Investigation of the Unusual Histone-Like Doublets of the *Melbournevirus* Host Organism *Acanthamoeba castellanii*

by

Aubrey Wolfe

Undergraduate Honors Thesis

Department of Biochemistry

University of Colorado Boulder

March 28th, 2022

Thesis Committee

Advisor: Dr. Karolin Luger, BCHM

Dr. Jeffrey Cameron, BCHM

Dr. Jennifer Martin, MCDB

Dr. Jennifer Knight, MCDB

Portions within the introduction section are to be published in a review article currently being written with the goal of submission for review spring 2022. My graduate student mentor within the Luger lab is first author, and I will be second author. Portions shared between my thesis and the review article are my own writing; analysis of data within Table 1 is the contribution of both myself and my graduate student mentor, Chelsea Toner.

Abstract

The organization of DNA into nucleosomes with histone proteins has been long been thought to be an exclusive feature of eukaryotes, and more recently archaea. The breadth of histone-encoding entities has been expanded by the discovery of histone-like sequences within the genomes of Nucleocytoplasmic Large DNA Viruses, or giant viruses. Recent studies of the amoebae-infecting giant virus, *Marseillevirus,* demonstrate the formation of nucleosome-like particles with its "fused" histone doublets. Intriguingly, unpublished data has revealed the expression of similarly fused histone-like doublets within the amoeba host organism, *Acanthamoeba castellanii.* The similarity of these histone-like doublets to *Marseillevirus* doublets, and to canonical eukaryotic histones, was unknown. Whether or not the amoeba histone-like doublets maintain the same functional capacity to form nucleosome-like particles is unclear. Multiple sequence alignment and secondary structure prediction demonstrate the characteristic histone structure of amoeba doublets despite a divergence in sequence identity between *Marseillevirus* and eukaryotic histones. Homology models constructed from eukaryotic histones further demonstrate the ability of the amoeba doublets to fit within an *in-silico* eukaryotic nucleosome particle. These results support the idea that the histone-like doublets of *Acanthamoeba castellanii* have the potential to form nucleosome-like particles *in vitro* and further enlarges the realm of histone diversity, both in structure and function.

Introduction

The year 2003 marked the discovery of what are now termed giant viruses, or Nucleocytoplasmic Large DNA Viruses (NCLDV) which have since transformed our previous understanding of viral size and complexity, and arguably, of our entire conceptualization of viruses. NCLDV are a class of double-stranded DNA viruses, divergent from classical viruses via their life cycles, and large capsid and genome sizes[1]. NCLDV genomes range from 100 kilobase pairs (kbp) to greater than 2.5 megabase pairs (Mbp), lying well within the range of some bacteria which harbor genomes anywhere from 130 kbp to 14 Mbp[1,2]. For comparison, Herpes Simplex Viruses contain some of the largest viral genomes apart from NCLDV, yet they are no larger than 230 kbp[3]. Additionally, a prominent distinction between NCLDV and traditional viruses is their replication cycles that can include both cytoplasmic and nuclear phases, where traditional viruses undergo replication processes in either the nucleus or the cytoplasm[2]. The International Committee on Taxonomy of Viruses (ICTV) currently classifies NCLDV into seven families: *Ascoviridae*, *Asfarviridae*, *Iridoviridae*, *Marseilleviridae*, *Mimiviridae*, *Phycodnaviridae*, and *Poxviridae,* while 21 proposed NCLDV remain unclassified[4]. Various phylogenetic analyses have informally assigned these unclassified NCLDV to additional proposed families[4,5]. 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 within virus replication and virion morphogenesis[6]. As the genomic diversity of NCLDV parallels the continued expansion of

Table 1. Representation of predicted histone sequences in both unclassified and classified Giant Virus families.

Virus Family	Encoded Histones	Genome Size (kb)	Replication Location	Host Taxonomy (Range)
Clandestinovirus (unclassified)	H3 H4 Doublet H2B-H2A H1/H5	582	Nucleus	*Vermamoeba vermiformis*
Marseille virus (classified)	H2B-H2A H4-H3	347-403	Nucleus cytoplasm	Acanthamoeba; probably, also algae
Medusavirus (unclassified)	H2A H1 H2B H3 H4	381	Nucleus	Amoeba
Mollivirus (unclassified)	H2A H2B H3 H4	652	Nucleus cytoplasm	Amoeba
Phycodnaviridae (classified)	DSLPV1 demonstrated H3 variant	155-474	Nucleus cytoplasm	Alveolata, Chlorophyta, Haptophyta, Stramenopiles

their class, the number of conserved genes amongst them dwindles. However, phylogenies have been constructed around 40 nearly universal genes. Of these genes, three are conserved amongst all NCLDV, namely major capsid protein (MCP), DNA polymerase, and Poxvirus Late Transcription Factor VLTF3[7]. The *Nucleocytoviricota* phylum was established by the ICTV based on the double jelly-roll major capsid proteins (DJR-MCPs)[4] and the main features of the phylogeny have remained remarkably stable over several decades, suggesting that the major branches of the NCLDV are already firmly defined.

Apart from the set of conserved genes, NCLDV genomes vary widely between families. Yet the enzymatic repertoires encoding a suite of proteins for replication, transcription, and translation is perhaps their greatest unifying feature, separating them from other double-stranded and single-stranded DNA viruses[6]. Within the last decade, the discovery of several NCLDV encoding histone-like proteins has contributed to our growing understanding of the mysterious nature of these viruses while further expanding the realm of histone-encoding entities within the biosphere. Currently, of the proposed and classified NCLDV, only five giant virus families encode various forms of putative histones (Table 1).

The Luger lab has focused on one NCLDV belonging to the *Marseilleviridae* (MSRV) family, *Melbournevirus* (MV), which encodes putative homologues of the four eukaryotic histones (H3, H4, H2A, and H2B) in the form of fused "doublets"[8]. That is, histone H2B is fused to H2A, and H4 is fused to H3 (Figure 1). Viral histone doublets exist in mature *Marseillevirus* virions wherein they are expected to participate in genome organization[9–11]. Work in the Luger lab has demonstrated the ability of these histone-like proteins to form nucleosome-like particles (NLPs) *in vitro*. Gene knock-out experiments from our collaborator, Dr. Chantal Abergel, showed that both H2B-H2A and H4-H3 doublets are essential for viral fitness[8]. In conjunction with the fact that MSRV fused histone doublets share less than 30% sequence identity with eukaryotic histones and are highly conserved within the MSRV family, it is proposed that the histone doublets have evolved to fulfill roles other than (or in addition to) DNA compaction[8] . A question that remained unanswered is how the viral histones are transported and targeted to viral DNA within MV cytoplasmic

viral factories (VF; foci in the amoeba cytoplasm, where viral transcription, replication, and assembly take place) during infection of its host, *Acanthamoeba castellanii (A. castellanii)*. To further understand the fate of viral histones upon infection, I investigated the viral life cycle, viral proteomic potential, and viral-host interactions through surveying published peer-reviewