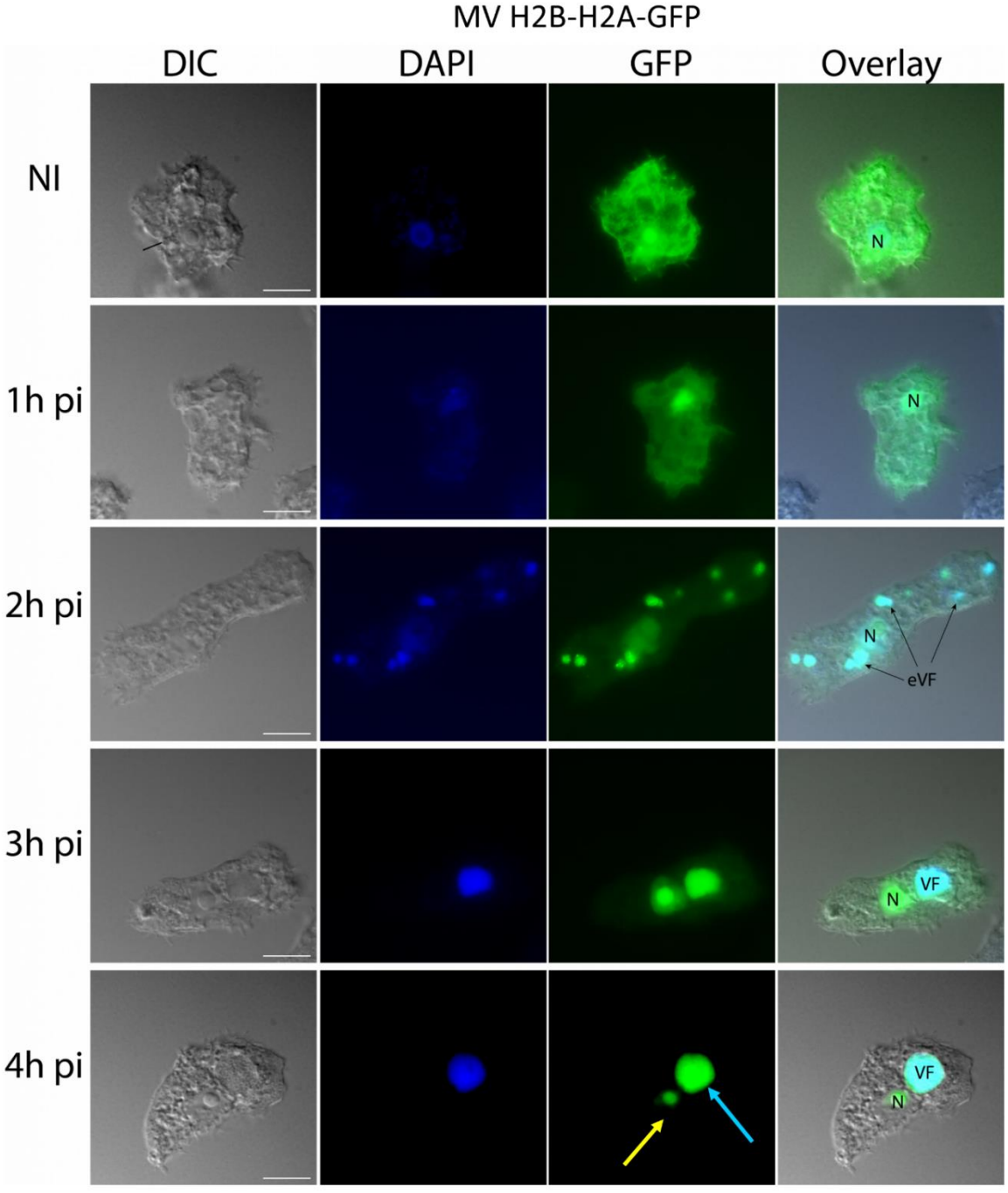

MV H2B-H2A-GFP
DIC
DAPI
GFP
Overlay
NI
1h pi
2h pi
3h pi
4h pi
N
N
N
eVF
N
VF
N
VF

Figure 2.1D

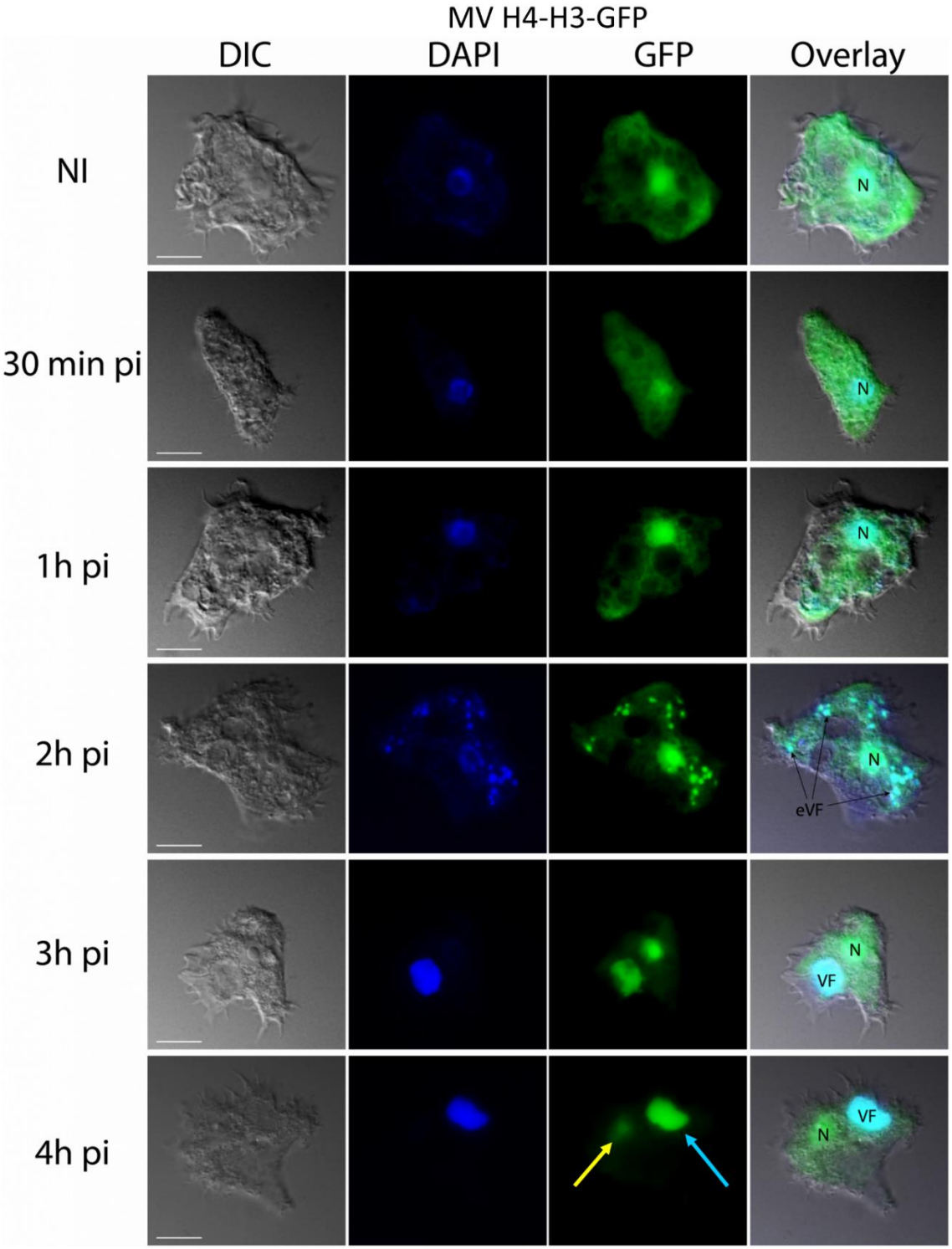

MV H4-H3-GFP
DIC
DAPI
GFP
Overlay
NI
30 min pi
1h pi
2h pi
3h pi
4h pi
N
eVF
VF

Figure S2.1E

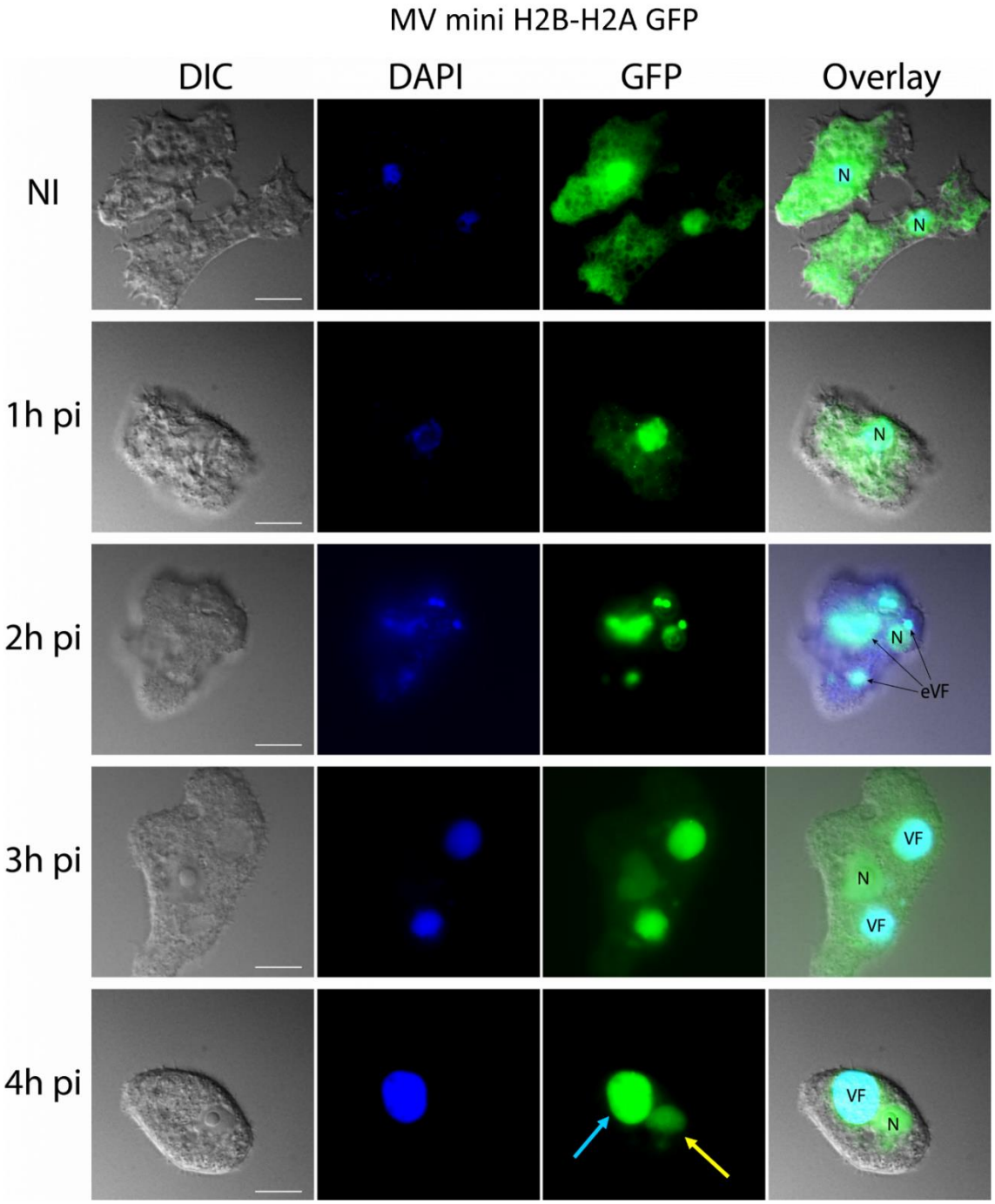

MV mini H2B-H2A GFP
DIC
DAPI
GFP
Overlay
NI
1h pi
2h pi
3h pi
4h pi
N
N
N
N
eVF
VF
N
VF
VF
N

Figure S2.1F

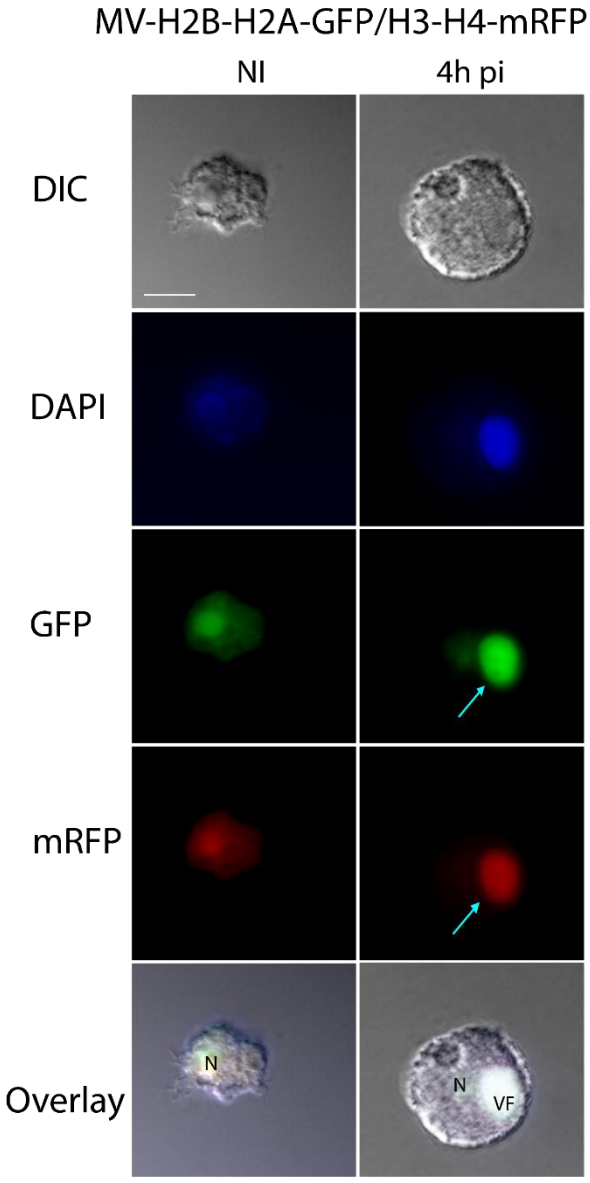

MV-H2B-H2A-GFP/H3-H4-mRFP
NI
4h pi
DIC
DAPI
GFP
mRFP
Overlay
N
N
VF

Figure 2.1G

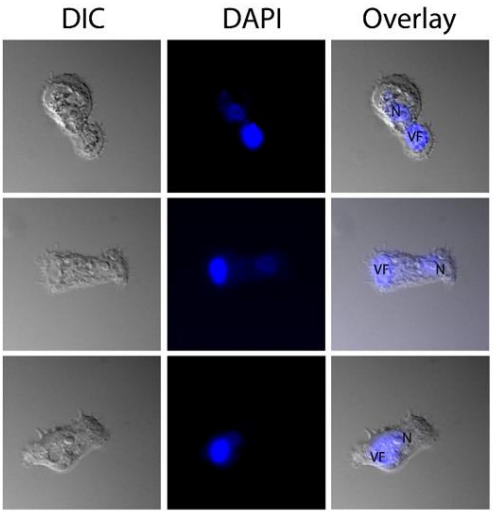


Figure 2.1H

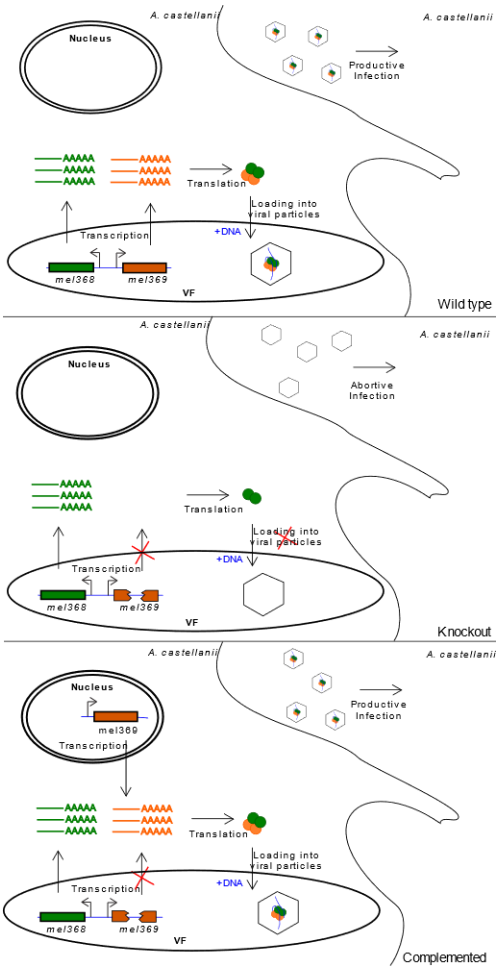


Figure. S2.1 Localization of fluorescently labeled Melbournevirus histones in virus-infected Amoeba.

(A) Complete sequence alignment of H2B-H2A (top) and H4-H3 (bottom) from Marseilleviridae, the host Acanthamoeba castellanii, and Xenopus laevis. Predicted α helices of H2B-H2A (light red and yellow) and H4-H3 (light green and blue) Melbournevirus histone doublets were generated using HHPRED's Quick 2D prediction web server. Histone dimer pairs H2B-H2A and H4-H3 of Acanthamoeba castellanii and Xenopus laevis were each aligned against their respective Marseilleviridae histone doublets using HHPRED's multiple sequence alignment tool, ClustalΩ. Conservation of each specific residue in each alignment is denoted by blue shading, with greater conservation being represented by darker blue. Known R-D clamp, R-T pairs, and DNA binding residues are indicated for Xenopus laevis histone pairs with their conservation within Melbournevirus histones. Marseilleviridae histones are 16-33% conserved to Xenopus laevis histones (shown on far right of alignment). Known α helices from the histone fold domain in Xenopus laevis are shown as dark colored tubes; H2B are red, H2A are yellow, H4 are green, H3 in blue, and additional helices in grey. Logo plot demonstrating residue conservation among the alignments provided by ClustalΩ tool is shown below. Light microscopy fluorescence images (scale bar 10 µm) of A. castellanii cells transfected with B) GFP-A. castellanii-H2A, C) MV-H2B-H2A-GFP, D) MV-H4-H3-GFP and E) MV-miniH2B-H2A-GFP, non-infected (NI) and infected with Melbournevirus at 30 min, 1h, 2h, 3h and 4h pi. A) GFP-A. castellanii-H2A concentrates only in the nucleus (N) of the non-infected cells and all along the infectious cycle. C) MV-H2B-H2A-GFP, D) MV-H4-H3-GFP and E) MV-miniH2B-H2A-GFP are scattered in the entire cell (including the nucleus) in the non-infected cells. Between 1h and 2h pi, the viral histones start accumulating in the early viral factories (eVF). At 4h pi, the fluorescence is predominantly concentrated in the mature viral factory (VF). DAPI staining remains at the nucleus all along the infection but the intense fluorescence in the late VF hides the staining of the nucleus at 4h pi. F). MV H2B-H2A-GFP and MV H4-H3-mRFP are co-localized in the viral factory, along with the viral DNA at 4h pi. G). DAPI staining of Acanthamoeba cells infected with Melbournevirus at 4h pi. The host nucleus does not disappear upon the infectious cycle. Due to the large accumulation of viral DNA in the viral factories, exposure time for the DAPI fluorescence was optimized to visualize both the viral factory (VF) and the nucleus (N) of the cell (typically between 400ms and 1s). H) Schematic depicting the strategy for analysis of histone gene knockouts.

Figure S2.2 A, B, C, D

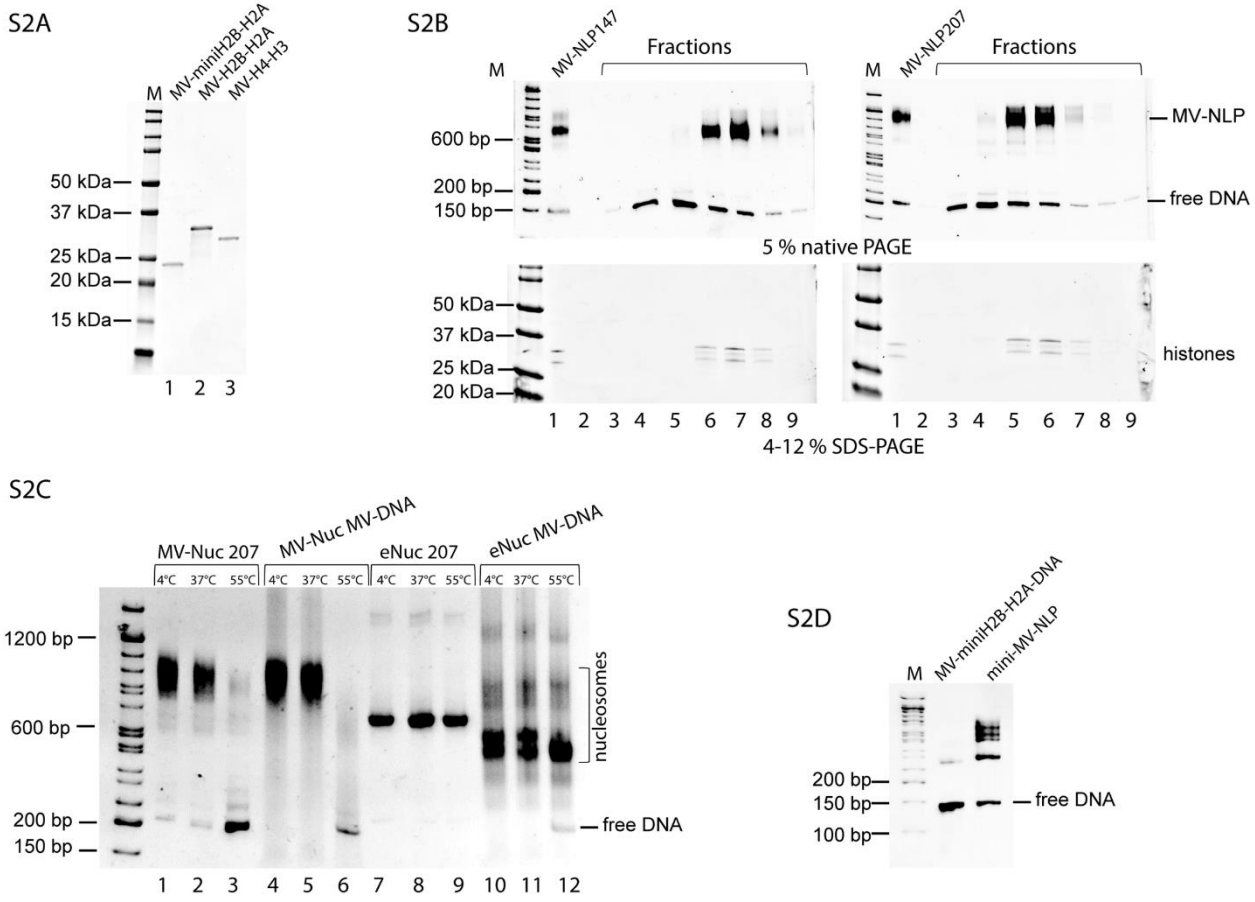


Figure 2.2 E, F

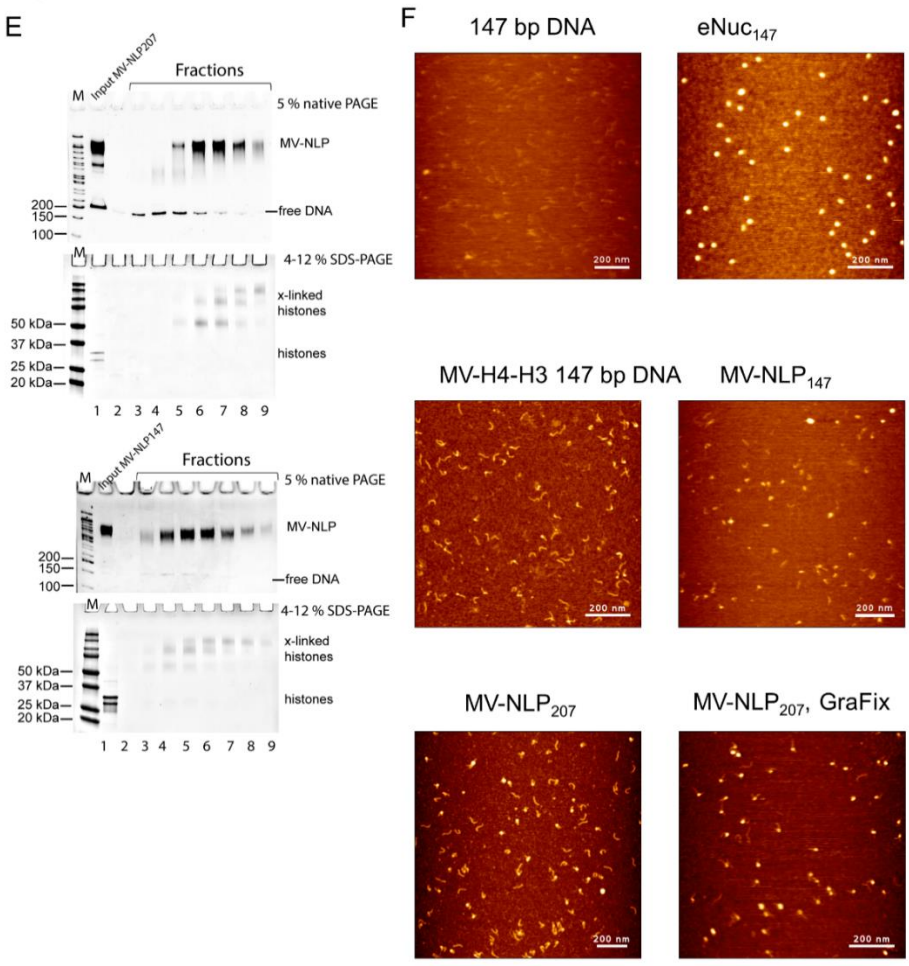


Figure. S2.2 MV-histones form nucleosome-like particles.

(A) SDS-PAGE of purified Melbournevirus (MV) histone doublets. (B) Sucrose gradient sedimentation of MV NLPs with 147 or 207 bp DNA. The compositions of MV NLPs were analyzed by native- and SDS-PAGE. (C) MV-NLP and eNuc were reconstituted on Widom '601' DNA and Melbournevirus native DNA, respectively. The MV-NLP and eNuc were heat treated at 37 and 55 °C. (D) Native PAGE of reconstituted MV mini-NLP (mini H2B-H2A instead of H2B-H2A) with 147 bp DNA. (E) GraFix of MV-NLPs with 207 bp DNA (top two panels) and 147 bp DNA (bottom two panels. Native- and SDS- PAGE of the crosslinked MV-NLP fractions representing successful crosslinking, compared to the native MV-NLP input. (F) Representative AFM images: Samples were diluted in TCS buffer and applied to APTES coated mica, rinsed with water and imaged on a NanoWizard Sa with a TAP300-GD cantilever. Samples include 147 bp DNA only, $eNuc_{147}$, MV-H4-H3 with 147 bp DNA, $MV\text{-}NLP_{147}$, $MV\text{-}NLP_{207}$ and $MV\text{-}NLP_{207}$ GraFix.

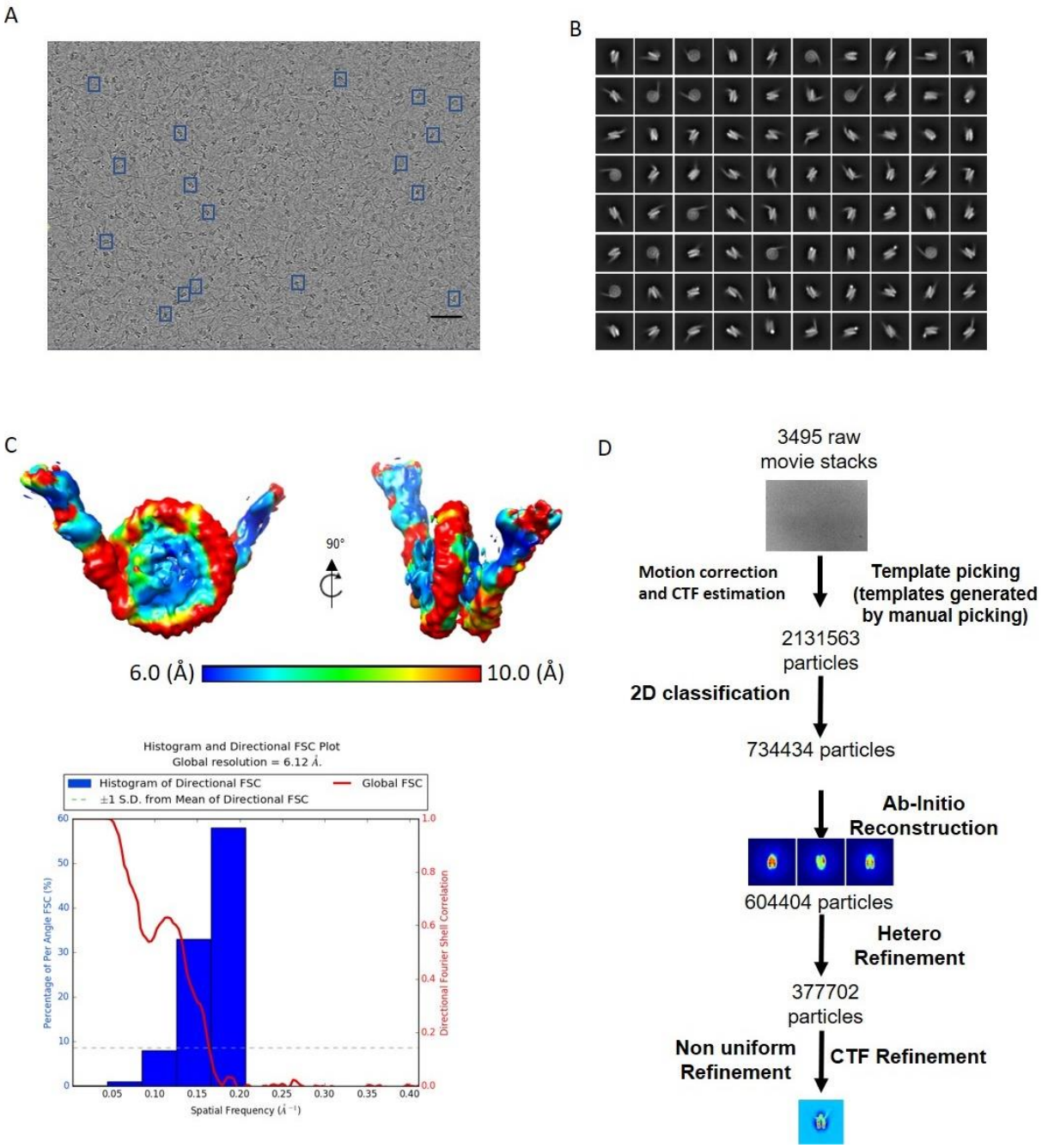


Figure. S2.3 Cryo EM data analysis of MV-NLP207.

(A) Raw micrograph of the Grafix-treated MV-NLP$_{207}$. Scale bar is 50 nm. (B) 2D class averages generated from the dataset. (C) 3D structure of the MV NLP. Local resolution map and FSC curve. (D) Data processing strategy flow chart.

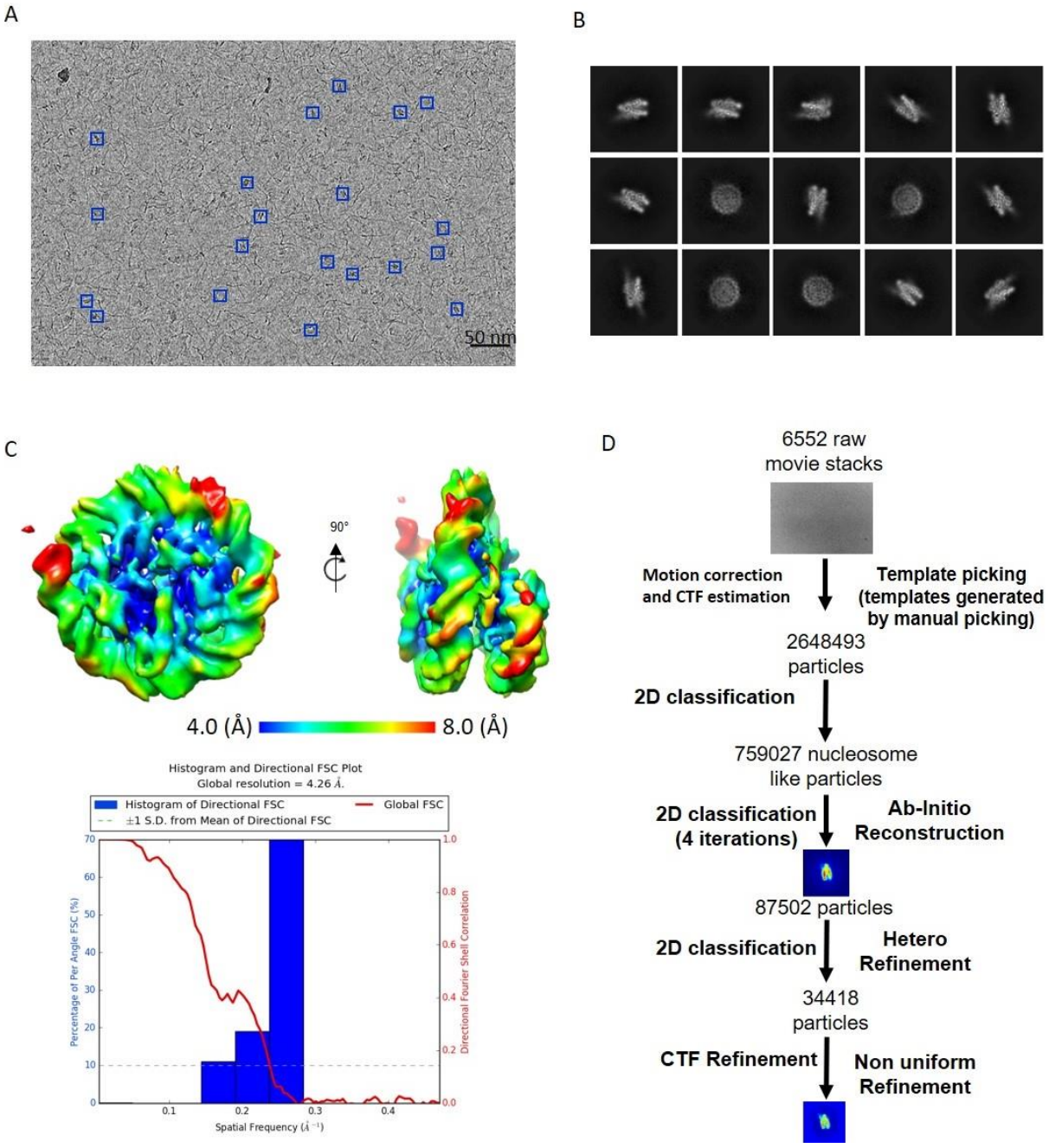


Figure. S2.4 Cryo EM data analysis of MV-NLP147.

(A) Raw micrograph of the GraFix treated MV-NLP$_{147}$. Scale bar is 50 nm. (B) 2D class averages generated from the dataset. (C) 3D structure of the MV NLP, and local resolution map and FSC curve. (D) Data processing strategy flow chart.

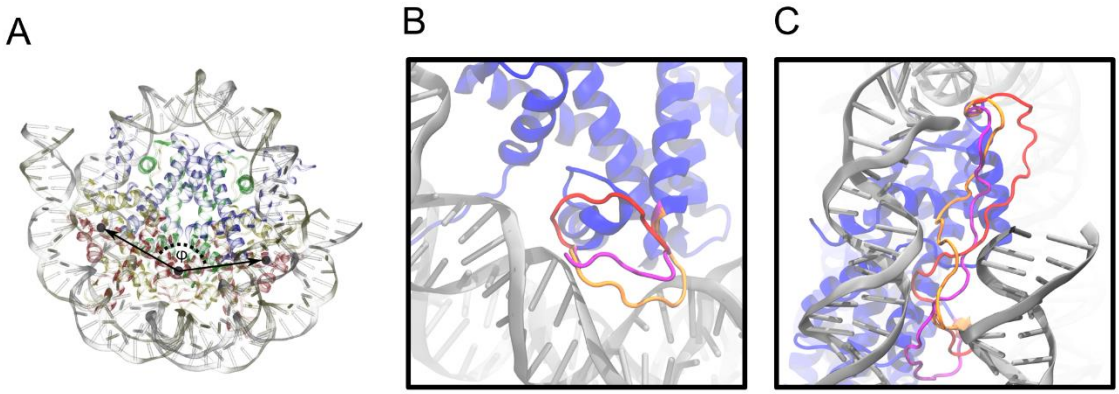


Figure. S2.5. MV-nucleosome histone dimer arrangement and histone connectors.

(A) Description of ϕ, the pseudo-dihedral angle measuring the relative orientations of the H2B-H2A subunits, between MV-NLP and eNuc. Larger magnitude values correspond to a dimer arrangement that is angled further away from the dyad axis. Fictitious particle locations are shown as black spheres, and histones (H3 in blue, H4 in green, H2A in yellow, and H2B in red) and DNA (silver and gold ribbons) are semi-transparent to allow for better visibility of each particle's location. (B) Overlay of de novo loop models with feasible conformations connecting H2B to H2A and (C) H4 to H3 in the MV-histone doublets. Loop configurations were generated using Modeller (v9.20) and then placed in the nucleosome context. After manually removing physically non-relevant (knots) conformations and clashes identified with a cutoff distance of 0.8 Å using CPPTRAJ of the Amber MD package (v18), the three best loops are shown in different colors (cross correlation range against the simulated map is 0.614-0.623).

Table S2.1. Primers used in the MV histone knockout generation cloning and analysis.

Primer ID	Purpose	Sequence
HB112	5' homology arm for recombination 368	tctcactaagacatatgacagcgtgccttccaaggt
HB113	5' homology arm for recombination 368	gcccttgctcaccatgacaaaaccgcggatggga
HB114	3' homology arm for recombination 368	cgcgtggtacctctaaaggccgccgagctcatgc
HB115	3' homology arm for recombination 368	ccgggtcgactctagagttggatttgaagtttttccagtcg
HB120	5' homology arm for recombination 369	tctcactaagacatacgagctgcatcgcaagaa
HB121	5' homology arm for recombination 369	gcccttgctcaccatagatttgatggttttagttccctct
HB122	3' homology arm for recombination 369	ttcacgcgtggtaccaggaccttggaaggcacgc
HB123	3' homology arm for recombination 369	ccgggtcgactctagattagtatccggggctcattcc
HB145	Histone 368 KO genotyping	Cgttttcctcagcgaggaggaag
HB146	Histone 368 KO genotyping	Ctggtgacggccttggtc
HB147	Histone 369 KO genotyping	Atgcatgagctcggcggcc
HB148	Histone 369 KO genotyping	Cagaggacgatgcttgttgtctcgtc
HB155	pEF1 (Nourseo) for transcomplementation	Taatctagagtcgacccggaac
HB156	pEF1 (Nourseo) for transcomplementation	Atgggtggattggtggtgttg
HB157	pEF1-mel368 (Nourseo) for transcomplementation	caccaatccacccatatgtcaaaagcaggcaaaaagg
HB158	pEF1-mel368 (Nourseo) for transcomplementation	gtcgactctagattattatagcggaccatgttccagc
HB159	pEF1-mel369 (Nourseo) for transcomplementation	caccaatccacccatatggcgacacaaaaggaaaccac
HB160	pEF1-mel369 (Nourseo) for transcomplementation	gtcgactctagattattagtatccggggctcattccttg
GFP Fo	Histone KO genotyping	Aggtgaacttcaagatccgc
GFP Rev	Histone KO genotyping	Gaacttgtggccgtttacgtc

Table S2.2 Analysis of AFM data.

Samples were imaged by AFM with heights (nm) of nucleosome like particles (NLP) recorded separately from free DNA. Average heights and standard deviations were calculated from the number of incidences (n). In addition to the average height of the NLP, the amount of free DNA (as a percentage of particles on the slide) is a good indication of the stability of MV-NLP. When not prepared by GraFix, NLP heights generally correspond with heights of the MV-Tetrasome (p-values from unpaired t-test), whereas GraFix preparation yields heights in agreement with the $eNuc_{147}$ system.

	MV-H4-H3-DNA (MV-tetrasome)		*$eNuc_{147}$*		*$MV\text{-}NLP_{147}$*		*$MV\text{-}NLP_{207}$*		*$MV\text{-}NLP_{207}$ GraFix*	
Particle Class	Free DNA	MV-Tetrasome	Free DNA	eNuc	Free DNA	MV-NLP	Free DNA	MV-NLP	DNA	MV-NLP
Average Height (nm)	0.60	1.35	0.61	2.09	0.50	1.21	0.78	1.55	0.64	2.25
Heights σ (nm)	0.07	0.33	0.21	0.52	0.12	0.91	0.08	0.40	0.19	0.63
n	25	26	2	41	29	46	41	37	31	120
% Particle	43.9	45.6	4.3	87.2	32.6	51.7	50.6	45.7	20.0	77.4
p-value (MV-Tet)	---	---	---	< 0.0001	---	0.4528	---	0.0403	---	< 0.0001
p-value ($eNuc_{147}$)	---	< 0.0001	---	---	---	< 0.0001	---	< 0.0001	---	0.1452

Table S2.3. Summary of cryoEM data collection and refinement.

	MV-NLP$_{147}$ (EMDB-24238) (PDB 7N8N)	MV-NLP$_{207}$
Data collection and processing		
Magnification	64,000	29000
Voltage (kV)	300	200
Electron exposure (e–/Å^2)	50	80
Defocus range (μm)	0.8-2.0	1.0-2.5
Pixel size (Å)	1.065	1.219
Symmetry imposed	C1	C1
Initial particle images (no.)	2,648,493	2,131,563
Final particle images (no.)	34,418	377,702
Map resolution (Å)	4.17	6.1
FSC threshold	0.143	0.143
Map resolution range (Å)	3.9-6.9	6.0-10.0
Refinement		
Initial model used (PDB code)	1AOI, 3LZ0	
Model resolution (Å)	4.1	
FSC threshold	0.143	
Model resolution range (Å)	4.1-7.3	
Map sharpening *B* factor (Å^2)		
Model composition		
Non-hydrogen atoms	11125	
Protein residues	789	
Nucleotide	250	
Ligands	0	
B factors (Å^2)		
Protein	n/a	
Nucleotide	n/a	
Ligand	n/a	
R.m.s. deviations		
Bond lengths (Å)	0.006	
Bond angles (°)	0.840	
Validation		
MolProbity score	2.38	
Clashscore	22.42	
Poor rotamers (%)	0.00	
Ramachandran plot		
Favored (%)	90.38	
Allowed (%)	9.62	
Disallowed (%)	0.00	

Table S2.4. Comparison of MV-NLP147 with eNuc147.

"Global" root-mean-squared deviation (RMSD) values represent changes in positioning throughout the nucleosome-like complex and are calculated for each domain after least-squares fitting the cores of the MV-NLP$_{147}$ and eNuc$_{147}$ complexes, whereas "local" RMSDs are values calculated when least-squares fitting individual domains prior to calculation and thereby represent the degree of internal re-ordering of the domain. Distances are calculated as the separation of geometric centers of backbone positions (C, C_α, and N atoms), except for the H2A L1 Loops "closest distance", which defines the smallest distance between heavy atoms in the two moieties. The H2B α_C orientation angle is calculated around the pseudo-dihedral defined in Figure S2.5A.

RMSD Values (Å)	Global	Local
Particle Core	2.7	---
DNA (central 120 bp)	3.1	3.1
Octamer Fold	2.3	2.3
H4-H3 Tetramer	2.1	1.2
H3 α_N-helix (Chain A)	2.3	0.3
H3 α_N-helix (Chain E)	2.2	0.3
H2B-H2A Dimers (both)	2.5	2.1
H2B-H2A Dimer (Chain C,D)	2.5	1.6
H2B-H2A Dimer (Chain G,H)	2.5	1.6
H2B α_C-helix (chain D)	2.6	0.5
H2B α_C-helix (chain H)	3.4	0.9
H2A Docking Domain (Chain C)	8.5	3.9
H2A Docking Domain (Chain G)	6.2	3.3
Separation Distances (Å)	**eNuc$_{147}$**	**MV-NLP$_{147}$**
H4-H3 Dimers	32.6	33.3
H3-H3' 4HB Interface	10.1	10.4
H2B-H2A Dimers	36.3	38.6
H2B α_C-helices	49.1	52.2
H2A L1 Loops (centers)	7.2	9.7
H2A L1 Loops (closest distance)	2.9	5.3
Orientation Angle (°)	**eNuc$_{147}$**	**MV-NLP$_{147}$**
H2B α_2-helices	135.4	143.9

Movie S2.1

CryoEM electron density map of MV-NLP_{207} reveals its overall nucleosome-like shape and defined extended linker DNA, which includes a vertical and horizontal dyad axis rotation of MV-NLP_{207} with a density flare.

Movie S2.2

CryoEM electron density map of MV-NLP_{147} reveals its overall nucleosome-like shape, which includes a vertical and horizontal dyad axis rotation of MV-NLP_{147} with a density flare.

Movie S2.3

Overview of MV-NLP_{147} structure (model to map, **Fig 2.3A**), including a highlight of doublet connectors and H4 N-term domain (**Fig. 2.4**). The equivalent regions of H3, H4, H2A, and H2B are shown in blue, green, yellow, and red, respectively.

Chapter 3

The four *Medusavirus medusae* core histones form nucleosome-like structures with distinct features

3.1 Summary

The organization of DNA into nucleosomes is a ubiquitous and ancestral feature that was once thought to be exclusive to the eukaryotic domain of life. Intriguingly, several representatives of the Nucleocytoplasmic Large DNA Viruses (NCLDV) encode histone-like proteins that in Melbournevirus were shown to form nucleosome-like particles. *Medusavirus medusae* (MM), a distantly related giant virus, encodes all four core histone proteins and, unique amongst most giant viruses, a putative acidic protein with two domains resembling linker histone H1. Here we report the structure of nucleosomes assembled with MM histones, and highlight similarities and differences compared to eukaryotic nucleosomes and Melbournevirus nucleosomes. Our structure provides insight into how variations in histone tail and loop lengths are accommodated within the context of the nucleosome. We show that MM-histones assemble into tri-nucleosome arrays, and that the putative linker histone H1 does not function in compaction. These findings expand our understanding of viral histones and suggest that MM histones represent an instance in the evolutionary timeline of nucleosome architecture.

3.2 Introduction

Within the domain of eukaryotes, the compaction of genomic DNA by histones to form nucleosomes is an omnipresent and ancestral feature. The eukaryotic nucleosome core contains four unique histones (H2A, H2B, H3 and H4), each consisting of a structurally conserved histone fold (HF) that is common to all four core histones, as well as HF extensions (or loops) and highly charged histone tails that are unique to each histone. Two copies each of H2A–H2B and H3–H4 heterodimers assemble into octameric particles that wrap 145-150 bp of DNA to form a canonical nucleosome[25,99]. The histone fold domains of all eight histones are responsible for organizing ~120 bp of DNA, while the N-terminal α-helix of H3 organizes the terminal ~13 bp of DNA on either side[100]. HF extensions further define the surface of the nucleosome. Numerous protein-protein and protein-DNA interactions within the histone octamer produce a stable disc-shaped particle; however this conformation is more dynamic than originally suggested by the early crystal structures, and this dynamic behavior is essential for the regulatory function of chromatin[25,101].

The four core histone genes are among the most evolutionarily conserved sequences in the eukaryotic domain of life, suggesting that their incorporation into genomes was an early and essential event during eukaryogenesis[102–104]. There are several hypotheses to explain the enigmatic emergence of the eukaryotic nucleus. Based on the similarities in the information-processing machinery between archaea and eukaryotes, the dominant theory suggests that eukaryotes arose through the metabolic symbiosis of an archaeal host and a proteobacterium, where the gene encoding the single, tail-less histone fold protein of archaea diversified and expanded to give rise to the four core histone genes[42,45,102,105]. However, the discovery of Nucleocytoplasmic Large DNA

Viruses (NCLDV), some of which encode their own viral histone-like proteins, gave rise to the Viral Eukaryogenesis hypothesis. This hypothesis suggests that the early eukaryotic cell was a tripartite conglomerate of an archaeon, an alpha-proteobacterium, and a complex DNA virus (possibly represented by modern NCLDV)[6,106,107].

Many NCLDV or 'giant viruses' further support this hypothesis by demonstrating various intermediate dependencies on their hosts. They either replicate and assemble in a viral factory that is located in the cytoplasm, or they transiently recruit various host nuclear proteins to the viral factory in the cytoplasm for transcription of early genes[10,58]. Additional evidence supports the hypothesis that the eukaryotic nucleus may be viral in origin, as many NCLDV encode homologues of the critical m7G capping apparatus that is absent in most archaea[46]. In all eukaryotes, the nuclear membrane separates chromatin from the cytoplasm and ribosomes, facilitating the decoupling of transcription from translation with the m7G capping. These hallmark eukaryotic genes that are present in the virus represent one critical component of eukaryotic differentiation not widely seen in other domains of life[6]. Therefore, while most giant viral proteins were initially believed to have eukaryotic origin, this alternative hypothesis suggests that they may instead have contributed to modern eukaryotic features. As such, NCLDV may provide insight into the origin of the eukaryotic nucleus[44].

At least one modern histone-encoding NCLDVs (Melbournevirus, a member of the *Marseilleviridae*) organize their genome in closely packed eukaryotic-like nucleosomes, suggesting eukaryotic nucleosomes (eNuc) may have evolved from viral DNA packaging into virions[48,108,109]. *Medusavirus medusae* (MM), named for its ability to reduce the *Acanthamoeba castellanii* (*A. castellanii*) host to 'stone' through encystment, is one of the few NCLDV known to encode all four

core histones (H2A, H2B, H3 and H4) on separate genes, and also harbors a putative homologue of the linker histone H1[8,21]. Unlike the Melbournevirus, the MM genome enters the host nucleus to initiate DNA replication, while particle assembly and DNA packaging are carried out in the cytoplasm near the nuclear membrane[21]. The MM putative linker H1 is expressed with immediate early genes in the MM life cycle, suggesting a role in reshaping host transcription patterns[21,110]. In contrast, the four core MM histones are expressed at a later timepoint in the infection cycle, suggesting that they might have a role in compacting, protecting, and regulating the viral genome by forming nucleosome-like structures within the established viral factory[21,110]. As such, it is unclear where MM histones are assembled onto viral DNA. Compared to eukaryotic histones and Melbournevirus, MM histones contain notable differences, particularly in the length of histone tails and loops within the histone fold (**Figure 3.1A, B, S3.1A-C).** Additionally, the MM putative linker histone H1 has a dramatically more acidic isoelectric point (pI) than its eukaryotic counterpart, suggesting a lack of DNA binding ability (**Figure S3.1D-F**).

Here, we utilize cryogenic electron microscopy (cryo-EM) to reveal that the predicted core histones from MM form octamers that assemble with DNA into nucleosome-like particles (MM-NLPs). These NLP are characterized by unique accommodations for elongated loops and tails, and a more pronounced positively charged DNA interacting ridge compared to eukaryotic histone octamers. Additionally, we demonstrate that MM histones can form positioned tri-nucleosomes[111]. AlphaFold and AFM analysis demonstrates that the putative virally encoded linker histone H1 consists of two winged-helix domains with bimodal charge distribution, but does not promote chromatin compaction, suggesting an alternate virus-specific function. Together, the data

presented here suggest the importance of MM-NLPs in the evolutionary timeline of nucleosome architecture and advances our understanding of NCLDV chromatin organization.

3.3 Results

3.3.1 MM core histones form distinct, stable nucleosome-like particles irrespective of DNA sequence

The MM genome harbors genes for homologs of histones H2A, H2B, H3 and H4 (ORF 318, ORF 61, ORF 255 and ORF 254, respectively)[21]. Secondary structure predictions indicate that MM histones have canonical histone folds (α1–L1–α2–L2–α3), but that the α-helices are connected by longer loops (MM-H2B) or have longer tails (MM-H2B, MM-H2A, MM-H3) than eukaryotic histones (**Figure 3.1A**). Key eukaryotic nucleosome (eNuc) features are maintained through the H3 αN-helix (which helps organize DNA ends of nucleosomes), but MM-H2A diverges in sequence from eNuc H2A in the docking domain which tethers the H2A-H2B dimers to $(H3\text{-}H4)_2$ tetramers (**Figure 3.1A and S3.1**). MM histones share a 77.8% sequence identity with histones to its closest relative *Medusavirus stheno,* but less than 30% with eukaryotic histones and only 23% with the fused histones from *Marseilleviridae* or histones from MM's closest relative Clandestinovirus. This makes the MM histones unique amongst other known viral histones and more closely related to their amoeba host than other NCLDV (**Figure 3.1B and S3.1**). Many signature histone residues are conserved across viral histones including arginine side chains that extend into the DNA minor groove, the paired L1 loop (R-T pairs), and intermolecular HF stabilization (R-D clamp). However, MM histones have more predicted DNA-interacting residues in H2A and H2B relative in sequence to other known binding residues in eukaryotic histones (**Figure 3.1A and S3.1**).

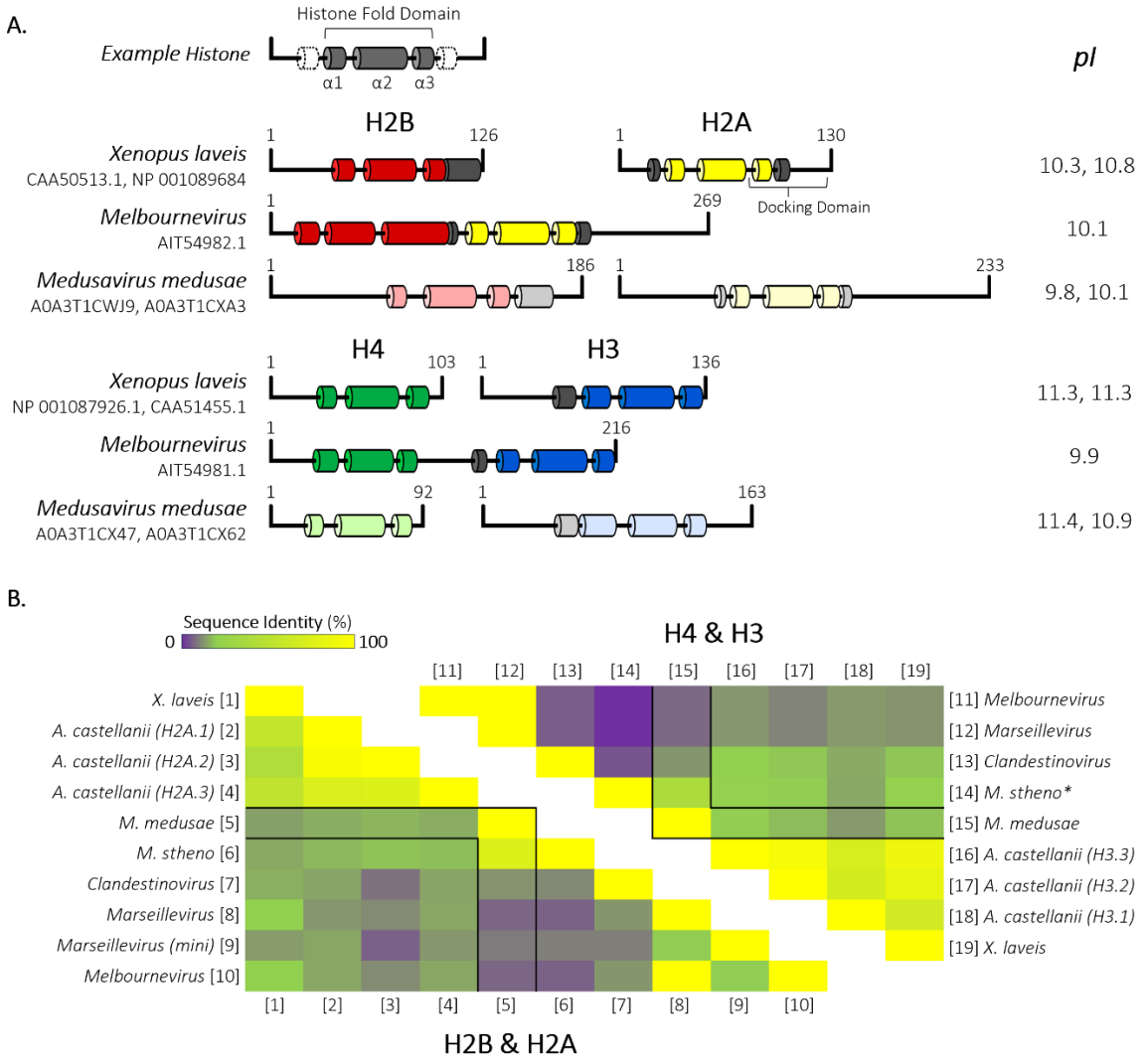


Figure 3.1 Secondary structure prediction and sequence alignment of *Medusavirus medusae* histones reveals conservation of key eukaryotic residues.

(A) Schematic of Eukarya (*Xenopus laevis*) and Nucleocytoviricota histones from Melbournevirus (MV) and Medusavirus (MM). Known *X. laevis* and MV α helices representative of the histone fold domain is represented in dark-colored tubes (H2B, red; H2A, yellow; H4, green; H3, blue; and additional helices, gray). α helices in MM histones were generated using HHpred's Quick 2D prediction web server (shown in lighter designated colors). Isoelectric points (pI) of each histone are shown to the right. (B) Heat map comparing percent identity of Eukarya and Nucleocytoviricota H2B-H2A histone sequences (left triangle) to each other. Equivalent H2B-H2A histone sequences are represented along the bottom side [1-10]. Heat map (right triangle) of percent identity of Eukarya and Nucleocytoviricota H4-H3 sequences to each other. Equivalent H4-H3 histone sequences are represented along the top side [11-19]. MM histones are outlined in black within both triangles. A comparison of different dimer pairs (H2B-H2A to H4-H3) sequence identity is not displayed. **M. stheno* H4-H3 alignment values determined from H3-H4 alignment shown in Figure S3.1C.

We expressed, purified, and refolded the four viral histone homologs into an octameric complex (**Figure 3.2A**). Utilizing the well-established salt-gradient nucleosome reconstitution

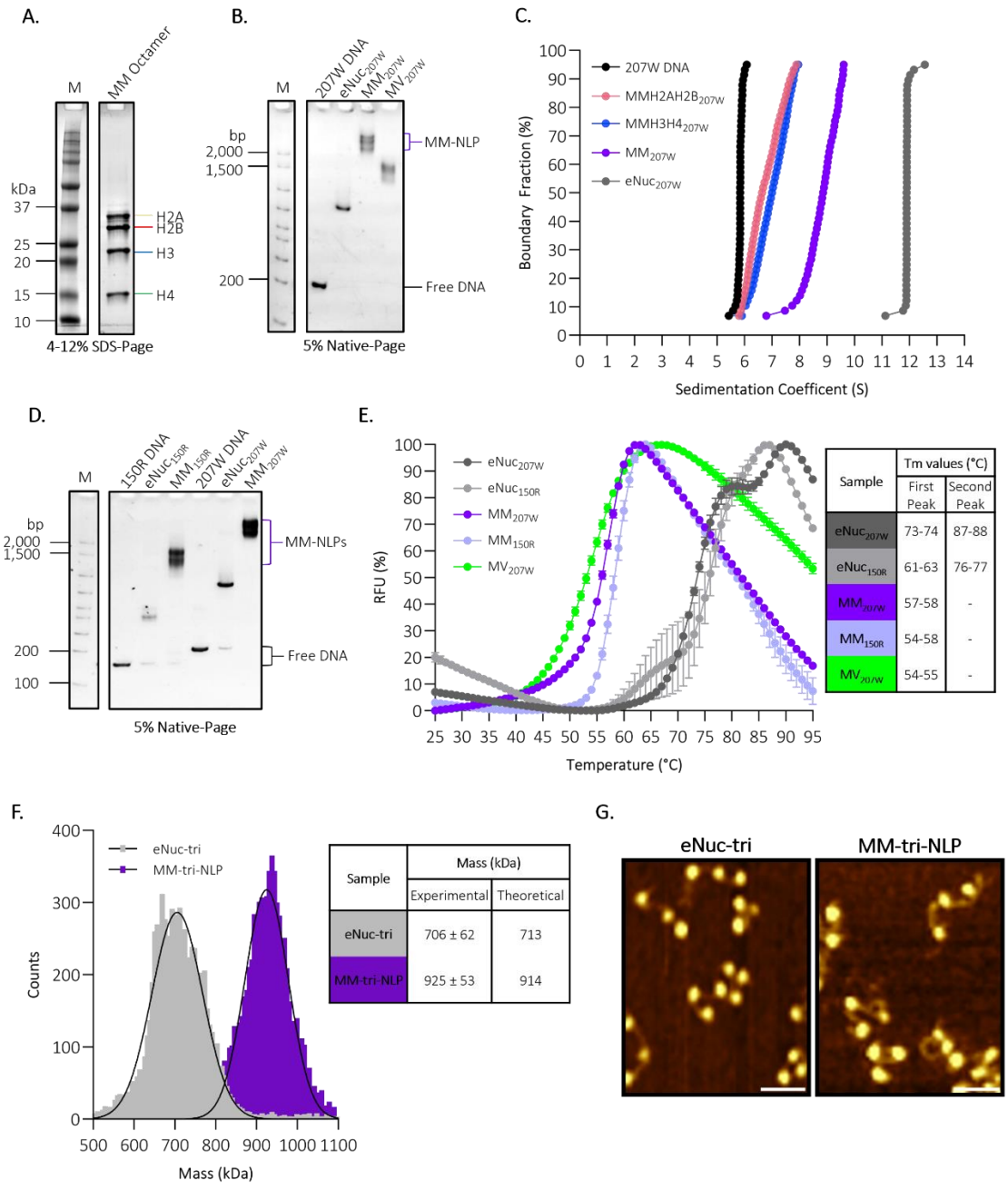


Figure 3.2 *Medusavirus medusae* histones and DNA assemble into stable mono nucleosome-like particles (NLP) and tri-NLP in vitro.

(A) SDS-PAGE of octameric refolded *Medusavirus medusae* (MM) core histones H2A, H2B, H3, and H4. (B) MM-NLP, eNuc, and Melbournevirus-NLP (MV-NLP) reconstituted with Widom '601' 207

bp DNA and analyzed by 5% Native-PAGE stained with SYBRGold (DNA visualization). (C) SV-AUC of reconstituted NLP and histone-DNA complexes. Van Holde-Weischet plot of eNuc on 207 bp DNA, histone-DNA complexes with H2A-H2B and H3-H42, and MM-NLP207. Quantitative evaluation is given in table 1. (D) 5% Native-PAGE of reconstituted MM-NLP and eNuc on 150 bp 'random sequence DNA' (50% G/C DNA (150R)), and on Widom '601' 207 bp DNA. (E) Thermal stability of MM-NLPs and eNucs shown in panel D including MV_{207W} from panel B (n=2). Averages of the replicates are shown and normalized with standard deviations denoted. Tm values of each MM-NLP and eNuc are shown in the inset. (F) Mass photometry analysis of eNuc-tri (gray) and MM-tri-NLP (purple). The solid lines represent the Gaussian function fit to the main species observed on particle counts versus molecular mass distribution histograms, with the estimated molecular weight (in kDa) corresponding to the respective mass at the center of each peak. Theoretical and measured molecular masses (± SD) are shown in the inset. (G) Representative AFM topography images of eNuc-tri and MM-tri-NLP. Scale bar = 50 nm.

protocol, increasing amounts of MM octamer were combined with 207 bp DNA (Widom "601" nucleosome positioning sequence) to obtain defined NLPs (**Figure 3.2B**)[60,112]. We confirmed that MM-NLPs on 207 bp DNA ($MM\text{-}NLP_{207W}$) contain a full complement of histones by analyzing sucrose gradient fractions by SDS-PAGE, as well as by analytical ultracentrifugation (see below) (**Figure 3.2C, Table 3.1 and S3.2**).

Table 3.1 S values (S(20,W)), frictional ratios (f/f0), and calculated molecular weights (with confidence intervals) of histone-DNA complexes from SV-AUC. All values were calculated using UltraScan[83].

Sample	S(20,W)	f/f0	Molecular Weight (kDa) experimental	Molecular Weight (kDa) theoretical	Protein Complex:DNA
207W	5.74 (5.70, 5.79)	3.59 (3.46, 3.77)	136.84 (122.39, 148.21)	127.82	-
$MM\text{-}H2AH2B_{207W}$	5.81 (5.77, 5.85)	2.95 (2.86, 3.06)	195.25 (186.43, 204.06)	180.22	2 : 1
$MMH3H4_{207W}$	5.95 (5.94, 5.96)	2.93 (2.91, 2.95)	187.89 (185.12, 188.14)	199.19	2 : 1
$MM\text{-}NLP_{207W}$	8.21 (6.97, 9.45)	2.51 (1.85, 3.18)	290.92 (237.99, 342.84)	304.00	1 : 1
$eNuc_{207W}$	12.10 (12.02, 12.15)	1.77 (1.69, 1.86)	253.57 (249.09, 266.22)	237.28	1 : 1

MM-NLPs migrate much higher on native gels compared to eukaryotic nucleosomes (eNuc), and even higher than Melbournevirus NLPs. Sedimentation velocity analytical ultracentrifugation (SV-AUC) allows the determination of macromolecule size and overall shape in solution from diffusion-corrected sedimentation values, which are proportional to particle mass and inversely proportional to viscous drag[83]. $eNuc_{207W}$ sediments at 12 S, comparable to previous $eNuc_{147W}$ sedimentation at 11 S[108]. $MM\text{-}NLP_{207w}$ sediments at ~ 8 S, caused by an increased viscous drag compared to eNuc, and similar to what was observed for Melbournevirus NLP_{207w} (**Table 3.1**). Reconstituting either MM H2A-H2B dimer or MM $(H3\text{-}H4)_2$ tetramer onto the same DNA gives rise to particles with the expected molecular weight and an even higher level of viscous drag (f/f_0). Importantly, the experimentally determined molecular mass of $MM\text{-}NLP_{207w}$ (290.92 ± 51.92 kDa) is in close agreement with the expected theoretical mass of an octameric protein complex with 207 bp DNA (304.00 kDa) (**Figure 3.2C and Table 3.1**).

To determine if MM-NLP formation is affected by either DNA sequence or length, we reconstituted MM octamer with a 150 bp 'random' DNA sequence ($MM\text{-}NLP_{150R}$). This DNA was designed to have a 50 % GC content over each 10 bp segment and does not harbor strong nucleosome positioning signals. Both the $MM\text{-}NLP_{150R}$ and $MM\text{-}NLP_{207W}$ migrate much higher than eNuc on native-PAGE, supporting the finding that MM histones form less compact particles irrespective of DNA sequence (**Figure 2D**). The stability of the two different MM-NLPs and eNucs on different sequences and their corresponding histone octamers, along with Melbournevirus NLP_{207W}, were tested in a thermal melting assay (25 ºC – 95 ºC). The MM octamer melts to fluoresce immediately, indicating instability, as opposed to the eukaryotic octamer that does not begin to dissociate until ~ 40 °C (**Figure S3A**). eNuc on both DNA fragments (207W and 150R)

demonstrate the characteristic peaks of histone dimer and tetramer release from the nucleosome, as previously reported[113]. In contrast, all viral NLPs (MM-NLP$_{150R}$, MM-NLP$_{207W}$, and Melbournevirus NLP$_{207W}$) melt in a single peak at much lower temperatures, suggesting the dissociation of the octamer from DNA in a single step, and irrespective of sequence and length. This indicates lower thermal stability of viral octamers and nucleosomes compared to their eukaryotic counterparts, as previously shown for Melbournevirus NLP (**Figure 3.2E**)[108].

To test if MM-NLP can assemble into higher-order chromatin, we generated a minimal chromatin construct using the MM octamer. We were able to successfully reconstitute MM tri-nucleosome-like particles (MM-tri-NLP), on a 3x copy of Widom 207 DNA[114]. As observed for the MM-NLP, MM-tri-NLPs migrate higher on a native gel relative to *X. laveis* tri-nucleosomes (eNuc-tri) (**Figure S3.3B**). The mass of the MM-tri-NLP was estimated in solution via mass photometry to be 925 kDa ± 53 kDa, highlighting the considerable size difference compared to eNuc-tri at 706 ± 62 kDa, (**Figure 3.2F**). MM-tri-NLPs have the same characteristic "beads on a string" structure as eukaryotic tri-nucleosomes, forming three distinct NLPs as visualized by Atomic Force Microscopy (AFM) and negative stain electron microscopy (**Figure 3.2G, Figure S3.3C**).

3.3.2 MM-NLP resemble eNuc with unique accommodations for longer histone tails and loops

Single-particle cryogenic electron microscopy (cryo-EM) was utilized to determine the structure of MM-NLP$_{207W}$. Samples were crosslinked through gradient fixation (GraFix) with glutaraldehyde[108], and compared to a native sample that had not been subjected to crosslinking. The migration in the sucrose gradient was unchanged upon crosslinking (**Figure S3.2**). Data for both particles was collected on a Titan Krios G3i, and following 3D reconstruction and refinement, we

obtained an electron density map of crosslinked MM-NLP$_{207W}$ at 4.3 Å resolution (**Figure 3.3 and S3.4A**). We observed electron density for all core histone helices and ~ 135 bp of DNA, allowing assignment of the MM-NLP$_{207W}$ core. Density for the MM-H2A tail region of the docking domain was not visible (**Figure S3.4**). Our initial homology model of MM histones in a nucleosome-like arrangement was refined before the final model was docked into the MM-NLP$_{207W}$ density, with strong agreement (cross correlation=0.705) (**Figure 3.3A**).

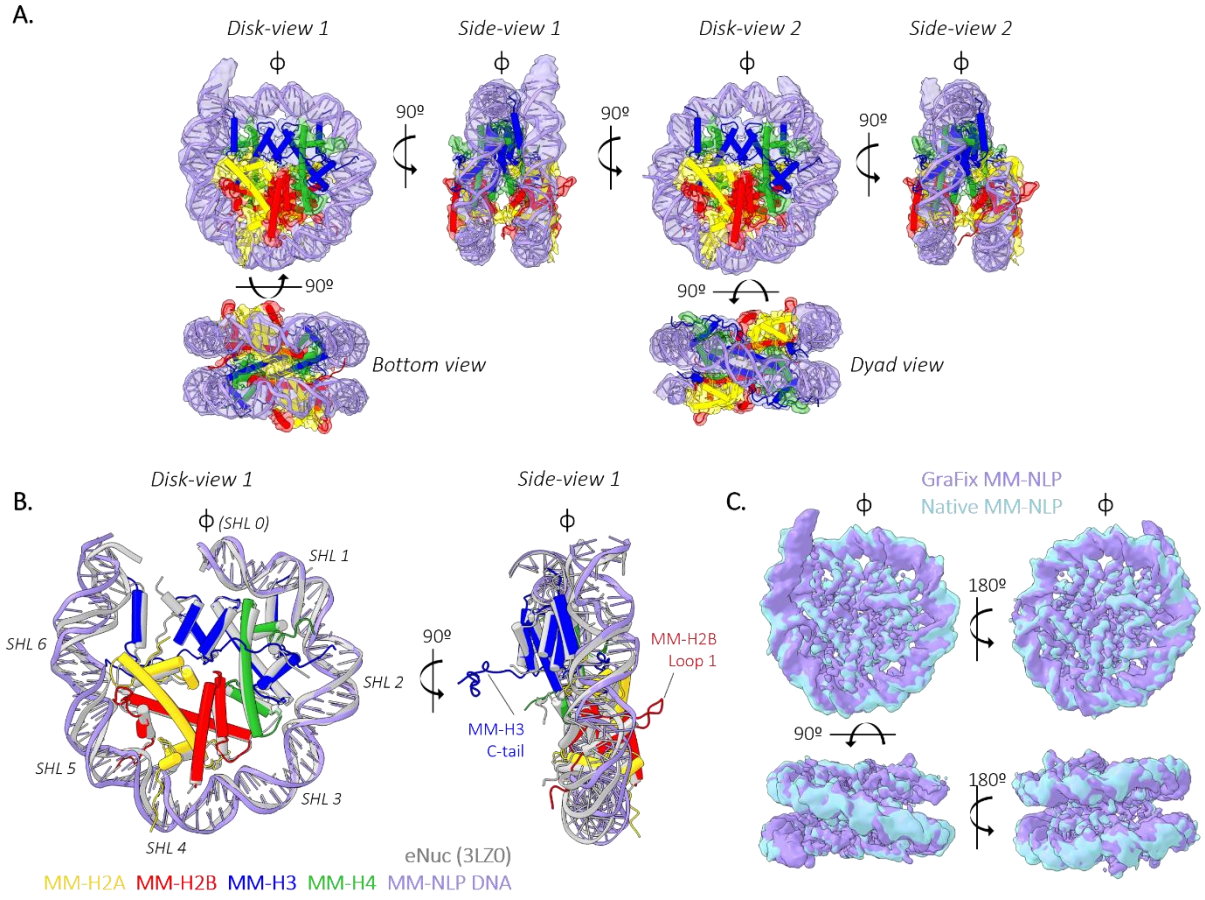


Figure 3.3 MM-NLP$_{207\ bp}$ closely resemble eukaryotic nucleosomes.

(A) Overview of MM-NLP$_{207W}$ structure. Individual histones H2A, H2B, H3 and H4 and their surrounding density are represented by yellow, red, blue, and green respectively. (B) Overlay of MM-NLP207 (H2A-yellow, H2B-red, H3-blue, H4-green, DNA-purple) with eNuc (gray) with only 71 bp of DNA, one H2A-H2B dimer, H3-H3' four helix bundle, H3' N-terminal helix, and a single H4 displayed for clarity. Superhelix locations (SHLs) are numbers from 0 to 6 starting at the

nucleosome dyad (ϕ). (C) Comparison of native (light blue) and GraFix (purple) MM-NLP207 electron densities.

Just like its eukaryotic counterpart, the MM-NLP_{207W} contains two copies each of H2A, H2B, H3 and H4 as an octameric core formed by HF regions, wrapped by ~ 130 bp of DNA (**Figure 3.3A**). However, the density of only one of the two H3 αN-helices and associated DNA is visualized at more stringent contour levels, underscoring the dynamic character of the ~13 penultimate base pairs of DNA which has been observed in several eukaryotic and viral nucleosome structures[115,116]. The overall geometry of the DNA superhelix and the layout of HF helices are near-identical between MM-NLPs and the eukaryotic nucleosome, with minor differences in the wrapping of DNA ends (**Figure 3B**). We also determined the structure of native (i.e. not crosslinked) Melbournevirus NLP_{207W} at 5.1 Å. This structure had a high correlation (0.949) with the crosslinked Melbournevirus $NLP_{207W,}$ confirming that there were no induced and potentially artificial conformations due to crosslinking (**Figure 3.3C and 3.4A)**.

MM histones have distinctly longer tails (H2A, H2B, H3) and loops (H2B) than eukaryotic histones. Most unusually, MM-H3 has a 29 amino acid long C-terminal tail, unlike any other H3 histone **(Figure 3.3B)**. Both of our structures reveal that this H3 tail extends from the end of α3 and lays across its partnered H4 α2-helix to reach the DNA minor groove at SHL ±1.5 (**Figure 3.4A and 3.4B**). In this position, it is in close contact with the H2A-H2B dimer, H4 α2, the H4 N-terminal tail, and H3 α1 through a combination of hydrophobic packing and hydrogen bonding interactions. The density for this tail is very pronounced, and observed in both the crosslinked and native structure, excluding crosslinking artifacts. The presence of the extended H3 C-terminal tail redirects the MM-H4 N-terminal tail to contact the DNA at SHL ± 1.5 (**Figure 3.4B**). The penultimate 10 amino acids

of the H3 C-terminal tail (containing three arginines and two lysines) are poised to interact with SHL ± 5.5 but are too disordered to be observed in the maps.

A second unique feature of MM histones is the extended H2B L1 loop connecting α1 and α2, which is 11 amino acids longer compared to eukaryotic histones. This loop, which in many respects resembles a β hairpin, has very well-defined density that protrudes out from the nucleosome disk by about ~ 15 Å (**Figure 3.3B**)[117]. Its base packs against H2B α2-L2 by forming a hydrophobic core centered around H2B F91 (**Figure 3.4C**). Together with α1, this extended loop forms a defined module that contributes to stabilizing the interaction with H2A, and uniquely defines the surface of the MM-NLP. Additionally, the α3 helix of MM-H2B is longer which allows the extended connector (4 amino acids) between H2B α3 and αC to contact the H2A docking domain and the extended H3 C-terminal tail, contributing to their unique interactions within the histone octamer (**Figure 3.4C**).

MM-NLPs exhibit a more pronounced positively charged electrostatic DNA-interacting ridge compared to eNuc. Full views of the ridge show a higher density of positive charge in MM-NLP, consistent with the predicted increase in DNA-interacting residues (**Figure 3.4D, Figure 3.1A, Figure S3.1**). The acidic patch, a localized region on each disk "face" of the nucleosome, is a well-established binding site for many chromatin-interacting proteins in eukaryotes[118]. The negative charge in this region of the MM-NLP is less pronounced (**Figure 3.4D**). Notably, the H3 C-terminal tail contributes to forming a pronounced "S" shaped acidic surface along the bottom side of the $(H3\text{-}H4)_2$ tetramer that is unique to MM histones (**Figure 3.4E**).

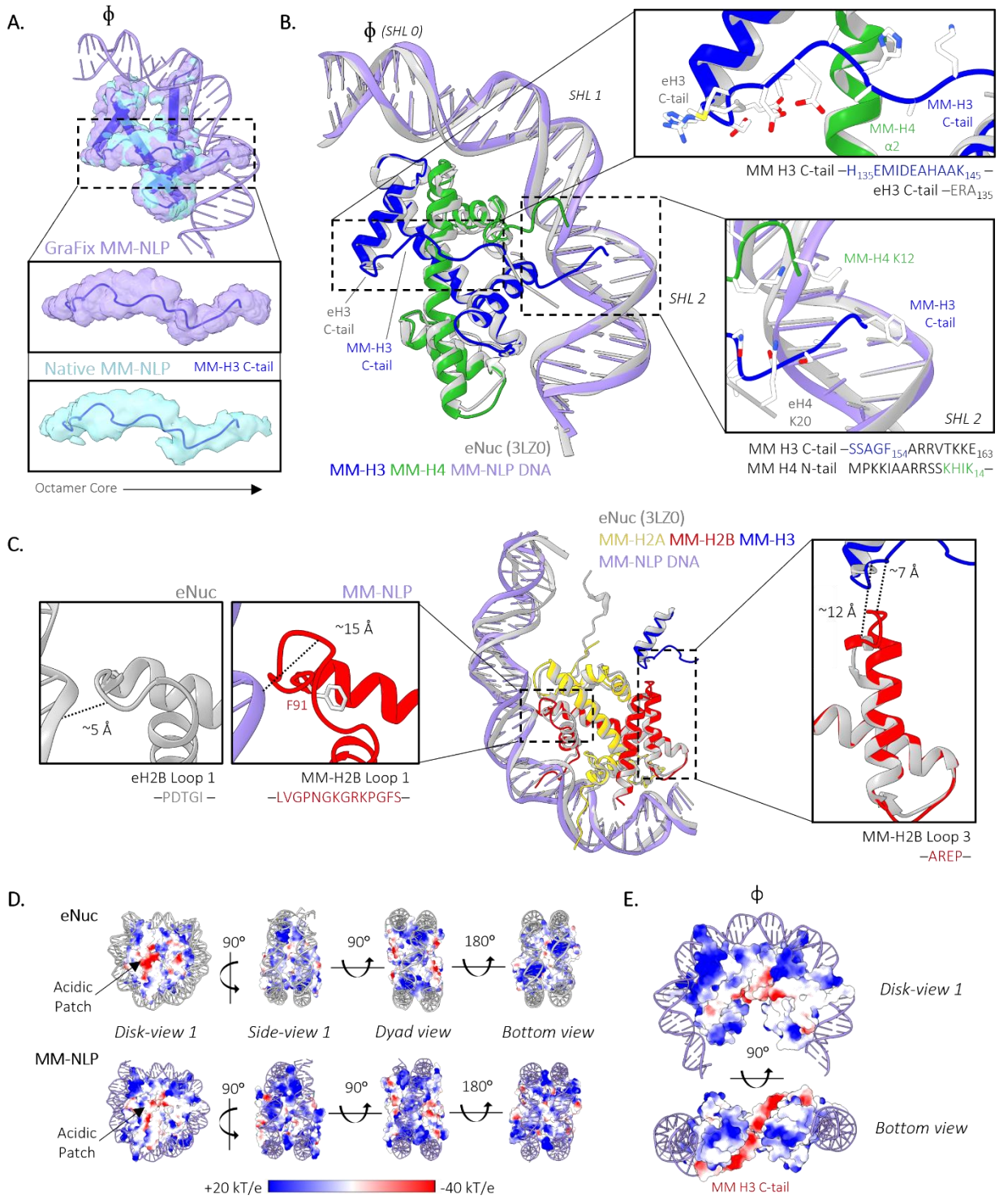


Figure 3.4 Comparison to eNuc reveals unique roles and accommodations for longer tails and loops.

(A) Comparison of crosslinked and native MM-NLP. Superposition of GraFix MM-H3 density and native MM-H3 density. The MM-H3 tail from H135 to F154 are displayed in the inset of each

density. (B) Superposition of MM-(H3-H4)2 tetramer with eukaryotic (H3-H4)2 tetramer and 30 bp of associated DNA. Close-ups are provided of MM H3 C-tail's orientation in MM-NLP in relation to the eukaryotic MM H3 C-tail and the MM-H3 C-tail extension to SHL 1.5. Corresponding residues of each tail are denoted below each close-up, with tail end residues that are not displayed in black. (C) Superposition of MM-H2A-H2B dimer aligned with eukaryotic H2A-H2B dimer (light gray) and 40 bp of associated DNA. Close-ups are provided of both eH2B and MM-H2B loop 1 along with the MM-H2B α3 and loop 3 extension towards the MM-H3 C-tail. Distances provided were measured from the backbone atoms of the closest residues from each loop to DNA or H3 C-tails. (D) Charged surface representation of histones from the MM-NLP_{207} and eNuc (PDB ID: 3LZ0). (E) Charged surface representation of MM tetramer with only 60 bp of DNA with associated histones.

3.3.3 Medusavirus NLPs are structurally similar to eukaryotic nucleosomes than to other viral NLPs

To date, the only other viral NLP for which structural information is available is the Melbournevirus nucleosome, where histones are fused into doublets (H2B-H2A and H4-H3)[48,108]. Overall, the two structures are rather similar to each other in the positioning of the histone fold elements and the DNA. An intriguing commonality between the viral NLPs is the presence of a histone tail that lays across the same region of the H4 α2-helix (**Figure 3.5A**). In MM-NLP, as shown above, the extended H3 C-terminal tail (unique to MM-H3) reaches over H4 α2, while the redirected N-terminal tail of H4 assumes the same function in Melbournevirus NLP but coming from the opposite direction. The tails contact similar positions on top of the H4 α2-helix (centered around H4 V50) (**Figure 3.5B, Figure 3.1B**). The interface consists of charge-charge interactions and hydrophobic packing between the H4 α2-helix and the respective tail, signifying an intriguing functional convergence between the two different viral particles (**Figure 3.5B**).

The charge distribution of the histone octamer surface differs between the two viral NLPs, with MM-NLP displaying a significantly more positively charged DNA interacting ridge compared to Melbournevirus NLP **(Figure S3.5)**. Superimposition of MM-NLP and Melbournevirus NLP with

eNuc allows us to highlight conserved histone-histone and histone-DNA interactions. Key features such as the H3 'R-D clamp' and the 'sprocket arginine' from H2A are present in MM-histones, and as such are conserved between viral and eukaryotic NLPs (Figure 3.5C). However, MM-NLP maintains the arginine minor groove interactions from H3 and H4 while Melbournevirus NLP does not have the equivalent residues, likely further contributing to wrapping DNA.

The histone core is held together by four-helix bundle interactions between H3 and H3', and between H2B and H4. MM-NLP differs in both of these interfaces from Melbournevirus NLP (Figure 3.5C and 3.5D) and eNuc. In MM-NLP, both interfaces are characterized by glutamine residues, and are less hydrophobic in nature than either Melbournevirus NLP or eNuc. Both viral nucleosomes lack the histidine-cysteine configuration in the H3-H3' four-helix bundle that is typical to nearly all eukaryotic nucleosomes, and that is hypothesized to convey copper reductase activity[119].

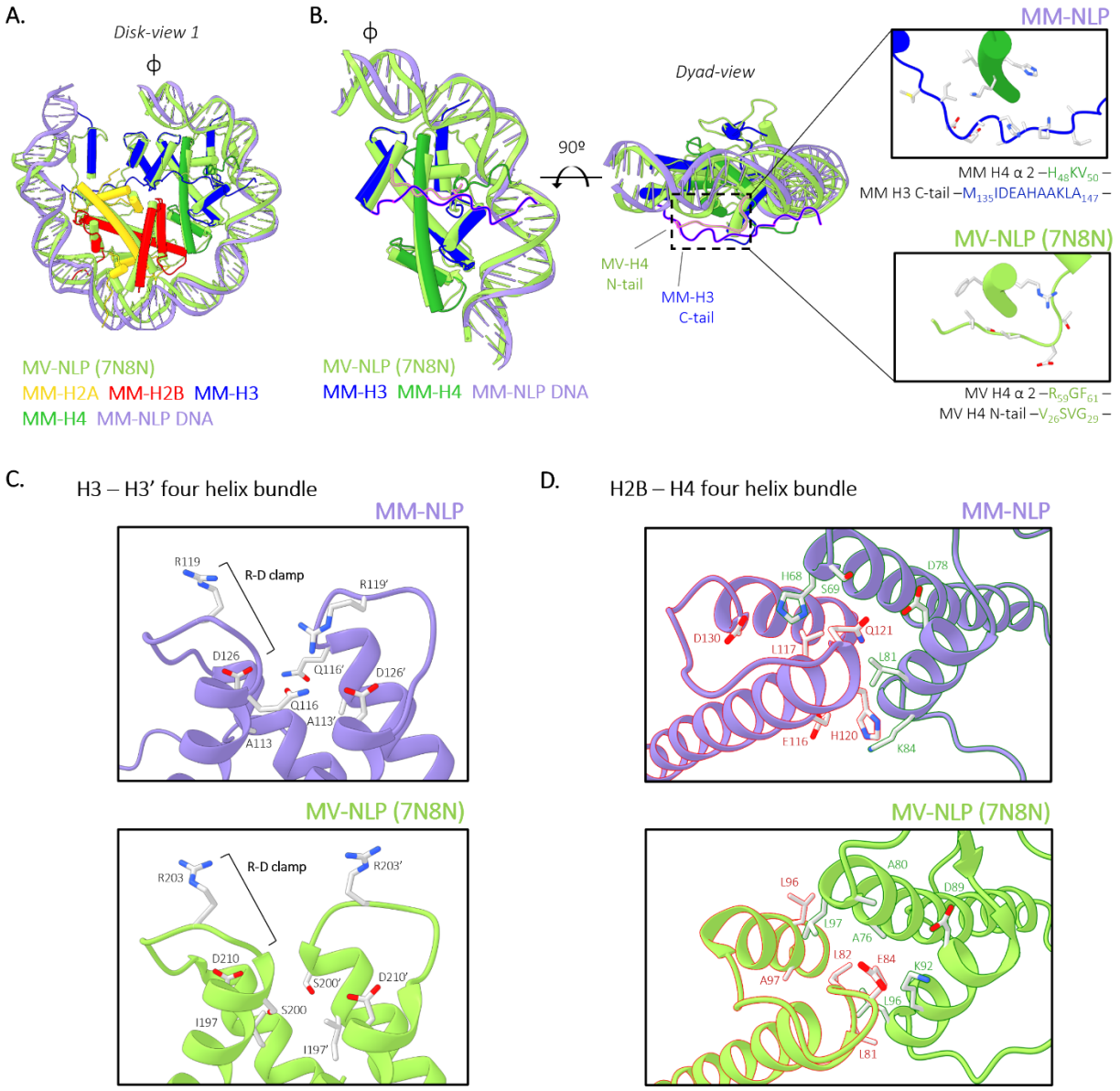


Figure 3.5 Structural comparison of viral nucleosome structures exposes distinct accommodations and variations in intermolecular interactions of viral NLP.

(A) Superposition of MM-NLP and Melbournevirus-NLP (MV-NLP; light green; PDB ID: 7N8N). (B) Superposition of H3-H4 histone dimer of MM-MVP and H4-H3 histone doublet of MV-NLP (light green). MM-H3 C-tail and MV-H4 N-tail are outlined in pink. Inset highlights the MM-H3 C-tail and MM-H4 N-tail pathway across the H4 α2 of each viral NLP. Residues within the inset are denoted below each close-up. (C) Residues contributing to the H3 – H3' four helix bundle interactions of MM-NLP and MV-NLP. The notable R-D clamp seen in eukaryotes is highlighted in both insets. (D) Residues contributing to the H2B – H4 four helix bundle interactions of MM-NLP and MV-NLP.

3.3.4 MM putative linker histone H1 does not compact MM tri-nucleosomes

Unique amongst histone encoding NCLDVs, MM encodes a putative linker histone H1 (ORF 106) with an uncharacteristically acidic pI of 5, sharing this unique feature with H1 of its host *A. castellanii* (**Figure 3.6A**)[21]. The putative MM linker histone H1 (MM-H1) has a low sequence conservation (12.86 %) compared to eukaryotic *X. laveis* H1.0 and *Gallus gallus* H5. Secondary structure predictions of MM-H1 suggest the presence of the canonical N-terminal winged-helix DNA-binding domain seen in *X. laveis* H1.0 and *A. castellanii* H1.1 (**Figure S3.1D**). Unexpectedly, we predict an additional winged-helix domain in both the MM-H1 and the host *A. castellanii* H1.1 following a ~15 amino acid loop, which is supported by each respective AlphaFold structure (**Figure 3.6A and S3.6**). Both winged-helix domains predicted for MM-H1 have a more negative charge

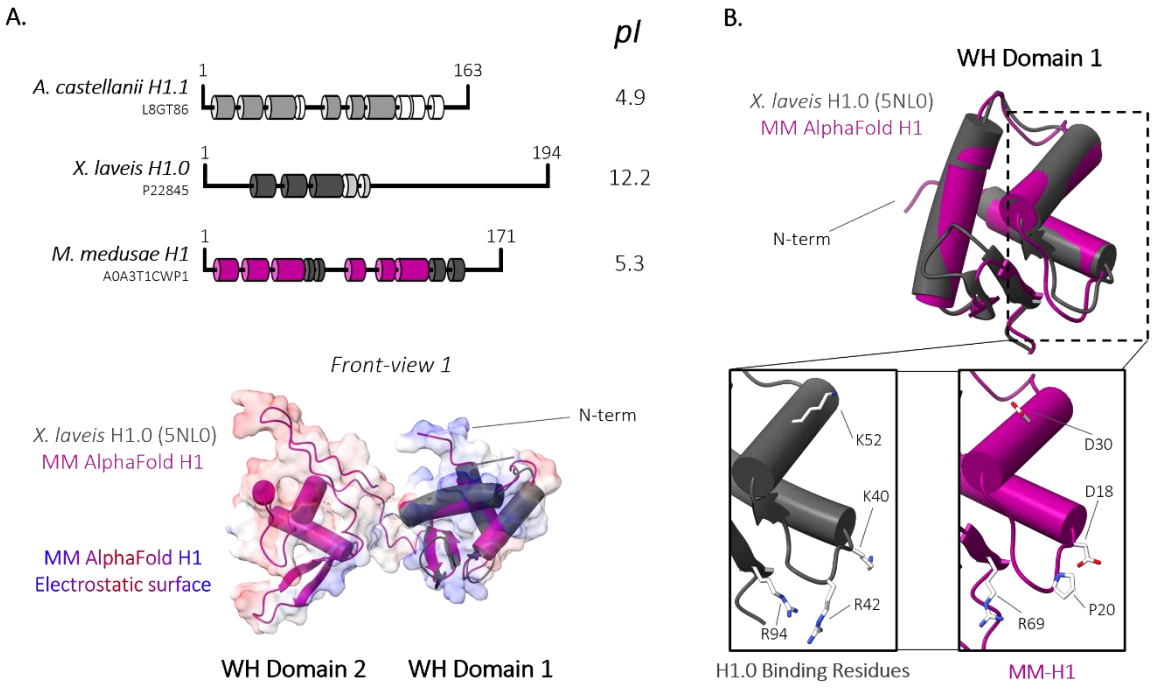


Figure 3.6 MM tri-nucleosomes are not compacted by the putative linker histone H1.

(A) *Acanthamoeba castellanii* (*A. castellanii*) H1.1, *Xenopus laveis* (*X. laveis*) H1.0, and *Medusavirus medusae* (MM) linker histones were aligned using HHpred's multiple sequence alignment tool

(ClustalΩ). Predicted *A. castellanii* α helices and beta sheets in *Mamonoviridae* histones were generated using HHpred's Quick 2D prediction web server (shown in dark pink). Known *X. laveis* H1.0 and *A. castellanii* α helices and beta sheets representative of wing helix (WH) domain is represented in dark-grey tubes and light-grey tubes, respectively. Predicted α helices and beta sheets in *Mamonoviridae* histones were generated using HHpred's Quick 2D prediction web server (shown in dark pink). Corresponding isoelectric points (pI) of each linker histone are provided to the right. Superposition of AlphaFold MM-H1 (dark pink) and *X. laveis* H1.0 (dark grey; 5NL0) within the charged surface representation of MM-H1. (B) Overlay of MM-H1 (shown in dark pink) and *X. laveis* H1 (dark grey; 5NL0). Inset: *X. laveis* H1 residues involved in chromatin compaction and corresponding residues in MM-H1.

compared to *X. laveis* H1.0, with the second predicted domain displaying a predominantly negative charge (Figure S3.6). When we superimpose the MM-H1 with *X. laveis* H1.0, it reveals that the four basic amino acids (K40, R42, K52 and R94) required to compact chromatin are not conserved in MM-H1[120]. Instead, the predicted MM-H1 has one equivalent basic residue (R69) and acidic residues (D18, P20 and D30), which suggests MM-H1 may lack the ability to function in compaction (Figure 3.6B).

To characterize the putative MM linker histone H1 (MM-H1), we expressed, purified, and refolded the protein for biochemical analysis (Figure S3.7A). Circular dichroism of refolded MM-H1 confirms a significant decrease in disorder when compared to *Mus musculus* H1.0 (eH1.0), with disorder representing approximately 25% and 64%, respectively. The observed decrease in disorder is accompanied by an observed increase in α-helix and β-sheets for MM-H1, as expected based on secondary structure predictions where MM-H1 contains a second winged-helix domain which is canonically disordered in eH1.0 (Figure S3.6)[121,122]. AFM has previously been used to demonstrate eNuc-tri compaction by eH1.0, which was observed by an increase in particle height, as well as an increase of visually compacted particles[34]. Here, we utilized AFM to probe the ability of the MM-H1 to compact tri-nucleosomes. As shown previously, eNuc-tri compaction was

observed with the addition of eH1.0 via an increase in the height distribution of tri-nucleosome particles (**Figure 3.7A, 3.7B and 3.7E**). This is expected based on structures of H1 bound nucleosome arrays, which show a zig-zag arrangement with non-consecutive nucleosomes 1 and 3 forming a stack and H1 bound to linker DNA near the dyad of each nucleosome[123]. This stacking likely explains the "bi-lobe" tri-nucleosomes we observe via AFM when tri-nucleosomes are incubated with eH1.0, which the stacked nucleosomes at the DNA ends (nucleosome 1 and 3) are higher than the single nucleosome (nucleosome 2) (**Figure 3.7A, 3.7B, and 3.7E**). However, the addition of the MM-H1 to eNuc-tri did not yield these distinct "bi-lobe" tri-nucleosomes or an increase in particle height (**Figure 3.7C, 3.7D, and 3.7E**). Additionally, compaction was directly assessed by measuring the total distance between each nucleosome in the tri-nucleosomes. Upon the addition of eH1.0 to eNuc-tri there was a significant decrease in total distance between nucleosomes, as they were compacted, that was not seen upon the addition of MM-H1 (**Figure 3.7F**). Using these two AFM analyses, no tri-nucleosome compaction was observed upon the addition of MM-H1. Together this data suggests that MM-H1 does not function in compaction under these conditions, as predicted by the lack of key compaction residues and acidic electrostatic surface (**Figure 3.6B and S3.6**).

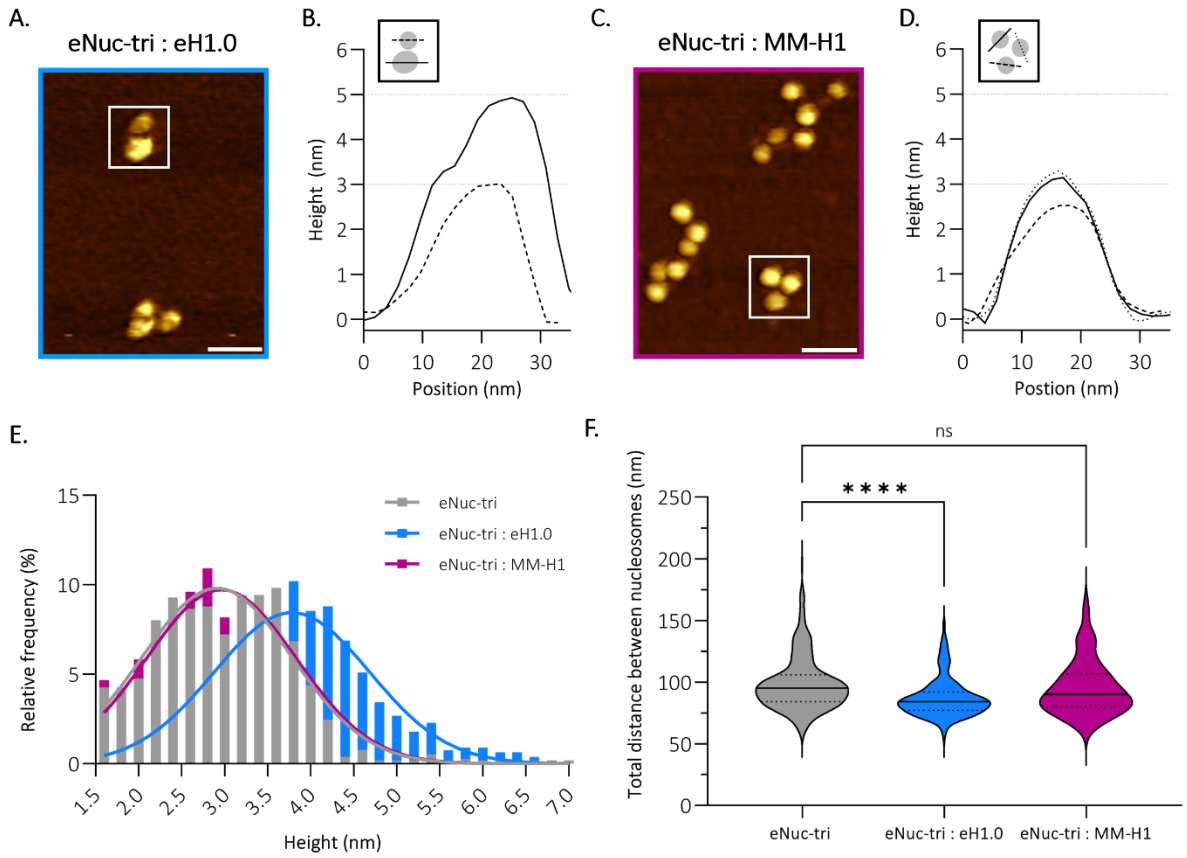


Figure 3.7 *Medusavirus medusae* linker histone H1 does not compact tri-nucleosomes.

(A) Representative AFM topography image of eNuc-tri:eH1.0 sample, imaged in air, with (B) height profile through indicated particles shown in graphical inset. (C) Representative AFM topography image of eNuc-tri:MM-H1 sample, imaged in air, with (D) height profile through indicated particles shown in graphical inset. Dashed lines at 3 and 5 nm shown for reference. Scale bar = 50 nm. (E) Histogram and Gaussian fitting of particle height for eNuc-tri alone (2.9 ± 0.9 nm, *N* = 774, grey) and after incubation with either eH1.0 (3.9 ± 0.9 nm, *N* = 784, blue) or MM-H1 (3.0 ± 0.9, *N* = 770, pink). (F) AFM compaction analysis for eNuc-tri alone (99 ± 22 nm, *N* = 260, grey) and after incubation with either eH1.0 (87 ± 16 nm, *N* = 245, blue) or MM-H1 (95 ± 22 nm, *N* = 249, pink). These values represent mean±s.d. and statistical significance was determined by unpaired two-tailed Student's *t*-test, **** represents P <0.0001, ns represents P >0.05.

3.4 Discussion and Conclusions

Once believed to be unique to the eukaryotic domain of life, the universe of histone-encoding organisms continues to expand to now include most archaea, some bacteria, and most

recently, several members of giant viruses (NCLDV). Structural analysis of non-eukaryotic histone-DNA complexes has demonstrated a remarkable diversity in histone-based DNA organization. For example, bacterial histone dimers bind DNA edge-on and filament around straight DNA, while archaeal histone homodimers wrap variable lengths of DNA into dynamic 'hypernucleosome slinkies'[53,54,124]. In contrast, giant viruses encode clearly recognizable homologs of the four eukaryotic core histones (H2A, H2B, H3, and H4) that assemble into nucleosome-like particles that wrap ~130 bp of DNA[108,48]. Phylogenetic analysis suggests that these histones diverged prior to the emergence of LECA[17,21,23,41,50]. Instances of histone doublets (or even triplets and quadruplets) have been described in NCLDV, which enforce specific pairing of histones and suggest a potential role in the origin of eukaryotic histone dimers/tetramers[41,42,99]. With some NCLDV only encoding one doublet pair, it is possible that NCLDV histones doublets have evolved independent functions in viral genome packaging; although it is unclear if these independent functions were developed before or after complete sets of core histone singlets were encoded in eukaryotes or even MM[23]. MM histones are placed at the root of most histone phylogenetic trees, and thus add to new potential theories regarding nucleosome evolution[6,21,44,102]. Their resemblance to eukaryotic core histones is plausibly a result of acquiring genes through horizontal gene transfer (HGT) post-LECA, however their sequence similarly is low compared to its host *A. castellanii* and even other viral histones. As such, the role of MM core histones in the origin of the nucleosome remains a mystery.

We show that key eukaryotic nucleosome features are conserved within MM-NLPs, despite the low conservation in amino acid sequence. This includes the overall geometry and arrangement of the DNA superhelix, which is achieved through the interactions between the DNA minor groove backbone and the main chain of antiparallel L1-L2 loops of the histones, as well as the N-termini

of histone fold α1 helices[25,48,108]. However, MM histones diverge in the lengths of tails and loops within the HF which distinctly shape the nucleosomal surface. For example, the MM-H2B loop 1 may act as an 'arm' for other interacting viral or host proteins, including putative nucleosome assembly, remodeling factors and histone chaperones, and/or may change the way in which nucleosomes pack against each other into higher order structures. A similar role for an extended loop has been observed for the centromeric histone variant CenpA[71].

The unique extended MM-H3 C-terminal tail lays across the H3-H4 histone fold. The path of the MM-H3 C-term tail, confirmed by very well-defined density in both the native and crosslinked complexes, encourages interactions with the elongated loops in MM-H2B and promotes packing against the MM-H3 α2 to provide a unique stabilization of the MM-NLP. While the H3 C-terminal tail is unique to MM, its pathway across H3 α2 mimics that of the Melbournevirus H4 N-term tail, which extends out and lays across the histone fold dimer in the exact same location but from the opposite direction. This conserved requirement of a 'stabilizing tail' (either from H3 in MM, or from H4 in Melbournevirus) suggests an ancient feature of the nucleosome that was later made obsolete in eukaryotic nucleosomes upon the utilization of histone tails for signaling and regulation.

Unlike most histone encoding NCLDV, Medusavirus encodes an acidic putative linker histone H1 that maintains a significant sequence identity to H1 encoded by the host *A. castellanii*. This is a much higher degree of identity than what is observed for the core MM histones (~30%) and suggests that the predicted H1 was shared through horizontal gene transfer at a later stage. The MM-H1 has the canonical winged-helix domain seen in *X. laveis* H1.0 but only one of the four established residues required for chromatin compaction. The MM-H1 also demonstrates an

additional winged-helix domain, as seen in the three linker histone genes of the host *A. castellanii*, which is extremely acidic and likely dissuades interactions with DNA. This is supported by our AFM analysis, in which no compaction is observed when MM-H1 is added to the tri-nucleosomes, as seen for eH1.0. This suggests an alternative and independent viral function of the MM-H1 beyond the anticipated compaction of chromatin.

During viral assembly of Melbournevirus, viral histones reside exclusively in the viral factories of the host cytoplasm to mature into virions upon Melbournevirus infection[108]. In contrast, MM maturation relies heavily on the host nucleus. Viral particles are independently produced in the cytoplasm, but viral DNA replication occurs in the nucleus, and only the assembled capsids that are close to the nuclear membrane are filled with DNA. This could allow the opportunity for the exchange of protein and DNA through open membrane gaps of the MM viral particle positioned near the nuclear membrane, to which MM histones may be used for chromatin organization between or within the host nucleus and the viral particle. Whether the DNA is already assembled into chromatin at the time of packaging, or whether nucleosome assembly takes place in the capsid remains to be determined[125]. The pronounced differences in the four-helix bundle interfaces holding together the $(H3\text{-}H4)_2$ and tethering the $(H3\text{-}H4)_2$ to the H2A-H2B dimer, might have evolved to prevent the formation of hybrid nucleosomes consisting of host and virally encoded histones, although this requires further investigation. While histones are essential for Melbournevirus fitness and infectivity, the importance of MM histones for MM fitness has yet to be explored[21,108]. As more metagenomes are discovered, more organisms will reveal encoded histones to fill in the gaps of the evasive evolution of the nucleosome.

3.5 Materials and Methods

3.5.1 Histone sequence alignment and secondary structure prediction

Predicted *Mamonoviridae* family and *Marseilleviridae* family core histone-like protein sequences were aligned with eukaryotic histone sequences with HHpred's Multiple Alignment using Fast Fourier Transform (MAFFT) with a 1.52 gap open penalty[8,21]. The sequence similarly and identity of *Medusavirus medusae* (MM), *Medusavirus stheno*, *Clandestinovirus, Marseillevirus* and Melbournevirus (MV) relative to each core eukaryotic histone were obtained using the Sequence Manipulation Suite webserver (SMS)[126]. To demonstrate structural conservation of the canonical core histone fold domain between *Eukarya* and *Mamonoviridae*, protein secondary structures were predicted using the MAFFT alignment on the HHpred's Quick 2D structural prediction webserver[74]. The isoelectric point of each protein was provided by protparam.

Predicted *Mamonoviridae* family and *Clandestinovirus* linker histone-like protein sequences were aligned against host *Acanthamoeba castellanii* H1*, Xenopus laveis* H1 and *Gallus gallus* H5) with HHpred's Multiple Alignment using Fast Fourier Transform (MAFFT) with a 1.52 gap open penalty[8,21]. The sequence similarly and identity of MM, *Medusavirus stheno* and *Clandestinovirus,* relative to each linker eukaryotic histone were obtained using the Sequence Manipulation Suite webserver (SMS)[126]. To demonstrate structural conservation of the canonical winged-helix domain between *Eukarya* and *Mamonoviridae*, protein secondary structures were predicted using the MAFFT alignment on the HHpred's Quick 2D webserver[74].

3.5.2 MM histone expression, purification, and refolding

MM ORF 318 (H2A), ORF 61 (H2B), ORF 255 (H3), ORF 254 (H4) and ORF 106 (H1) were each cloned into a pET-28a plasmid for expression and purification from *E.coli*., utilizing adaptations of well-established eukaryotic histone protocols[60,127]. Expression of each histone was performed in 6 L of Rosetta 2(DE3) *E.coli* cells with induction of 0.5 mM IPTG at OD:0.4-0.6 and growth at 14°C for 18 hrs. Inclusion bodies were isolated from the cells utilizing a minor adaptation from published protocols, including a 30 min stir of the cell suspension in wash buffer containing 50 µL of DNase I, 5 mM $MgCL_2$, 50 µg of Lysozyme and 10 mM $CaCl_2$ before tissumizing[60]. Isolated inclusion bodies were treated with 1 mL of 99.7% DMSO for 30 min before stirring with 40 mL of the denaturing lysis buffer (6 M Guanidinium HCL, 20 mM sodium acetate pH 5.2 and 200 mM NaCl) for 30 min. After stirring, samples were tissumized for 25 s intervals at 30% approximately 4-5 times until the viscosity of the lysate resembled water. The lysate was spun at 16,000 rpm for 30 min and the supernatant was filtered with a 0.4 µM syringe filter.

The filtered lysate was applied to a 5 mL His-Trap HP column in nickel loading buffer (8 M urea, 20 mM sodium acetate pH 5.2, 200 mM NaCl and 20 mM Imidazole) and eluted utilizing a gradient of the nickel elution buffer (8 M urea, 20 mM sodium acetate pH 5.2, 200 mM NaCl and 1 M Imidazole). Isolated fractions containing each MM core histone (H2A, H2B, H3 and H4) were combined and run over a TSK-SP cation exchange column using the SAUDE 200 buffer (8M urea, 20 mM sodium acetate pH 5.2, 200 mM NaCl, 5 mM βME and 1 mM EDTA) and eluted using a gradient of the SAUDE 1000 buffer (8 M urea, 20 mM sodium acetate pH 5.2, 1000 mM NaCl, 5 mM βME and 1 mM EDTA). For the putative linker histone H1, fractions from the His-Trap HP column were placed over a Mono-Q anion exchange column in SAUDE buffer. Eluted fractions, from

the TSK-SP or Mono-Q, containing each individual histone were combined and dialyzed into 5mM βME for lyophilization. All histones were lyophilized and stored at -20°C[127]. Confirmation of each histone protein ID was performed through LC-MS/MS (Jeremy Balsbaugh, UConn).

For protein refolding, putative linker histone H1 was dialyzed alone into refolding buffer (20 mM Tris-HCL pH 7.5, 2 M NaCl, 1 mM EDTA, 1 mM βME) overnight with multiple buffer changes. Precipitated protein was discarded through centrifugation and the supernatant was concentrated for application over a size-exclusion S200 equilibrated in refolding buffer. The sample was stored in 20% glycerol at -80°C. Conversely, the predicted core histones (H2A, H2B, H3 and H4) of MM were refolded together either as an octamer, as dimers (H2A-H2B), or as putative tetramer $(H3\text{-}H4)_2$ utilizing previously described protocols[127]. Samples eluted off gel filtration at expected volumes and were stored in 20% glycerol at -80 °C.

3.5.3 MM-NLP reconstitution

The DNA utilized to reconstitute the MM nucleosome-like particle (NLP) *in vitro* included the Widom 601 DNA at 147 bp (5'– ATCTGAGAATCCGGTGCCGAGGCCGCTCAA TTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGG GATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCCGAT–3'), 165 bp (5'– ATCGCCAGGCCTGAGAATCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACG CACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATA TATACATCCAGGCCTTGTGGAT –3') and 207 bp (5'– ATCTAATACTAGGACCCTATACGCGGCCGCATCGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGA CAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTA

GTCTCCAGGCACGTGTCAGATATATACATCGATTGCATGTGGATCCGAATTCATATTAATGAT –3') lengths; along with an in-house generated 150 bp 50% G/C content double stranded DNA (5'– GCTAGT CCGTCTTCTACTCTGAAATGAGCAGTCCTAGTCAGCAAGATCGCTCAGCCAACTTTCTACCAGCGCAACCCT AATCTACCCCATGAATGAAGCCGCACCCAAAACCGCATTCTAAGGAGTGACATTAACCCTCGGTGAGGATGT –3')[112]. MM octamer and DNA were mixed at a range of ratios of DNA to octamer and reconstituted by gradient dialysis (ideal ratio 1.0: 1.6). All eukaryotic histones (*Xenopus laevis*) utilized in reconstitution of eNuc controls were supplied by the Histone Source (Hataichanok Scherman, Colorado State University). Reconstituted samples were analyzed by 5% native-PAGE .

3.5.4 Sedimentation velocity analytical ultracentrifugation (SV-AUC)

To evaluate the homogeneity, molecular size, and molecular shape of MM-NLPs and MM-LE-tri-NLP in solution; we used SV-AUC with absorbance optics (λ = 280 nM). MM-NLP of various DNA lengths and MM-tri-NLP, all at 250 nM, were spun at 30,000 – 35,000 rpm at 20ºC in the Beckman XL-A ultracentrifuge using the An60Ti rotor (50 mM NaCl, 20 mM Tris-HCL, 1 mM EDTA, pH 7.5 and 1 mM DTT). Sedimentation velocity data analysis were performed in UltraScan III version 4.0 to determine sedimentation coefficients ($S_{(20,W)}$), frictional ratios (f/f_0) and molecular weights of each sample using established protocols [80,81,108]. Integral $S_{(20,W)}$ distributions plots are displayed using GraphPad Prism version 10.0.0 for Windows, GraphPad Software, Boston, Massachusetts USA, www.graphpad.com.

3.5.5 Thermal Stability Assay

MM-NLPs and eNucs were reconstituted on Widom 601 at 207 bp and Random 150 bp DNA as described above. MV-NLPs were also reconstituted on Widom 601 at 207 bp DNA as previously described[108]. Two replicates of each nucleosome (800nM) were incubated in the presence of SYPRO Orange for 1 min at 25 ° C, in 20 mM Tris-HCL (pH 7.5), containing 5 mM DTT and 50 mM NaCl. SYPRO Orange is provided SIGMA-ALDRICH as a 5000x concentrated solution (Catalog #: S5692). Each reaction was performed in a final volume of 20 µL, containing 2.5 µL of 62.5-fold diluted SYPRO Orange 5000x (final concentration 8x). Samples were increased by 1 °C and maintained for 1 min at the increased temperature. Fluorescence measurements were measured using a StepOnePlus Real-Time PCR unit (Applied Biosystems) after each 1 min incubation step. This sequential process was repeated for each 1 °C from 25 °C to 95 °C measuring using the Cal Orange filter (fluorescence emission maximum, 560 nm). Normalized plots displaying the relative fluorescence release of SYPRO Orange over 25 °C to 95 °C are displayed using GraphPad Prism version 10.0.0 for Windows, GraphPad Software, Boston, Massachusetts USA, www.graphpad.com. The thermal melting point (Tm) ranges of each NLP sample (n=2) measured in the thermal stability assay were determined by locating the temperature at the lowest point of the fluorescence derivative (-d/dU RFU).

3.5.6 Sucrose gradient sedimentation and gradient fixation (GraFix) crosslinking

A continuous 10-30% (w/v) sucrose gradient was prepared in a 13.2 mL Beckman polypropylene coulter tube (331372). To form the gradient, 6 mL of the top solution (50 mM NaCl, 20 mM HEPES, 1 mM EDTA, pH 7.5 and 10% sucrose) was added to the coulter tube and then 6 mL

of the bottom solution (50 mM NaCl, 20 mM HEPES, 1 mM EDTA, pH 7.5 and 30% sucrose) was slowly added from the bottom of the tube, pushing the top solution up until the demarcation line reached halfway. A gradient maker (BioComp Gradient Master) was then used to generate the continuous gradient. A 200 µL MM-NLP sample (3 µM) was then loaded on top of the gradient and spun at 4ºC for 18 hr at 30,000 rpm (Beckman, Rotor SW-41Ti). Samples were fractionated with absorbance optics (λ = 260 nm) for identification of complexes (BioComp Gradient Master). Desired complex fractions were determined using 5% native-PAGE and SDS-PAGE to be combined and dialyzed to remove sucrose (50 mM NaCl, 20 mM Tri-HCL, 1 mM EDTA, pH 7.5 and 1 mM DTT). Gradient fixation (GraFix) was performed utilizing the same procedure, with the addition 0.15% glutaraldehyde in the bottom solution to form a continuous gradient with crosslinker. After centrifugation, fractions were dialyzed to remove sucrose and quench the crosslinking reaction (50 mM NaCl, 20 mM Tri-HCL, 1 mM EDTA, pH 7.5 and 1 mM DTT).

3.5.7 Homology Modeling

Initial homology modeling of the MM histones ORF 318 (H2A), ORF 61 (H2B), ORF 255 (H3) and ORF 254 (H4) were constructed using SWISSMODEL, with *Xenopus laevis* histones in the context of a nucleosome (1AOI) as a reference[65,88,89]. To identify steric clashes in the initial model, CPPTRAJ of the Amber MD package (v18) with a cutoff distance of 0.8 Å between over-lapping atoms was utilized[91]. Each clash was manually addressed by modifying rotamers to reduce overlap and to achieve energy minimization in Chimera using default settings[61,86,128]. This initial MM nucleosome conformation was further refined by fitting into final 3D electron maps as described below.

3.5.8 Single particle cyro-electron microscopy (cryo-EM) and data processing

MM-NLP particles isolated from both sucrose gradient sedimentation (native) and GraFix (crosslinked) were concentrated to 2 µM using Amicon Ultra-4 centrifugal filters (Ultracel 30K, Millipore). C-Flat 1.2/1.3 (Cu) grids were glow discharged (Tergeo-EM Plasma Cleaner) at 40mA for 30 s before 4 µL of sample was applied to the grid and plunged into ethane using the Vitrobot Mark IV at 100% humidity at 4ºC with no wait time, a 2 s blot, and a blot force of 5. Micrographs of the MM native and crosslinked nucleosome particles were acquired with a nominal magnification of 64000x on a GEI Titan Krios (300 kV) outfitted with a Gatan K3 direct detection camera. The raw pixel size was 1.017 Å with movies captured in super-resolution mode maintaining an electron dose rate of 46.29 e/Å^2. The defocus range was-0.8 to-2.2 µM.

Both datasets were processed initially by cryoSPARC (v2.12.4) through motion correction and CTF estimation[84,129]. Exposures were curated to exclude sub-optimal characteristics by inspecting the CTF Fit resolution (Å), relative ice thickness and defocus range. Approximately 1000 particles were manually picked from the curated exposures to generate a picking model through Topaz (downsampling=16). For the crosslinked MM-NLP, Topaz picking yielded 451,081 particles which were subjected to two iterative rounds of 2D classification in order to discard bad particles[130]. Selected particles (162,284) were used to generate three *ab initio* models and the 3D model with the best directional distribution was selected to improve to the final resolution using homogeneous refinement, non-uniform refinement and local refinement (154,200 particles) (**Figure S3.3B**)[131,132]. For the native MM-NLP, Topaz picking yielded 72,768 particles and these particles were also subjected to two iterative rounds of 2D classification to remove undesirable

particles. This yielded 72,768 particles which were utilized to generate *an initio* model that was similarly subjected to homogeneous refinement, non-uniform refinement, and local refinement to improve resolution (Figure S3A).

3.5.9 Structural characterization of MM nucleosomes

Comparisons between the MM nucleosome (8UA7), MV nucleosome (7N8N) and eukaryotic nucleosome (1AOI) were conducted using ChimeraX[133–135]. Structures were aligned with the matchmaker tool in ChimeraX using the best-aligning pair of protein chains between the reference and matching structure. All residue and feature comparisons were determined by previously identified canonical eukaryotic histones features and figures were rendered with ChimeraX[25,133].

3.5.10 Structural prediction of putative linker histone MM-H1

The sequence of the putative linker histone MM-H1 was folded using AlphaFold (v.2.3.2)[136]. The highest ranked model was then aligned to the *X. laveis* linker histone H1 in PDB 5NL0 using ChimeraX.

3.5.11 MM-tri-NLP reconstitution

Tri-NLP reconstitution was performed using the Linker Ended Widom 601-207 bp x3 repeat DNA (LE-Tri) (5'– ATCTAATACTAGGACCCTATACGCGGCCGCATCGGAGAATC CCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCCCC

CGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCGATTGCATGTG
GATCCGAATTCATATTAATCATATCTAATACTAGGACCCTATACGCGGCCGCATCGGAGAATCCCGGTGCCGA
GGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAA
CCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCGATTGCATGTGGATCCGAATT
CATATTAATCATATCTAATACTAGGACCCTATACGCGGCCGCATCGGAGAATCCCGGTGCCGAGGCCGCTCAA
TTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGG
GATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCGATTGCATGTGGATCCGAATTCATATTAATGAT
–3’), and an in-house generated 500 bp 50% G/C content double stranded DNA (5’–
GCTAGTCCGTCTTCTACTCTGAAATGAGCAGTCCTAGTCAGCAAGATCGCTCAGCCAACTTTCTACCAGCGC
AACCCTAATCTACCCCATGAATGAAGCCGCACCCAAAACCGCATTCTAAGGAGTGACATTAACCCTCGGTGA
GGATGTCCATACAAGCACCTCCTACTACGGATCGAACCGTTAGTTCCCCAACTAAGTCCAAACCGTTAGACC
GCTTTCCGTACCATTCCGGTACTTATCTTCGCCACAACCTGAGACAATCCCAAGCTTAAGGCTCGACACAGA
CTGACGAAGGATATATCTCGCCCTAACCGTACCTCTATACCGCCATGAAGGAAGTGCCAAGTAGCCACAGAA
CCTTGGGATAGCAAGACTCTATGTCCCAGACCTCACTAACACCGAAGGAAAGTACCCACACAGACATCAGG
AAAACCCTCTGACCACTACGGCGAATGAAAAGTCCAGAGGACCAATACGTTACAGAGGCGACTGGATGT –
3’)[112]. MM octamer and DNA were mixed at a 1.0 : 4.8 ratio of DNA to octamer and reconstituted into tri-nucleosomes by gradient dialysis[60]. All eukaryotic histones (*Xenopus laevis*) utilized in reconstitution of LE-tri controls were supplied by the Histone Source (Hataichanok Scherman, Colorado State University). Reconstituted samples using the LE DNA were tested for quality by performing EcoRI restriction enzyme digestion to monitor the generation of mono-nucleosomes via EcoRI site within linker DNA between nucleosomes. Reconstituted samples were tested for quality using 4% native-PAGE and mass photometry to confirm a homogenous sample.

3.5.12 Mass Photometry

Mass photometry measurements were performed on a Refeyn TwoMP mass photometer (Refeyn Ltd). Glass coverslips were first cleaned with isopropanol, deionized water, and dried with N_2 gas, before being coated with a 0.01% Poly-L-Lysine solution for 20 seconds, rinsed with water, and dried with N_2 gas. To form a sample chamber, self-adhesive silicon gaskets were adhered to the top of the treated coverslip. For each measurement, the coverslip was placed on the oil-immersion objective lens, centered on a single well, and 13.5 ul of sample buffer (20 mM HEPES, pH 7.5, 100 mM KCl) was added to the well and the focal position of the glass surface was determined and held constant using an autofocus system. Nucleosome samples were first diluted to 100-200 nM, before a final 10-fold dilution onto the sample stage (final concentration of 10-20 nM). All dilutions were performed at room temperature in sample buffer (20 mM HEPES, pH 7.5, 100 mM KCl). A 60 second video was recorded immediately after the final dilution. A fresh well and dilution was used for each measurement and repeated at least three times for each sample. Tri-nucleosomes were diluted in buffer so that the number of detected events (particle counts) during the 60 second measurement was roughly 4,000 to 9,000 for an optimum data acquisition and processing. A known mass standard (β-amylase and thyroglobulin) was used to convert image contrast-signal into mass units. To calculate the molecular weight of the main species observed on the particle counts versus molecular mass distribution histograms we used the Gaussian function in the DiscoverMP software.

3.5.13 Atomic Force Microscopy (AFM)

MM-tri-NLP and eNuc-tri were reconstituted via salt gradient dialysis as previously described[34]. AFM slides were prepared by freshly cleaving the mica, treated with APTES for 30 minutes, rinsed with water, and then dried with nitrogen gas utilizing a 0.22 µm PES filter. Tri-nucleosomes samples were diluted in 20 mM Tris-HCl pH 7.5, 50 mM NaCl, and 1 mM EDTA to a final concentration of 1 ng/µl. Immediately after dilution, samples were applied to the APTES mica slide for 2 minutes, rinsed with water, and dried with filtered N_2 gas. For H1 samples, the respective tri-nucleosome was incubated with H1 in a 1:4 ratio for 30 minutes at room temperature, before being diluted 2-fold and deposited on the APTES surface (as described above). All samples were imaged in air on JPK/Bruker NanoWizard 3.0 with TAP300-Gold (Ted Pella) cantilevers. Images were collected at a scan size of 1×1 or 2×2 µm, a resolution of 512 × 512 pixels, and at 1-3 Hz.

AFM data was leveled, processed, and analyzed using Gwyddion software[137]. For particle height analysis, a mask at 1.5 nm was applied to identify particles, followed by extraction of the maximum height for each particle. Rare particles (>8 nm) were excluded as debris or aggregation. For compaction analysis, total distance between nucleosomes was measured in Image J by manually measuring perimeter of a triangle formed between the three nucleosomes. Of note, only tri-nucleosomes where three nucleosomes were visible were included in this analysis. Analysis and graphing of extracted data was completed.

3.5.14 Negative-stain electron microscopy (EM)

For negative staining, 4 µL of native of MM-LE-tri (50 nM) were applied onto glow-discharged 2 mm Cu 200 mesh grids (Electron Microscopy Services). Micrographs of negatively

stained images were recorded on the Tecnai T12 at 100 keV under low dose conditions (~20 e/Å^2) at a nominal magnification of x 96,000 on the 2k x2k AMT CCD side-mount and 4k x 4k FEI Eagle CCD bottom-mount cameras with a defocus range of 750 nm.

3.5.15 Circular Dichroism (CD)

Far-UV CD spectra of H1 proteins (0.5 mg/ml) were recorded at room temperature on a ChirascanPlus (Applied Photophysics Ltd, UK) spectrometer using quartz tubes (0.5 mm optical path length). The measurements were recorded in the 180-260 nm wavelength range with a 0.5 nm step size. All experiments were carried out in 0.02 M Sodium Phosphate and 0.2 M Sodium Fluoride at pH 7.5. For each H1 protein, five replicate CD spectra were averaged, baseline-corrected for signal contributions by the buffer. For secondary structure analysis of CD spectra, the DichroIDP program was used[138]. This CD analysis program was chosen as it is suitable for analyses of proteins containing significant amounts of disordered structures (as seen in eH1.0). Data was graphed using Graph Pad Prism (GraphPad Software, Boston, Massachusetts USA, www.graphpad.com).

3.6 Supplemental Materials

Figure S3.1A

Figure S3.1B

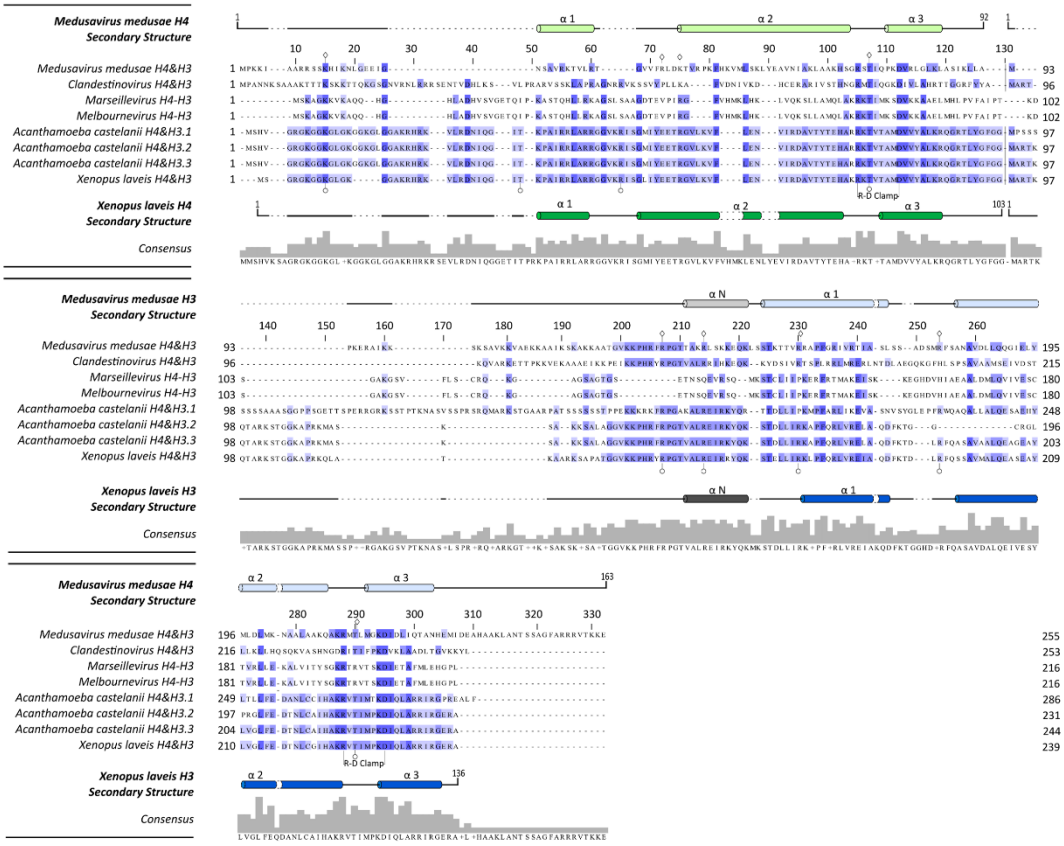

Medusavirus medusae H4
Secondary Structure
Medusavirus medusae H4&H3
Clandestinovirus H4&H3
Marseillevirus H4-H3
Melbournevirus H4-H3
Acanthamoeba castelanii H4&H3.1
Acanthamoeba castelanii H4&H3.2
Acanthamoeba castelanii H4&H3.3
Xenopus laveis H4&H3
Xenopus laveis H4
Secondary Structure
Consensus
Medusavirus medusae H3
Secondary Structure
Medusavirus medusae H4&H3
Clandestinovirus H4&H3
Marseillevirus H4-H3
Melbournevirus H4-H3
Acanthamoeba castelanii H4&H3.1
Acanthamoeba castelanii H4&H3.2
Acanthamoeba castelanii H4&H3.3
Xenopus laveis H4&H3
Xenopus laveis H3
Secondary Structure
Consensus
Medusavirus medusae H4
Secondary Structure
Medusavirus medusae H4&H3
Clandestinovirus H4&H3
Marseillevirus H4-H3
Melbournevirus H4-H3
Acanthamoeba castelanii H4&H3.1
Acanthamoeba castelanii H4&H3.2
Acanthamoeba castelanii H4&H3.3
Xenopus laveis H4&H3
R-D Clamp
Xenopus laveis H3
Secondary Structure
Consensus

Figure S3.1C

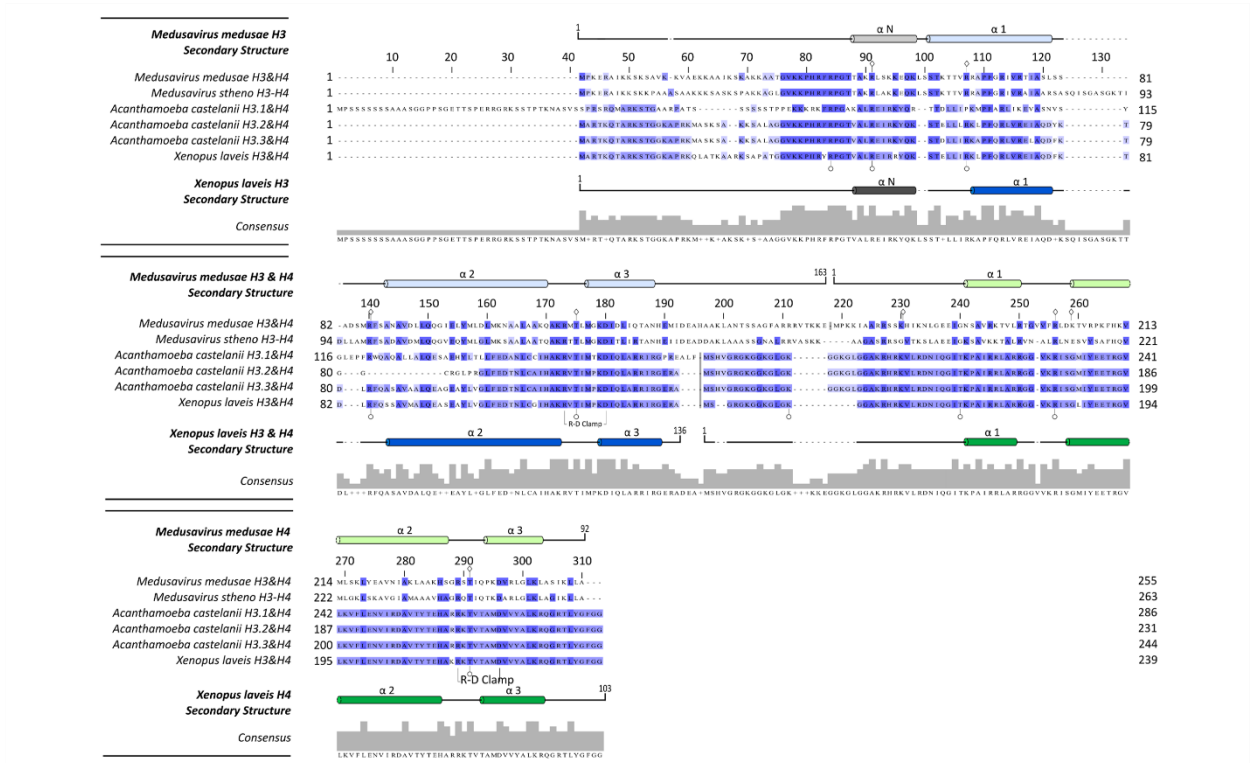

Medusavirus medusae H3
Secondary Structure
Medusavirus medusae H3&H4
Medusavirus stheno H3-H4
Acanthamoeba castelanii H3.1&H4
Acanthamoeba castelanii H3.2&H4
Acanthamoeba castelanii H3.3&H4
Xenopus laveis H3&H4
Xenopus laveis H3
Secondary Structure
Consensus
Medusavirus medusae H3 & H4
Secondary Structure
Xenopus laveis H3 & H4
Secondary Structure
Medusavirus medusae H4
Secondary Structure
Xenopus laveis H4
Secondary Structure
R-D Clamp

Figure S3.1 D, E, F

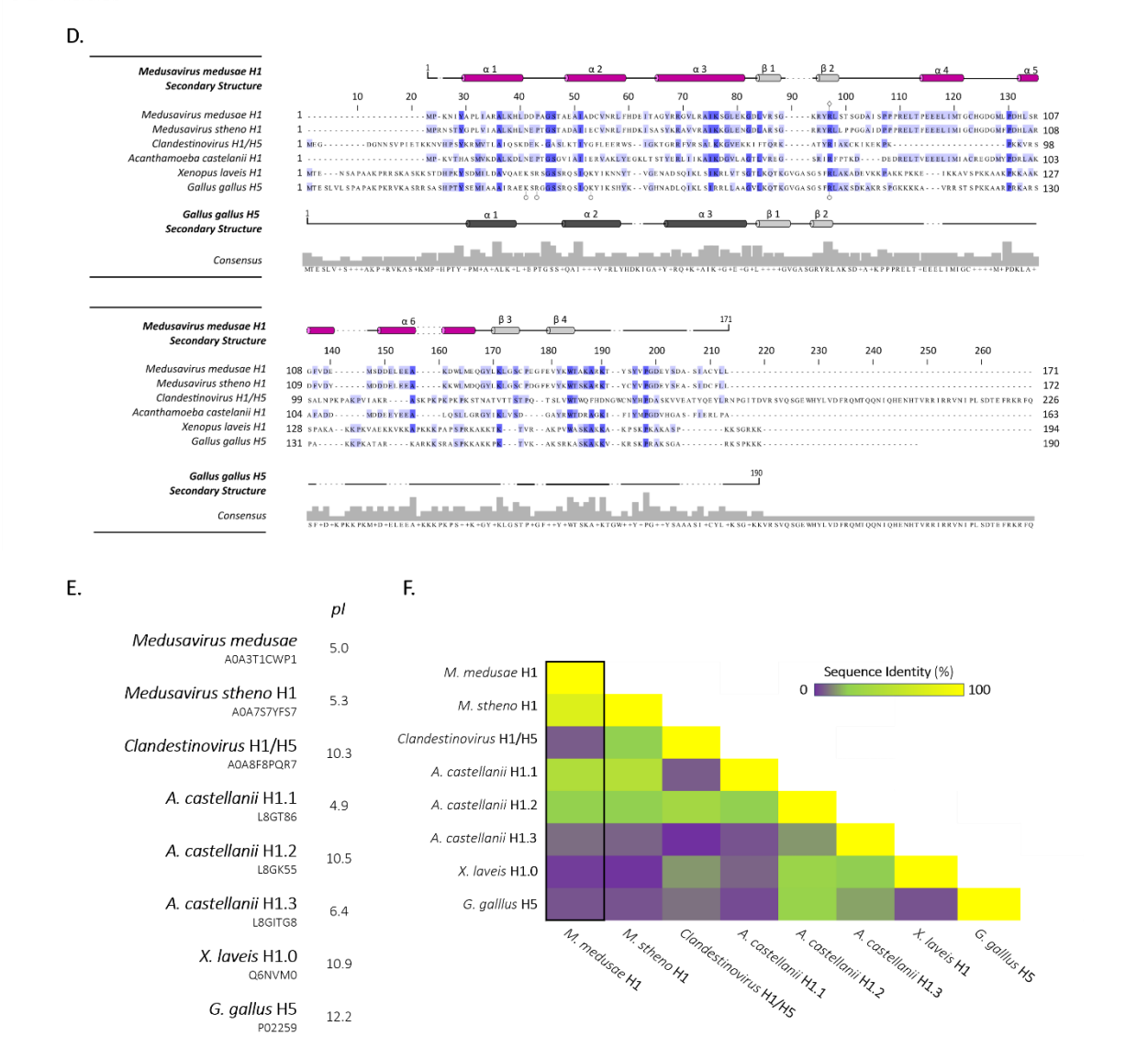


Figure S3.1. Complete sequence alignment and secondary structure prediction of *Medusavirus medusae* histones.

Viral histone dimer pairs (or doublets) H2B-H2A and H4-H3 were aligned against *A. castellanii* and *X. laevis* histones using HHPRED's multiple sequence alignment tool, ClustalΩ. Conservation of each residue within the alignment is represented by blue shading, where darker blue signifies a greater conservation. Known α helices of *X. laveis* are shown in dark colored tubes (H2B-red, H2A-yellow, H4-green, H3-blue). Predicted α helices of MM (light colored tubes) were generated using HHPRED's Quick 2D prediction web server. (A) Complete sequence alignment of H2B and H2A viral histones (*Mamonoviridae* and *Marseilleviridae* families) against viral host histones *A. castellanii,* and *X.laevis*. (B) Complete sequence alignment of H4 and H3 viral histones (excluding *M. stheno*

doublet) against viral host *A. castellanii*, and *X, laevis*. (C) Complete sequence alignment of H3 and H4 viral histones (including *Medusavirus stheno* H3-H4 doublet) against viral host *A. castellanii*, and *X. laevis*. Differs from previous alignment in order of histone pairs (H3-H4 instead of H4-H3). (D) Viral putative linker histone H1 was aligned against *A. castellanii, X. laevis*, and *Gallus gallus* H5. Known α helices of *X. laevis* H1 are shown in dark grey colored tubes. Predicted α helices of MM (pink) were generated using HHPRED's Quick 2D prediction web server. (E) The isoelectric point (pI) of each predicted and known linker histone H1. (F) Heat map comparing percent identity of predicted viral linker histone H1/H5 and eukaryotic H1/H5 sequences. MM-putative H1 is outlined in black. Related to Figure 3.1.

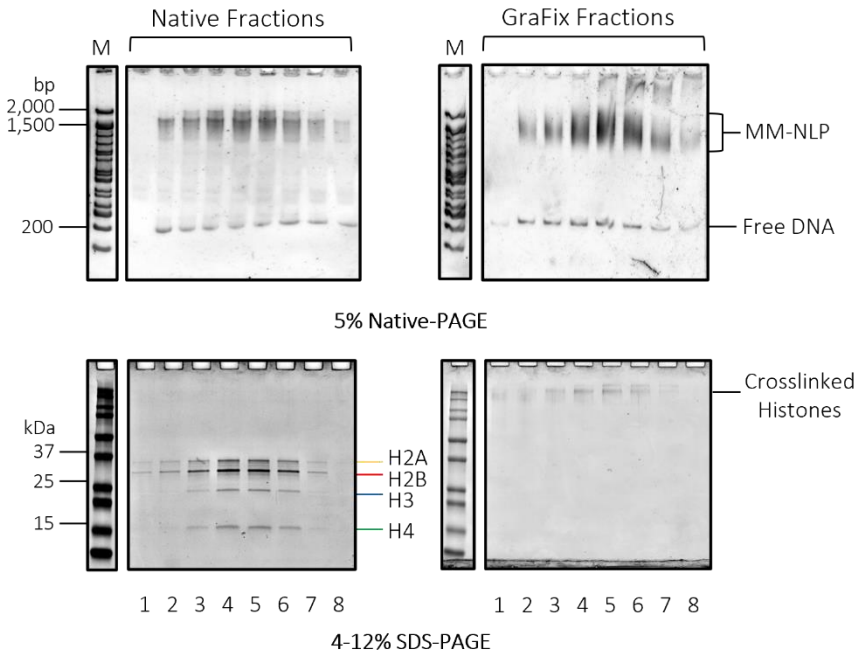


Figure S3.2 MM-NLP preparation for Cryo-EM.

Sucrose gradient sedimentation and GraFix of MM-NLP with 207 bp DNA (MM-NLP$_{207W}$). Fractions of each were analyzed by 4-12% SDS-PAGE and 5% Native-PAGE to determine composition of particles. Related to Figure 3.2 and 3.3.

‘

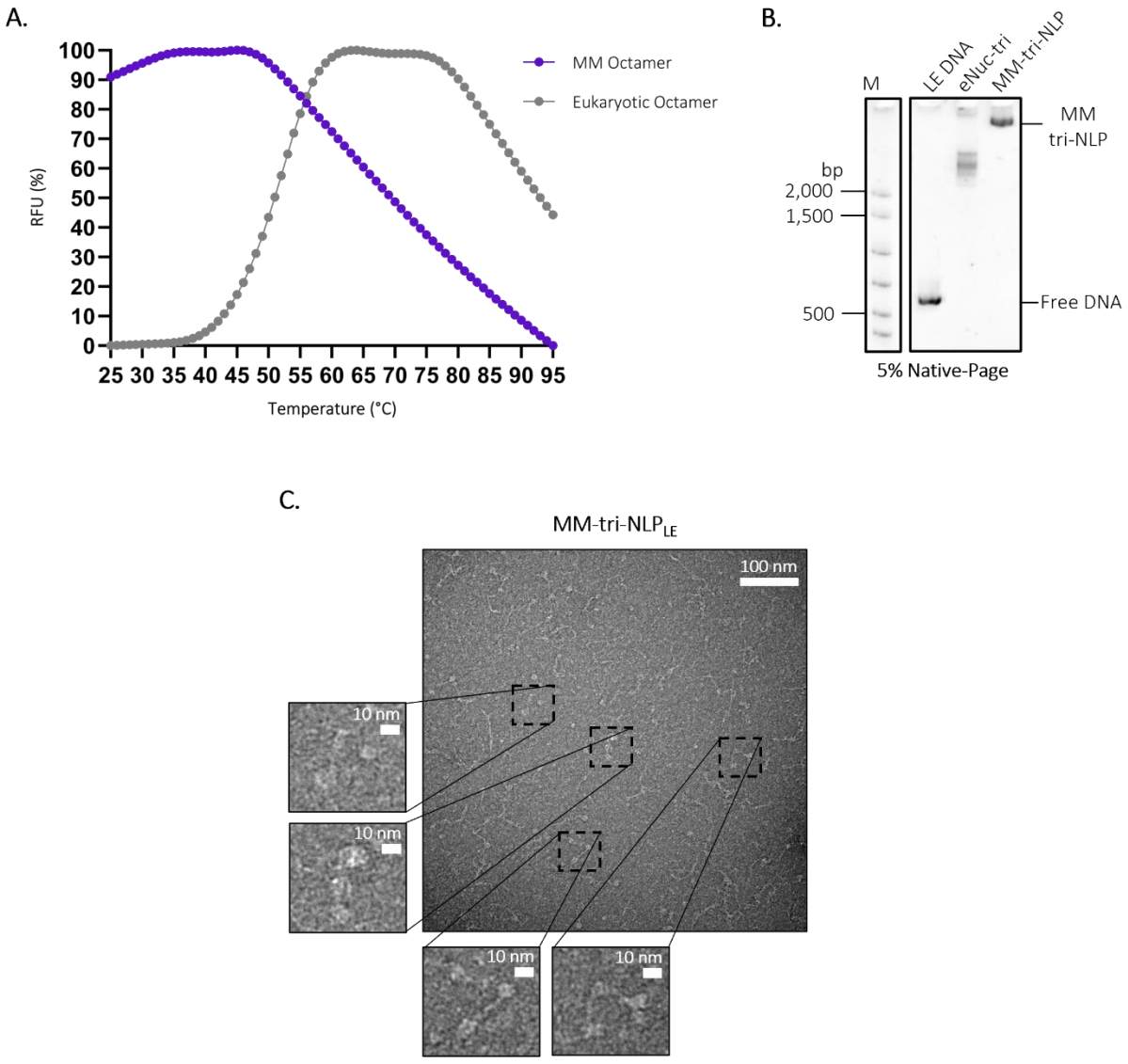


Figure S3.3 Biochemical analysis of *Medusa medusae* octamers and tri-nucleosomes.

(A) Thermal shift stability of MM and eukaryotic octamer utilized in formation of NLP. The raw relative fluorescence units were normalized for plotting. (B) MM tri-nucleosomes (MM-tri-NLP) and eukaryotic tri-nucleosomes (eNuc-tri) on LE DNA. (C) Micrograph of negatively stained MM-tri-NLP$_{LE}$. Related to Figure 3.2.

Figure S3.4A

A.

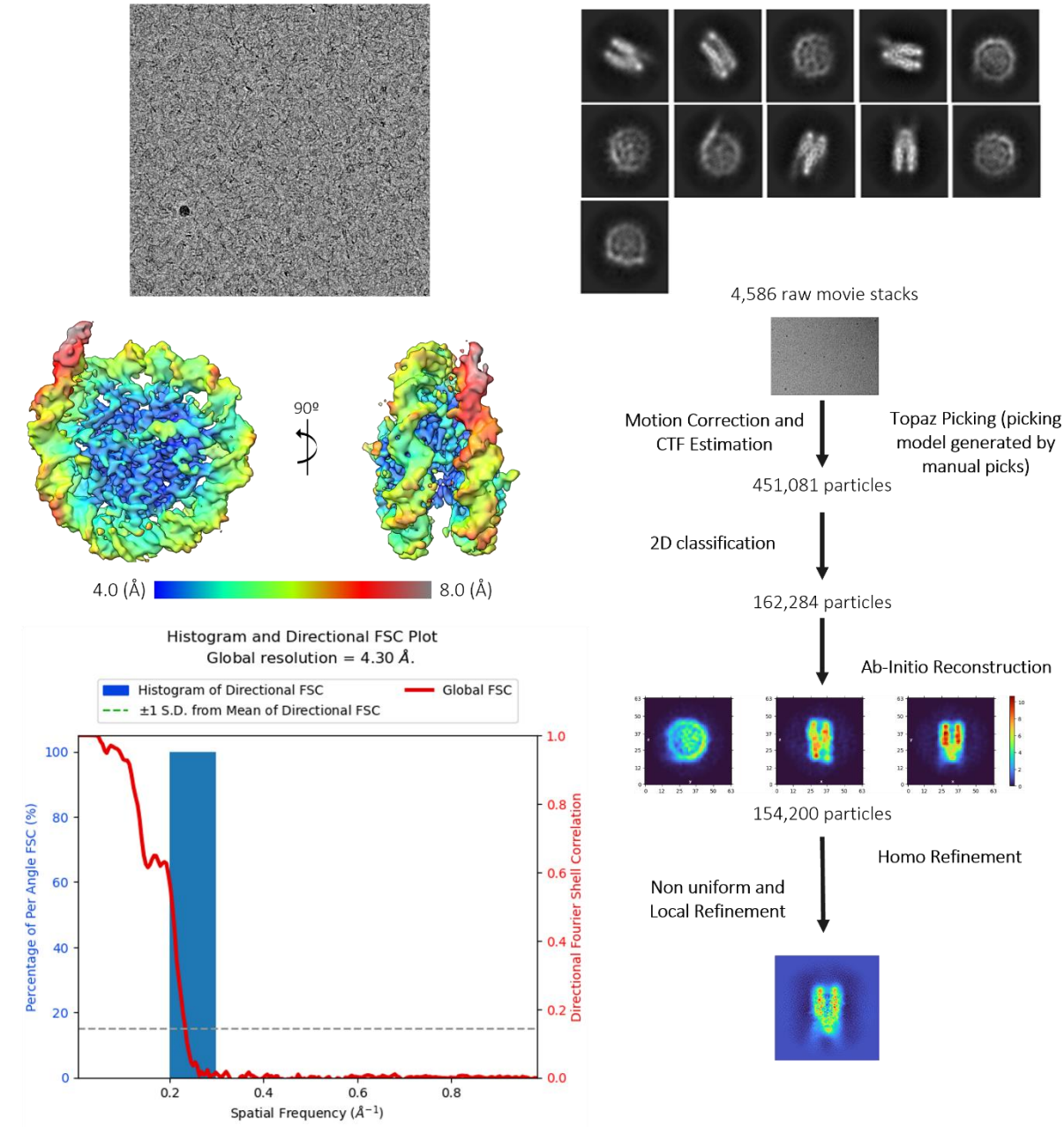


Figure S3.4B

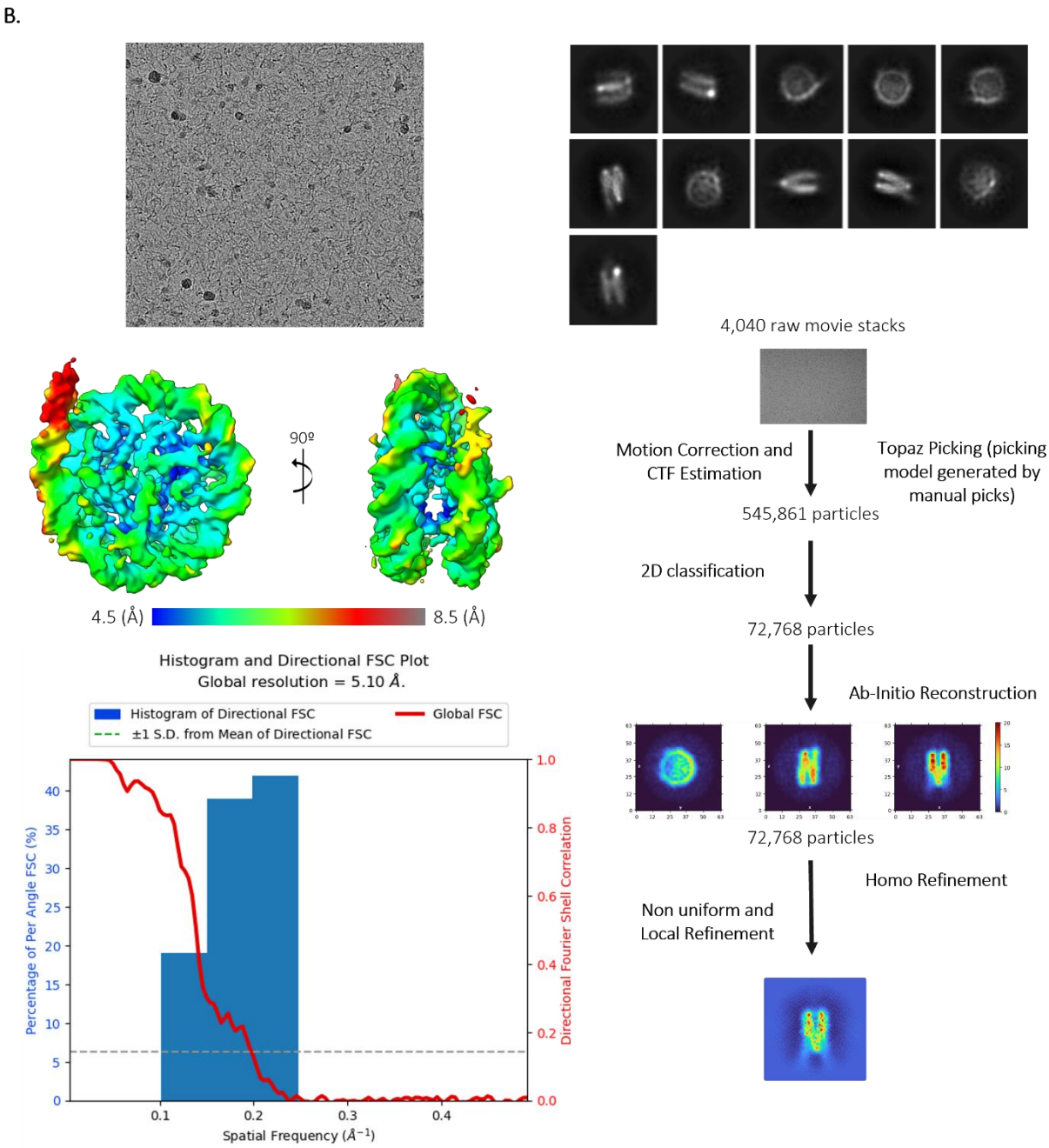


Figure S3.4 Cryo-EM analysis of Native and GraFix MM-NLP$_{207}$ bp.

(A) Raw micrograph of GraFix MM-NLP$_{207}$, 2D class averages generated from represented dataset, 3D structure of MM-NLP with local resolution map, FSC curve, and CryoSPARC data processing flow chart. (B) Raw micrograph of native MM-NLP$_{207}$, 2D class averages generated from represented dataset, 3D structure of MM-NLP with local resolution map, FSC curve, and CryoSPARC data processing flow chart. Related to Figures 3.4 and 3.5.

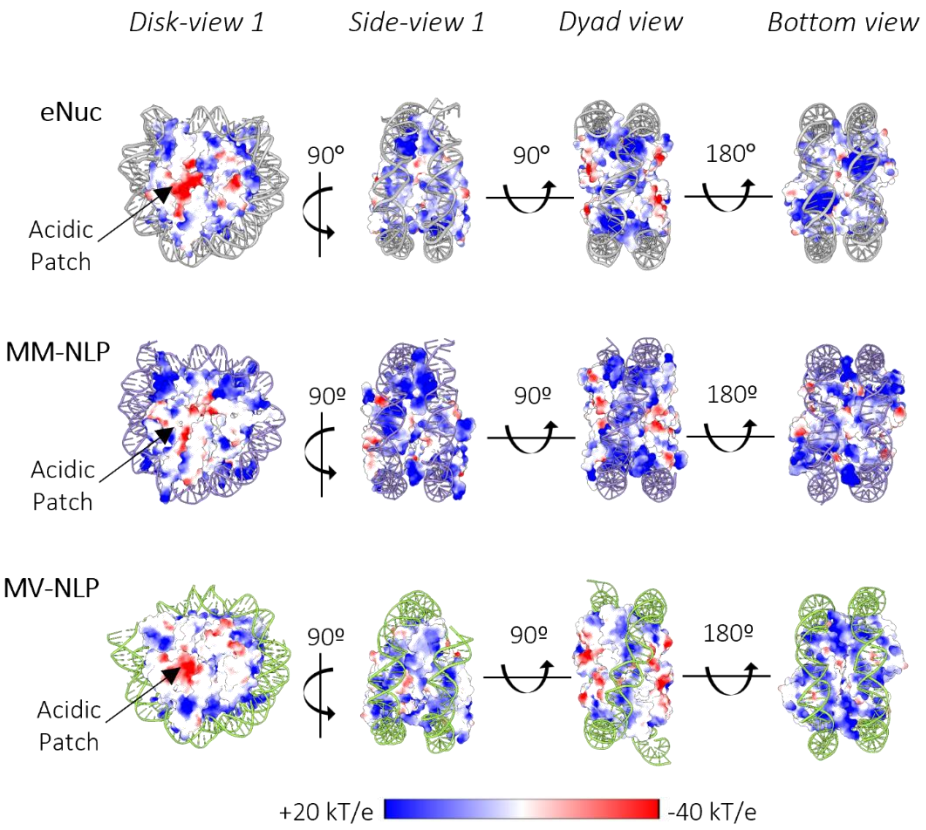


Figure S3.5 Electrostatic surface representation comparison of eNuc to viral NLPs.

Charged surface representation of histones from the eNuc (PDB ID: 3LZ0), MM-NLP and MV-NLP (PDB ID: 7N8N) in different orientations. Related to Figure 3.5.

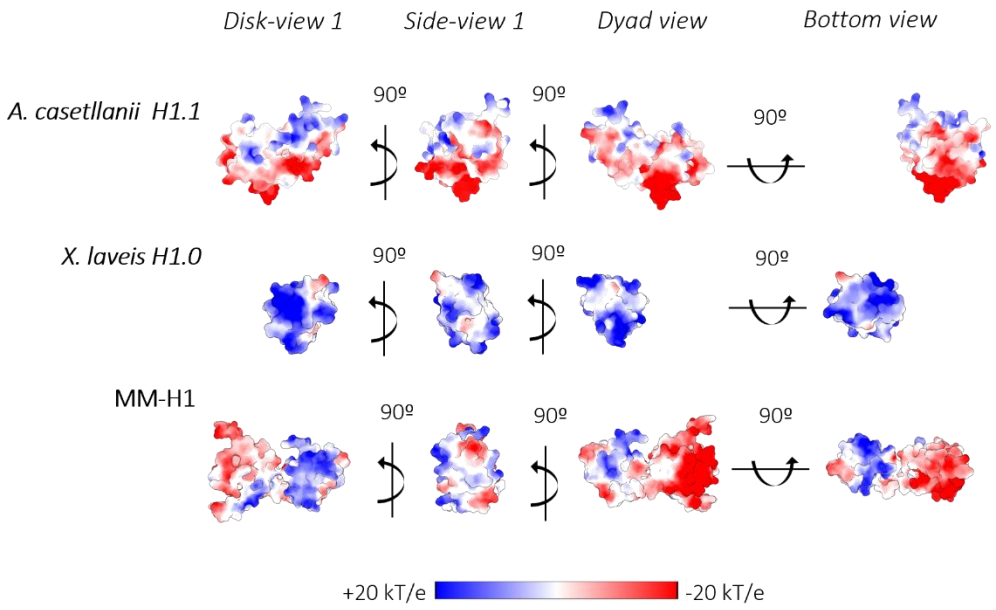


Figure S3.6 Electrostatic surface representation comparison of eNuc to viral NLPs.

Charged surface representation of *Acanthamoeba castellanii* H1.1, *Xenopus laveis* H1.0, and MM-H1 with rotational views. Coordinates for *Xenopus laveis* H1 were acquired from 5NL0. Coordinates for *Acanthamoeba castellanii* H1.1 and MM-H1 were determined through AlphaFold. Related to Figure 3.6.

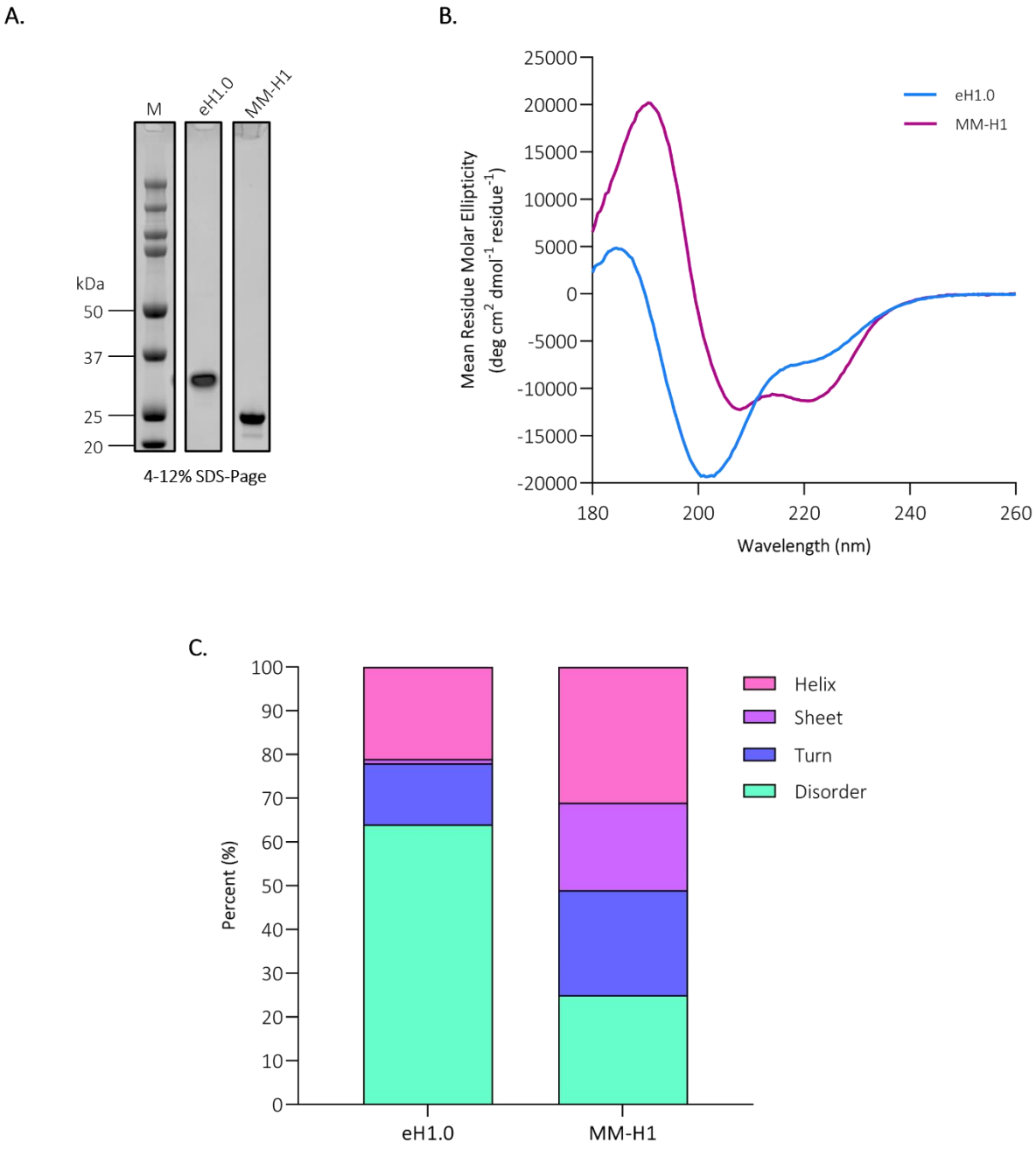


Figure S3.7 Biochemical analysis of *Mus musculus* and *Medusavirus medusae* histone H1.

(A) Purified MM-putative linker histone H1 (MM-H1) and *Mus musculus* H1.0 (eH1.0). (B) CD spectra of purified eH1.0 (blue) and MM-H1 (pink). (C) Secondary structure estimation based on experimental CD data (shown in B), using DichroIDP. Related to Figure 3.7.

Chapter 4

Additional Work

4.1 Hybrid Nucleosomes

4.1.1 *Medusavirus medusae* dependency on the host nucleus

The known NCLDV proteomes surprisingly maintain diverse sets of proteins associated with translation, DNA maintenance, and metabolism. As such, the giant viruses display a range of dependencies of their host nucleus for DNA replication[22]. The replication of Pandoraviruses and Mollivirus includes a complete disorganization of the nuclear membrane to release all nuclear replication machinery to the cytoplasm, while Marseilleviruses replicate in the cytoplasm by transient recruitment of specific nuclear machinery to the viral factory[58,139,140]. Notably, the *Medusavirus medusae* (Medusavirus) has demonstrated a distinctive dependence on the host *Acanthamoeba castellanii* (Amoeba) nucleus by replicating their viral DNA within the host nucleus over the first 14 hours of infection, before shuttling viral DNA to empty capsids located near the nuclear membrane (**Figure 1.3**). While Medusavirus does encode many eukaryotic homologs demonstrated in other NCLDV, it uniquely lacks the DNA-dependent RNA polymerase, mRNA capping enzyme, and topoisomerase II seen in other NCLDV; which is predicted to contribute to its need for the host nucleus[21]. In fact, recent RNA-sequencing of the Amoeba host suggests that there is remodeling of the host nuclear environment upon Medusavirus infection. Early

transcription of Medusavirus genes include a virally-encoded linker histone homologue H1 (MM-H1), while the core Medusavirus histones are transcribed at a later stage together with components of the virus factory[110].

When the Medusavirus histones are expressed, it is not currently known how they are transported and where through the time course of infection. Our most recent viral nucleosome structure reveals that Medusavirus histones indeed do assemble into nucleosome-like structures. Considering the notable dependence of Medusavirus on the host nucleus for viral DNA replication, Medusavirus histones brings about new speculation regarding their involvement in viral infection. Do they contain nuclear localization signals that transport them into the host nucleus to interact with viral DNA? Do they interact with host DNA? When are these viral histones assembled into nucleosomes and where? Ultimately these considerations start to probe if Medusavirus histones are able to interact with host histones to form nucleosomes.

4.1.2 Hybrid Octamer Refolding

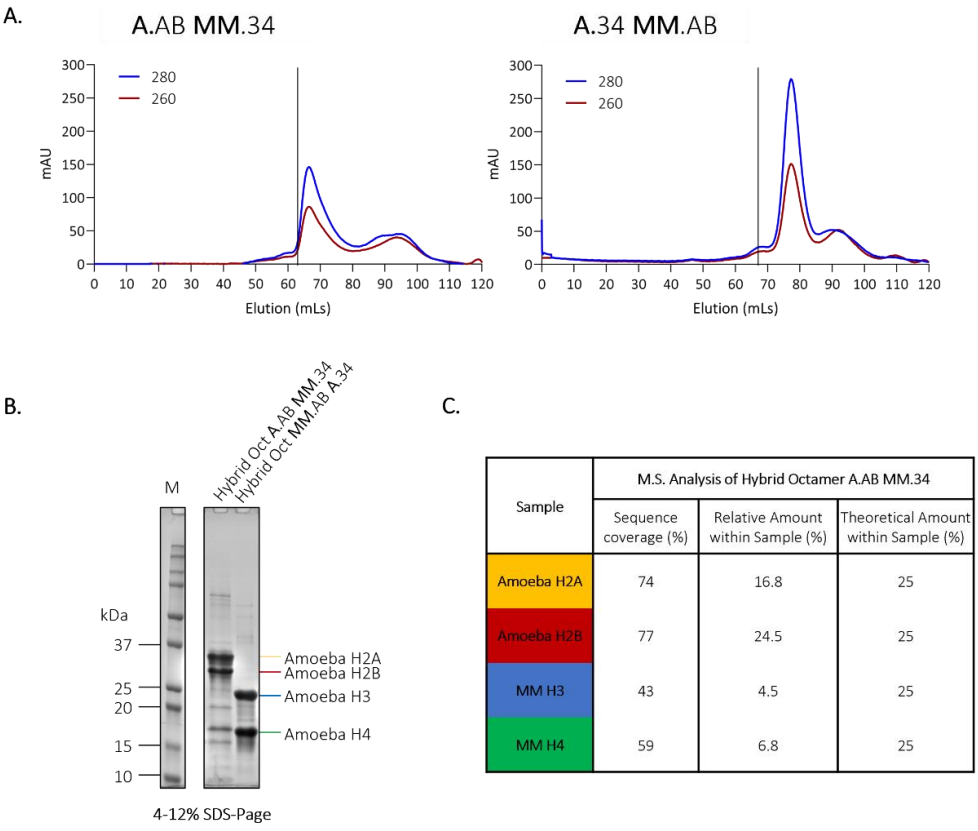
In vitro, as well as during chaperone-mediated nucleosome assembly, nucleosomes are assembled on ~150 bp of DNA utilizing histone H2A-H2B dimers and $(H3\text{-}H4)_2$ tetramers[26,127]. As all of Medusavirus DNA replication takes place within the host nucleus, it's possible that host histones are assembled onto viral DNA. This makes it difficult to keep host histones removed from the assembly process within the Medusavirus viral factory and overall maturation. Incorporation of host histones likely yields no benefit for the virus, suggesting they may maintain a process that avoids viral genome organization by host histones. In this scenario, the Medusavirus histones could combat host histones in the viral factories by making it difficult for the heterotypic histone

dimers/tetramers to coexist in the nucleosome context. These defined 'hybrid nucleosomes' are nucleosomes that are assembled by histones from different species, e.g. through the contribution of one or two H2A-H2B dimers from one organism and a $(H3\text{-}H4)_2$ tetramer from another organism. Understanding the capabilities of hybrid nucleosomes to form between Medusavirus and Amoeba allows us to further understand how Medusavirus utilizes histones during infection.

To investigate if histones from Amoeba and Medusavirus are compatible, we wanted to test the capability for MM histone dimers/tetramers to assemble with its host Amoeba histone dimers/tetramers into a nucleosome *in vitro*. We first attempted to refold the two different hybrid combinations, MM H2A-H2B and Amoeba H3-H4 (MM.AB A.34) versus Amoeba H2A-H2B and MM H3-H4 (A.AB MM.34), into histone octamers. Preliminary refolding yielded high amounts of precipitation for each octameric combination, which is not seen for homotypic histone octamers[60]. Following removal of precipitated protein, each refolded hybrid was placed over a sizing column (s200) to isolate each soluble product based on its expected octamer elution. Both of the refolded octamers eluted later than expected, suggesting that they each were either refolded as dimers or tetramers (**Figure 4.1A**).

The fractions from the largest 280 peak from each hybrid octamer were combined, concentrated, and run on SDS-Page for identification of refolded components. With each of the hybrid octamer combinations, only Amoeba histones primarily prevail after each refolding combination attempt (**Figure 4.1B**). The A.AB MM.34 hybrid octamer suggested some promise for the presence of both Amoeba and Medusavirus histones based on its gel filtration elution, however mass spectrometry analysis showed low amounts of Medusavirus H3 and H4 with low sequence coverage (**Figure 4.C**). This indicates that while Medusavirus H3 and H4 may be present in the

refolded A.AB MM.34 hybrid octamer, it is in minimal amounts and likely degrading within the sample. We conclude that these histones cannot be refolded together into an octamer *in vitro* (**Figure 4.1**). Therefore, we shifted our focus to test whether nucleosomes could be assembled from refolded Medusavirus and Amoeba dimer and tetramers, respectively, using salt dialysis. Within the cell, nucleosome assembly functions through the sequential deposition of $(H3\text{-}H4)_2$ tetramers followed by H2A-H2B dimers, and this order is mimicked in our salt-dependent deposition assembly protocols[60].



Sample	M.S. Analysis of Hybrid Octamer A.AB MM.34		
	Sequence coverage (%)	Relative Amount within Sample (%)	Theoretical Amount within Sample (%)
Amoeba H2A	74	16.8	25
Amoeba H2B	77	24.5	25
MM H3	43	4.5	25
MM H4	59	6.8	25

Figure 4.1 Refolding of hybrid MM and Amoeba octamers.

(A) S200 Gel filtration elution of refolded hybrid viral and eukaryotic octamers. Hybrid octamer **A**.AB and **MM**.34 contained host *A.castellani* (Amoeba) core histones H2A, H2B, along with MM H3, and H4. While hybrid octamer **MM**.AB and **A**.34 contained MM core histones H2A, H2B, along with host Amoeba H3, and H4. The absorbance at 280 and 260 is given across each elution profile

and the expected elution volume based on the expected molecular weight of each hybrid octamer is indicated by the black bar. (B) SDS-PAGE of hybrid octamers from combined fractions under the largest peak of each octamer elution in panel A. (C) Table of M.S. analysis on hybrid octamer **A**.AB **MM**.34. Sequence coverage of each histone is provided along with the relative amount of each determined by the normalized average precursor intensity. Theoretical amounts of each histone in a hybrid octamer are shown.

4.1.3 AlphaFold of Hybrid Octamers

Before pursuing hybrid nucleosome reconstitutions from dimers and tetramers of Medusavirus and Amoeba, we wanted to examine the ability of the two hybrid octamers (MM.AB A.34 and A.AB MM.34) to exist through AlphaFold predictions. Using the histone sequences of Medusavirus and Amoeba, we tested the combinations of the different H2A-H2B and H3-H4 histones from Medusavirus and Amoeba to further probe the capability of interactions with one another (**Figure 4.2**). While the histone tails of each octamer have a decreased confidence score

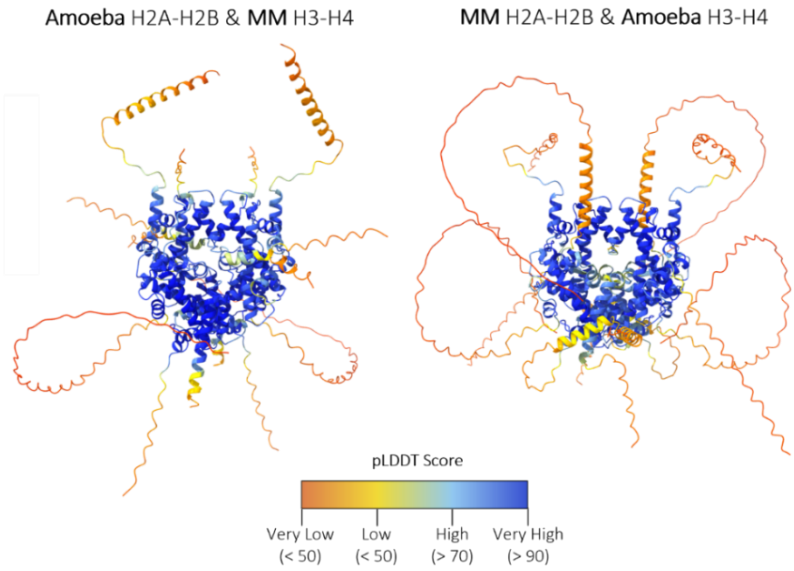


Figure 4.2 AlphaFold of hybrid MM and Amoeba octamers.

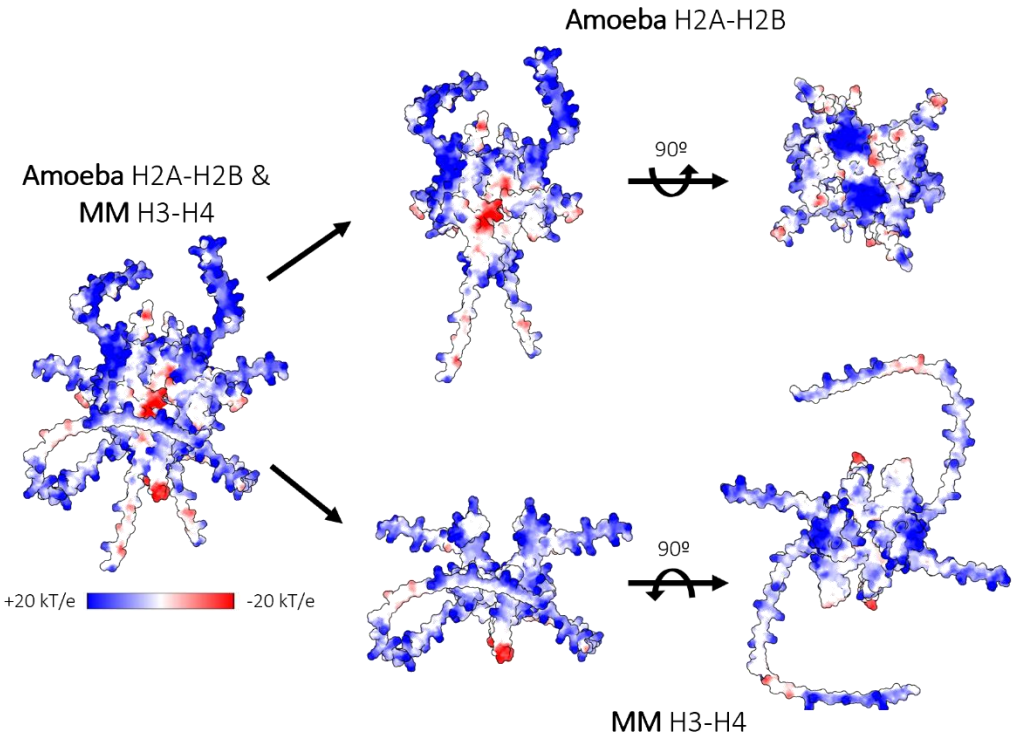
AlphaFold prediction of hybrid viral and eukaryotic octamers. Hybrid octamer A.AB and MM.34 contained host *A. castellani* (Amoeba) core histones H2A, H2B, along with MM H3, and H4. While hybrid octamer MM.AB and A.34 contained MM core histones H2A, H2B, along with host Amoeba

H3, and H4. The highest ranked model from each AlphaFold prediction is shown. The per residue confidence score (pLDDT) is provided along each hybrid octamer model with the higher confidence in blue and lower confidence in orange.

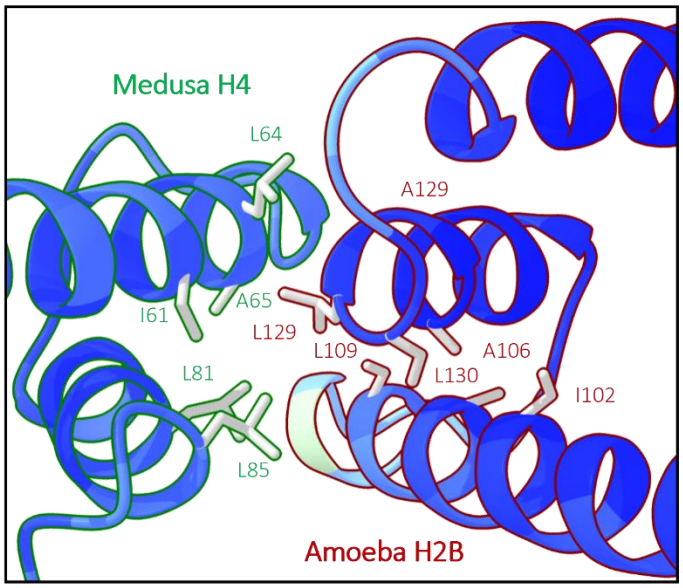
(orange), the octamer core of each hybrid remains very well predicted (blue, Figure 4.2). This suggests that while the histones do not seem to favor being refolded together *in vitro*, it is possible that dimer and tetramer complexes of Medusavirus and Amoeba may associate with one another to form hybrid nucleosomes.

With electrostatic surface representations of each dimer/tetramer and a closer look at the H4-H2B four-helix bundle, the potential interfaces of each heterotypic octamer were investigated deeper (Figure 4.3). The Amoeba dimer and Medusavirus tetramer core interface each maintain two hydrophobic regions that are opposite one another, creating a hydrophobic octamer core (Figure 4.3A). In fact, the Medusavirus H4 and Amoeba H2B demonstrate the characteristic H4-H2B four-helix bundle with strong hydrophobic interactions demonstrated throughout (Figure 4.3B). Similarly, the Medusavirus dimer and Amoeba tetramer demonstrate a similar hydrophobic core as seen for the A.AB and M.34 hybrid octamer (Figure 4.3C). However, the four-helix bundle between Amoeba H4 and Medusavirus H2B demonstrates less overall hydrophobic interactions than seen in the A.AB and M.34 hybrid octamer (Figure 4.3D). We conclude that the Amoeba and MM dimers/tetramers can theoretically assemble into heterotypic octamers together *in vitro*.

A.


Amoeba H2A-H2B
Amoeba H2A-H2B &
MM H3-H4
90º
90º
+20 kT/e
-20 kT/e
MM H3-H4

B.


Medusa H4
L64
A129
I61
A65
L129
L109
A106
L130
I102
L81
L85
Amoeba H2B

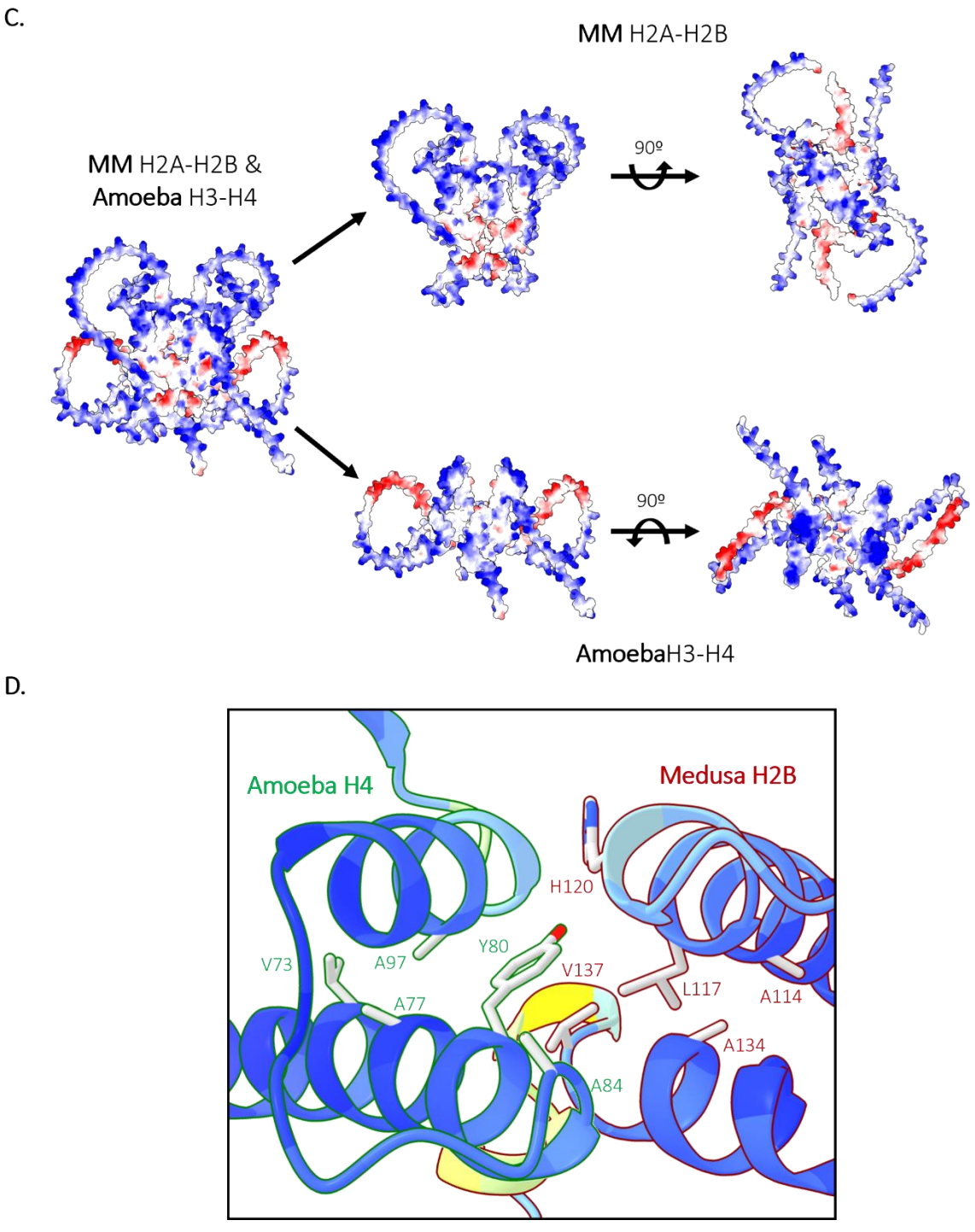


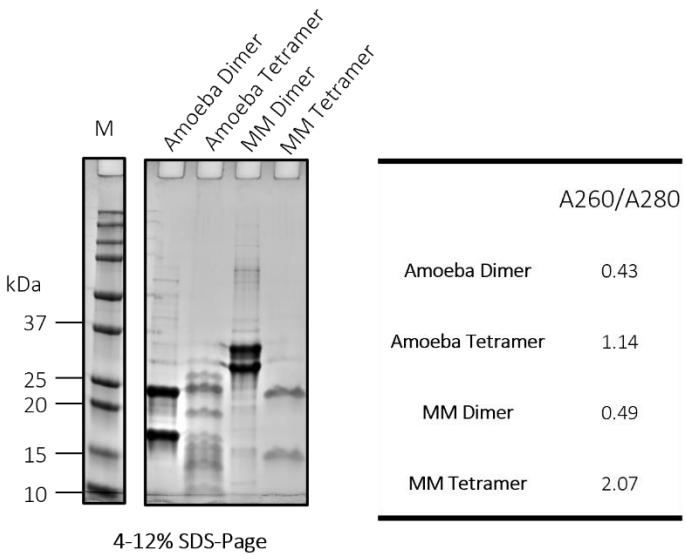
Figure 4.3 Interfaces of each hybrid Amoeba and MM octamer

(A) Electrostatic surface representation of hybrid octamer A.AB and MM.34 containing host A.castellani (Amoeba) core histones H2A, H2B, along with MM H3, and H4 (AlphaFold prediction). The histones from each organism are separated and their interacting electrostatic surfaces of the Amoeba dimer and MM tetramer are shown. (B) H4-H2B four-helix bundle interface of hybrid

octamer A.AB and MM.34. The histones are colored by their per residue confidence score (pLDDT) as shown in Figure 4.2. MM H4 is outlined in green while Amoeba H2B is outlined in red. Residues contributing to interactions are shown. (C) Electrostatic surface representation of hybrid octamer MM.AB and A.34 containing MM core histones H2A, H2B, along with host Amoeba H3, and H4 (AlphaFold prediction). The histones from each organism are separated and their interacting electrostatic surfaces of the MM dimer and Amoeba tetramer are shown. (D) H4-H2B four-helix bundle interface of hybrid octamer A.AB and MM.34. The histones are colored by their per residue confidence score (pLDDT) as shown in Figure 4.2. Amoeba H4 is outlined in green while MM H2B is outlined in red. Residues contributing to interactions are shown.

4.1.4 Ongoing effort to form hybrid nucleosomes

With the validation that Medsusavirus dimers/tetramer and Amoeba dimers/tetramers can, in principle, associate through AlphaFold, we attempted to purify and refold them individually (**Figure 4.4**). However, only the dimers from Medusavirus and Amoeba demonstrated successful refolding with low DNA contamination. Both the Medsauvirus and Amoeba tetramer initial refolding were unsuccessful because they were both hampered by high levels of degradation or DNA contamination (**Figure 4.4**). While not ideal, it is not entirely surprising that there are difficulties with the tetramer refolding of Medusavirus and Amoeba. Previous refolding of eukaryotic histone tetramers with reduced hydrophobic interactions indicated some difficulty with tetramer stability[141]. As the Medusavirus tetramer electrostatic surface suggests, there is an increase in charge within the Medusavirus tetramer that may decrease its overall stability without its corresponding Medusavirus dimers and DNA (**Figure 3.4E**). In addition, the H3 – H3' four helix bundle in Medusavirus demonstrates decreases hydrophobic interactions compared to eukaryotic nucleosomes and likely impacts the tetramer stability as well (**Figure 3.5C**). Various refolding attempts of the Medusavirus and Amoeba tetramers with cleaner individual histone preps are currently ongoing.



	A260/A280
Amoeba Dimer	0.43
Amoeba Tetramer	1.14
MM Dimer	0.49
MM Tetramer	2.07

Figure 4.4 Refolding of MM and Amoeba histones as dimers and tetramers.

Amoeba H2A and H2B histones were refolded as a dimer similarly to MM H2A and H2B histones. Amoeba H3 and H4 were refolded as a tetramer similarly to MM H3 and H4 histones. The ratio of A260/A280 of each histone refolding is provided (ideal ratio < 0.8).

4.2 Materials and Methods

4.2.1 Histone expression, purification, and hybrid octamer refolding

Medusavirus ORF 318 (H2A), ORF 61 (H2B), ORF 255 (H3), ORF 254 (H4) and Amoeba ORF (H2A), ORF (H2B), ORF (H3), and ORF (H4) were each cloned into a pET-28a plasmid for expression and purification from *E.coli*., utilizing adaptations of well-established eukaryotic histone protocols[60,127]. Purification and refolding were performed for each hybrid octamer combination (MM.AB A.34 and A.AB MM.34) as described in Section 3.5.2 (MM histone expression, purification, and refolding).

4.2.2 AlphaFold of hybrid octamers

The Medusavirus and Amoeba histones sequences of each hybrid combination (MM.AB A.34 and A.AB MM.34) was folded using AlphaFold-Multimer (v.2.3.2)[136]. The highest ranked model is displayed with each AlphaFold provided per residue confidence scoring (pLDDT) through ChimeraX[25,133].

4.2.2 Dimer and tetramer refolding

Medusavirus and Amoeba histone dimers/tetramers were refolded as described in in Section 3.5.2 (MM histone expression, purification, and refolding).

Chapter 5

Conclusions and Future Directions

5.1 Conclusions

Previously it was not known how virus-encoded histones organized DNA, and whether they are essential for viral fitness. Here, I demonstrated that Melbournevirus histone doublets and Medusavirus histones assemble into nucleosome-like particles with unique accommodations for their histone fusions, varied sizes of loops within the histone fold, and increased tail lengths. While these viral nucleosomes broadly resemble eukaryotic nucleosomes, they do maintain noticeable differences in their distinctive tails that lay across each H4 α 2 and key eukaryotic interactions within the four-helix bundles that are the key contacts that holds together the $(H3\text{-}H4)_2$ tetramer and tethers it to the H2A-H2B dimers. As a result, these nucleosomes are overall less stable than eukaryotic nucleosomes and likely much more dynamic. Additionally, I demonstrated that Medusavirus histones can assemble into tri-nucleosomes, suggesting that these viral histones are in fact involved in chromatin organization. Whether the Medusavirus histones organize their own viral genome or the host *Acanthamoeba castellanii* (Amoeba) genome is still unknown, but Melbournevirus histones have been demonstrated to densely package the viral genome[109]. Lastly, I found that the early transcribing putative linker histone H1 in Medusavirus does not perform its

predicted role of chromatin compaction. Since the Medusavirus H1 is remarkably similar to the Amoeba H1 and share several very unusual properties not found in the linker histones from other species, it is likely that the Medusavirus acquired the H1 gene from Amoeba and further evolved to serve its own purpose.

The work presented here demonstrates the first instance of virally encoded histones forming a nucleosome, including that they are essential to the virus, and describes the only two known viral nucleosome structures to date. While the contribution or relationship of these viral histones in eukaryogenesis is still debated, we showed that both the Melbournevirus and Medusavirus nucleosome-like particles are undoubtably similar to eukaryotic nucleosomes[47,6,41]. This may suggest horizontal gene transfer between the virus and host, and it is entirely possible that these viral nucleosomes represent the origin of the 'modern' eukaryotic nucleosome. However, it is more likely that an ancestral NCLDV was involved in infecting a proto-eukaryote with early versions of histones, and that the virus acquired them from their hosts to further evolve them for their own purpose. As more NCLDV genomes are discovered and more viral histones revealed in their genomes and metagenomes, we can hope to expand upon the potential role of viral histones in eukaryogenesis and the organization of chromatin by giant viruses[42].

5.2 Future Directions

While the role of essential Melbournevirus histones in viral infection has preliminarily been explored through localization during infection, Medusavirus histones have yet to be investigated *in vivo*. For Melbournevirus, although we know that virally encoded histone rapidly accumulates in the cytoplasmic viral factory, it is not known when, where, and how they are deposited onto viral

DNA. Does this happen prior to capsid assembly or is the DNA transported in its histone-free form, with the histones waiting in the capsid? If that is the case, how does assembly function in the very close confines of the capsid? How does the virus know how much histone is needed? For Medusavirus, even less is known as no viral strains with tagged histones are available. How are they transported into the nucleus, how are they assembled, and how (if at all) is the separation between viral and host histones accomplished? Localization experiments as the ones performed by our collaborators for core histones of Melbournevirus would provide insight into the viral function of the acidic Medusavirus H1. However, the genetics and biology of these viral systems are still prohibitively difficult to manipulate.

To address the question of assembly and potential viral-specific histone chaperones or nucleosome assembly factors, a pull-down assay utilizing Melbournevirus or Medusavirus histone specific antibodies could provide insight into what viral histones interact with at different time points during infection. This is hampered by the difficulties in generating viable strains of virus with tagged histones (Melbournevirus seems to resent this greatly), and by the difficulties preparing antibodies that specifically recognize viral over host histones. Lastly, it is not known whether any of these viral histones can be regulated through post-translational modifications. Mass spectrometry analysis of viral histones extracted from viral particles could provide insight into modification events to regulate the genome. However, the process of removing intact chromatin from virions has proven difficult but not impossible (although rather invasive techniques have to be used). As such, analysis of chromatin in the virion will likely come with many challenges[109].

Bibliography

1. Athira, M. T. & Antony, S. P. The Tiny Giants: Overview of Giant Viruses. *Ecol. Genet. Genomics* **29**, 100210 (2023).
2. Scola, B. L. *et al.* A Giant Virus in Amoebae. *Science* **299**, 2033–2033 (2003).
3. *Lesser Known Large dsDNA Viruses*. vol. 328 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009).
4. Rampersad, S. & Tennant, P. Replication and Expression Strategies of Viruses. in *Viruses* 55–82 (Elsevier, 2018). doi:10.1016/B978-0-12-811257-1.00003-6.
5. Colson, P., De Lamballerie, X., Fournous, G. & Raoult, D. Reclassification of Giant Viruses Composing a Fourth Domain of Life in the New Order ***Megavirales***. *Intervirology* **55**, 321–332 (2012).
6. Bell, P. J. L. Eukaryogenesis: The Rise of an Emergent Superorganism. *Front. Microbiol.* **13**, 858064 (2022).
7. Schulz, F., Abergel, C. & Woyke, T. Giant virus biology and diversity in the era of genome-resolved metagenomics. *Nat. Rev. Microbiol.* **20**, 721–736 (2022).
8. Zhang, R., Takemura, M., Murata, K. & Ogata, H. "Mamonoviridae", a proposed new family of the phylum Nucleocytoviricota. *Arch. Virol.* **168**, 80 (2023).
9. Koonin, E. V. & Yutin, N. Evolution of the Large Nucleocytoplasmic DNA Viruses of Eukaryotes and Convergent Origins of Viral Gigantism. in *Advances in Virus Research* vol. 103 167–202 (Elsevier, 2019).

10. Mönttinen, H. A. M., Bicep, C., Williams, T. A. & Hirt, R. P. The genomes of nucleocytoplasmic large DNA viruses: viral evolution writ large. *Microb. Genomics*.
11. Subramaniam, K. *et al.* A New Family of DNA Viruses Causing Disease in Crustaceans from Diverse Aquatic Biomes. *mBio* **11**, e02938-19 (2020).
12. Andreani, J. *et al.* Cedratvirus, a Double-Cork Structured Giant Virus, is a Distant Relative of Pithoviruses. *Viruses* **8**, 300 (2016).
13. Needham, D. M. *et al.* A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proc. Natl. Acad. Sci.* **116**, 20574–20583 (2019).
14. Clouthier, S. *et al.* A new species of nucleo-cytoplasmic large DNA virus (NCLDV) associated with mortalities in Manitoba lake sturgeon Acipenser fulvescens. *Dis. Aquat. Organ.* **102**, 195–209 (2013).
15. Andreani, J. *et al.* Pacmanvirus, a New Giant Icosahedral Virus at the Crossroads between Asfarviridae and Faustoviruses. *J. Virol.* **91**, e00212-17 (2017).
16. Abrahão, J. *et al.* Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nat. Commun.* **9**, 749 (2018).
17. Rolland, C. *et al.* Clandestinovirus: A Giant Virus With Chromatin Proteins and a Potential to Manipulate the Cell Cycle of Its Host Vermamoeba vermiformis. *Front. Microbiol.* **12**, 715608 (2021).
18. Bajrai, L. H. *et al.* Isolation of Yasminevirus, the First Member of Klosneuvirinae Isolated in Coculture with Vermamoeba vermiformis, Demonstrates an Extended Arsenal of Translational Apparatus Components. *J. Virol.* **94**, e01534-19 (2019).
19. Andreani, J. *et al.* Orpheovirus IHUMI-LCC2: A New Virus among the Giant Viruses. *Front. Microbiol.* **8**, 2643 (2018).
20. Colson, P., La Scola, B. & Raoult, D. Giant Viruses of Amoebae: A Journey Through Innovative Research and Paradigm Changes. *Annu. Rev. Virol.* **4**, 61–85 (2017).

21. Yoshikawa, G. *et al.* Medusavirus, a Novel Large DNA Virus Discovered from Hot Spring Water. *J. Virol.* **93**, e02130-18 (2019).

22. Brandes, N. & Linial, M. Giant Viruses—Big Surprises. *Viruses* **11**, 404 (2019).

23. Talbert, P. B., Armache, K.-J. & Henikoff, S. Viral histones: pickpocket's prize or primordial progenitor? *Epigenetics Chromatin* **15**, 21 (2022).

24. Smith, M. M. Histone structure and function. *Curr. Opin. Cell Biol.* **3**, 429–437 (1991).

25. Luger, K. Crystal structure of the nucleosome core particle at 2.8 A˚resolution. **389**, (1997).

26. McGinty, R. K. & Tan, S. Nucleosome Structure and Function. *Chem. Rev.* **115**, 2255–2273 (2015).

27. Phillips, E. O. N. & Gunjan, A. Histone variants: The unsung guardians of the genome. *DNA Repair* **112**, 103301 (2022).

28. Tsunaka, Y., Furukawa, A. & Nishimura, Y. Histone tail network and modulation in a nucleosome. *Curr. Opin. Struct. Biol.* **75**, 102436 (2022).

29. Peng, Y., Li, S., Onufriev, A., Landsman, D. & Panchenko, A. R. Binding of regulatory proteins to nucleosomes is modulated by dynamic histone tails. *Nat. Commun.* **12**, 5280 (2021).

30. Burgess, R. J. & Zhang, Z. Histone chaperones in nucleosome assembly and human disease. *Nat. Struct. Mol. Biol.* **20**, 14–22 (2013).

31. Hammond, C. M., Strømme, C. B., Huang, H., Patel, D. J. & Groth, A. Histone chaperone networks shaping chromatin function. *Nat. Rev. Mol. Cell Biol.* **18**, 141–158 (2017).

32. Tyagi, M., Imam, N., Verma, K. & Patel, A. K. Chromatin remodelers: We are the drivers!! *Nucleus* **7**, 388–404 (2016).

33. *Histone Mutations and Cancer*. vol. 1283 (Springer Singapore, Singapore, 2021).

34. White, A. E., Hieb, A. R. & Luger, K. A quantitative investigation of linker histone interactions with nucleosomes and chromatin. *Sci. Rep.* **6**, 19122 (2016).

35. Kalashnikova, A. A., Rogge, R. A. & Hansen, J. C. Linker histone H1 and protein–protein interactions. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1859**, 455–461 (2016).

36. Talbert, P. B. & Henikoff, S. Histone variants at a glance. *J. Cell Sci.* **134**, jcs244749 (2021).

37. Gad, W. & Kim, Y. A viral histone H4 encoded by Cotesia plutellae bracovirus inhibits haemocyte-spreading behaviour of the diamondback moth, Plutella xylostella. *J. Gen. Virol.* **89**, 931–938 (2008).

38. Kim, Y. & Kumar, S. Persistent expression of Cotesia plutellae bracovirus genes in parasitized host, Plutella xylostella. *PLOS ONE* **13**, e0200663 (2018).

39. Boyer, M. *et al.* Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl. Acad. Sci.* **106**, 21848–21853 (2009).

40. Rodrigues, R. A. L. *et al.* Analysis of a Marseillevirus Transcriptome Reveals Temporal Gene Expression Profile and Host Transcriptional Shift. *Front. Microbiol.* **11**, 651 (2020).

41. Talbert, P. B., Henikoff, S. & Armache, K.-J. Giant variations in giant virus genome packaging. *Trends Biochem. Sci.* **48**, 1071–1082 (2023).

42. Irwin, N. A. T. & Richards, T. A. *Self-Assembling Viral Histones Unravel Early Nucleosome Evolution*. http://biorxiv.org/lookup/doi/10.1101/2023.09.20.558576 (2023) doi:10.1101/2023.09.20.558576.

43. Rivera, M. C. & Lake, J. A. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431**, 152–155 (2004).

44. Livingstone Bell, P. J. Viral Eukaryogenesis: Was the Ancestor of the Nucleus a Complex DNA Virus? *J. Mol. Evol.* **53**, 251–256 (2001).

45. Talbert, P. B., Meers, M. P. & Henikoff, S. Old cogs, new tricks: the evolution of gene expression in a chromatin context. *Nat. Rev. Genet.* **20**, 283–297 (2019).

46. Bell, P. J. L. Evidence supporting a viral origin of the eukaryotic nucleus. *Virus Res.* **289**, 198168 (2020).

47. Takemura, M. Medusavirus Ancestor in a Proto-Eukaryotic Cell: Updating the Hypothesis for the Viral Origin of the Nucleus. *Front. Microbiol.* **11**, 571831 (2020).

48. Valencia-Sánchez, M. I. *et al.* The structure of a virus-encoded nucleosome. *Nat. Struct. Mol. Biol.* **28**, 413–417 (2021).

49. Thomas, V. *et al.* Lausannevirus, a giant amoebal virus encoding histone doublets. *Environ. Microbiol.* **13**, 1454–1466 (2011).

50. Erives, A. J. Phylogenetic analysis of the core histone doublet and DNA topo II genes of Marseilleviridae: evidence of proto-eukaryotic provenance. *Epigenetics Chromatin* **10**, 55 (2017).

51. Watanabe, R., Song, C., Kayama, Y., Takemura, M. & Murata, K. Particle Morphology of Medusavirus Inside and Outside the Cells Reveals a New Maturation Process of Giant Viruses. *J. Virol.* **96**, e01853-21 (2022).

52. Luger, K. & Richmond, T. J. The histone tails of the nucleosome. *Curr. Opin. Genet. Dev.* **8**, 140–146 (1998).

53. Bowerman, S., Wereszczynski, J. & Luger, K. Archaeal chromatin 'slinkies' are inherently dynamic complexes with deflected DNA wrapping pathways. *eLife* **10**, e65587 (2021).

54. Mattiroli, F. *et al.* Structure of histone-based chromatin in Archaea. *Science* **357**, 609–612 (2017).

55. Oh, J. *et al.* Genome Wide Nucleosome Mapping for HSV-1 Shows Nucleosomes Are Deposited at Preferred Positions during Lytic Infection. *PLOS ONE* **10**, e0117471 (2015).

56. Lieberman, P. M. Chromatin organization and virus gene expression. *J. Cell. Physiol.* **216**, 295–302 (2008).

57. Polisky, B. & McCarthy, B. Location of histones on simian virus 40 DNA. *Proc. Natl. Acad. Sci.* **72**, 2895–2899 (1975).

58. Fabre, E. *et al.* Noumeavirus replication relies on a transient remote control of the host nucleus. *Nat. Commun.* **8**, 15087 (2017).

59. Okamoto, K. *et al.* Cryo-EM structure of a Marseilleviridae virus particle reveals a large internal microassembly. *Virology* **516**, 239–245 (2018).

60. Dyer, P. N. *et al.* Reconstitution of Nucleosome Core Particles from Recombinant Histones and DNA. in *Methods in Enzymology* vol. 375 23–44 (Elsevier, 2004).

61. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).

62. Kastner, B. *et al.* GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008).

63. Chen, Y. *et al.* Asymmetric unwrapping of nucleosomal DNA propagates asymmetric opening and dissociation of the histone core. *Proc. Natl. Acad. Sci.* **114**, 334–339 (2017).

64. Chua, E. Y. D., Vasudevan, D., Davey, G. E., Wu, B. & Davey, C. A. The mechanics behind DNA sequence-dependent properties of the nucleosome. *Nucleic Acids Res.* **40**, 6338–6352 (2012).

65. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).

66. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinforma.* **54**, (2016).

67. Peng, Y., Markov, Y., Goncearenco, A., Landsman, D. & Panchenko, A. R. Human Histone Interaction Networks: An Old Concept, New Trends. *J. Mol. Biol.* **433**, 166684 (2021).

68. Dorigo, B., Schalch, T., Bystricky, K. & Richmond, T. J. Chromatin Fiber Folding: Requirement for the Histone H4 N-terminal Tail. *J. Mol. Biol.* **327**, 85–96 (2003).

69. Morrison, E. A., Bowerman, S., Sylvers, K. L., Wereszczynski, J. & Musselman, C. A. The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome. *eLife* **7**, e31481 (2018).

70. Hodges, A. J. *et al.* Histone Sprocket Arginine Residues Are Important for Gene Expression, DNA Repair, and Cell Viability in *Saccharomyces cerevisiae*. *Genetics* **200**, 795–806 (2015).

71. Pentakota, S. *et al.* Decoding the centromeric nucleosome through CENP-N. *eLife* **6**, e33442 (2017).

72. Zhou, M. *et al.* Structural basis of nucleosome dynamics modulation by histone variants H2A.B and H2A.Z.2.2. *EMBO J.* **40**, e105907 (2021).

73. Gowripalan, A., Smith, S., Stefanovic, T. & Tscharke, D. C. Rapid poxvirus engineering using CRISPR/Cas9 as a selection tool. *Commun. Biol.* **3**, 643 (2020).

74. Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018).

75. Bateman, E. Expression plasmids and production of EGFP in stably transfected Acanthamoeba. *Protein Expr. Purif.* **70**, 95–100 (2010).

76. Blanca, L., Christo-Foroux, E., Rigou, S. & Legendre, M. Comparative Analysis of the Circular and Highly Asymmetrical Marseilleviridae Genomes. *Viruses* **12**, 1270 (2020).

77. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

78. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. (2011).

79. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

80. Demeler, B. *et al.* Characterization of Size, Anisotropy, and Density Heterogeneity of Nanoparticles by Sedimentation Velocity. *Anal. Chem.* **86**, 7688–7695 (2014).

81. Gorbet, G. *et al.* A Parametrically Constrained Optimization Method for Fitting Sedimentation Velocity Experiments. *Biophys. J.* **106**, 1741–1750 (2014).

82. Brookes, E. H., Demeler, B. & Rocco, M. Developments in the US-SOMO Bead ModelingSuite: New Features in the Direct Residue-to-Bead Method, Improved Grid Routines, andInfluence of Accessible Surface Area Screening. *Macromol. Biosci.* **10**, 746–753 (2010).

83. Edwards, G. B., Muthurajan, U. M., Bowerman, S. & Luger, K. Analytical Ultracentrifugation (AUC): An Overview of the Application of Fluorescence and Absorbance AUC to the Study of Biological Macromolecules. *Curr. Protoc. Mol. Biol.* **133**, e131 (2020).

84. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

85. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

86. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

87. Emsley, P. & Cowtan, K. *Coot* : model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

88. Bienert, S. *et al.* The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Res.* **45**, D313–D319 (2017).

89. Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* **30**, S162–S173 (2009).

90. Sali, A. & Blundell, T. L. Comparative Protein Modeling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **234**, 779–815 (1993).

91. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).

92. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

93. Adams, P. D. *et al. PHENIX* : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

94. McGreevy, R., Teo, I., Singharoy, A. & Schulten, K. Advances in the molecular dynamics flexible fitting method for cryo-EM modeling. *Methods* **100**, 50–60 (2016).

95. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

96. Chan, K.-Y., Trabuco, L. G., Schreiner, E. & Schluten, K. Cryo-electron microscopy modeling by the molecular dynamics flexible fitting method. *Biopolymers* **97**, 678–686 (2012).

97. Huang, J. & Mackerell Jr, A. D. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–45 (2013).

98. Vasudevan, D., Chua, E. Y. D. & Davey, C. A. Crystal Structures of Nucleosome Core Particles Containing the '601' Strong Positioning Sequence. *J. Mol. Biol.* **403**, 1–10 (2010).

99. Malik, H. S. & Henikoff, S. Phylogenomics of the nucleosome. *Nat. Struct. Mol. Biol.* **10**, 882–891 (2003).

100. Luger, K. & Richmond, T. J. DNA binding within the nucleosome core. *Curr. Opin. Struct. Biol.* **8**, 33–40 (1998).

101. Zhou, K., Gaullier, G. & Luger, K. Nucleosome structure and dynamics are coming of age. *Nat. Struct. Mol. Biol.* **26**, 3–13 (2019).

102. Koonin, E. V. Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20140333 (2015).

103. Wolffe, A. P. & Pruss, D. Chromatin: Hanging on to histones. *Curr. Biol.* **6**, 234–237 (1996).

104. Patwal, I., Trinh, H., Golden, A. & Flaus, A. *Histone Sequence Variation in Divergent Eukaryotes Facilitates Diversity in Chromatin Packaging*. http://biorxiv.org/lookup/doi/10.1101/2021.05.12.443918 (2021) doi:10.1101/2021.05.12.443918.

105. Da Cunha, V., Gaia, M., Nasir, A. & Forterre, P. Asgard archaea do not close the debate about the universal tree of life topology. *PLOS Genet.* **14**, 1–5 (2018).

106. Villarreal, L. P. Evolution of Viruses. in *Encyclopedia of Virology* 174–184 (Elsevier, 2008). doi:10.1016/B978-012374410-4.00706-8.

107. Koonin, E. V. & Yutin, N. Origin and Evolution of Eukaryotic Large Nucleo-Cytoplasmic DNA Viruses. *Intervirology* **53**, 284–292 (2010).

108. Liu, Y. *et al.* Virus-encoded histone doublets are essential and form nucleosome-like structures. *Cell* **184**, 4237-4250.e19 (2021).

109. Bryson, T. D. *et al.* A giant virus genome is densely packaged by stable nucleosomes within virions. *Mol. Cell* **82**, 4458-4470.e5 (2022).

110. Zhang, R., Endo, H., Takemura, M. & Ogata, H. RNA Sequencing of Medusavirus Suggests Remodeling of the Host Nuclear Environment at an Early Infection Stage. *Microbiol. Spectr.* **9**, e00064-21 (2021).

111. Kenzaki, H. & Takada, S. Linker DNA Length is a Key to Tri-nucleosome Folding. *J. Mol. Biol.* **433**, 166792 (2021).

112. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).

113. Taguchi, H., Horikoshi, N., Arimura, Y. & Kurumizaka, H. A method for evaluating nucleosome stability with a protein-binding fluorescent dye. *Methods* **70**, 119–126 (2014).

114. Muthurajan, U. *et al.* In Vitro Chromatin Assembly. in *Methods in Enzymology* vol. 573 3–41 (Elsevier, 2016).

115. Harp, J. M., Hanson, B. L., Timm, D. E. & Bunick, G. J. Asymmetries in the nucleosome core particle at 2.5 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 1513–1534 (2000).

116. Zhou, B.-R. *et al.* Revisit of Reconstituted 30-nm Nucleosome Arrays Reveals an Ensemble of Dynamic Structures. *J. Mol. Biol.* **430**, 3093–3110 (2018).

117. DuPai, C. D., Davies, B. W. & Wilke, C. O. A systematic analysis of the beta hairpin motif in the Protein Data Bank. *Protein Sci.* **30**, 613–623 (2021).

118. Skrajna, A. *et al.* Comprehensive nucleosome interactome screen establishes fundamental principles of nucleosome binding. *Nucleic Acids Res.* **48**, 9415–9432 (2020).

119. Attar, N. *et al.* The histone H3-H4 tetramer is a copper reductase enzyme. *Science* **369**, 59–64 (2020).

120. Cirillo, L. A. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J.* **17**, 244–254 (1998).

121. Barbero, J. L., Franco, L., Montero, F. & Moran, F. Structural studies on histones H1. Circular dichroism and difference spectroscopy of the histones H1 and their trypsin-resistant cores from calf thymus and from the fruit fly Ceratitis capitata. *Biochemistry* **19**, 4080–4087 (1980).

122. Höllmüller, E. *et al.* Interactome of Site-Specifically Acetylated Linker Histone H1. *J. Proteome Res.* **20**, 4443–4451 (2021).

123. Dombrowski, M., Engeholm, M., Dienemann, C., Dodonova, S. & Cramer, P. Histone H1 binding to nucleosome arrays depends on linker DNA length and trajectory. *Nat. Struct. Mol. Biol.* **29**, 493–501 (2022).

124. Hocher, A. *et al.* Histones with an unconventional DNA-binding mode in vitro are major chromatin constituents in the bacterium Bdellovibrio bacteriovorus. *Nat. Microbiol.* **8**, 2006–2019 (2023).

125. Fukaya, S. & Takemura, M. Kinetic Analysis of Acanthamoeba castellanii Infected with Giant Viruses Quantitatively Revealed Process of Morphological and Behavioral Changes in Host Cells. *Microbiol. Spectr.* **9**, e00368-21 (2021).

126. Stothard, P. The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences. *BioTechniques* **28**, 1102–1104 (2000).

127. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Preparation of nucleosome core particle from recombinant histones. in *Methods in Enzymology* vol. 304 3–19 (Elsevier, 1999).

128. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247–260 (2006).

129. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

130. Bepler, T., Kelley, K., Noble, A. J. & Berger, B. Topaz-Denoise: general deep denoising models for cryoEM and cryoET.

131. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

132. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).

133. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

134. Meng, E. C. *et al.* UCSF ChimeraX : Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).

135. Pettersen, E. F. *et al.* UCSF ChimeraX : Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

136. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

137. Nečas, D. & Klapetek, P. Gwyddion: an open-source software for SPM data analysis. *Open Phys.* **10**, (2012).

138. Miles, A. J., Drew, E. D. & Wallace, B. A. DichroIDP: a method for analyses of intrinsically disordered proteins using circular dichroism spectroscopy. *Commun. Biol.* **6**, 823 (2023).

139. Philippe, N. *et al.* Pandoraviruses: Amoeba Viruses with Genomes Up to 2.5 Mb Reaching That of Parasitic Eukaryotes. *Science* **341**, 281–286 (2013).

140. Legendre, M. *et al.* In-depth study of *Mollivirus sibericum* , a new 30,000-y-old giant virus infecting *Acanthamoeba*. *Proc. Natl. Acad. Sci.* **112**, (2015).

141. Banks, D. D. & Gloss, L. M. Folding mechanism of the $(H3–H4)_2$ histone tetramer of the core nucleosome. *Protein Sci.* **13**, 1304–1316 (2004).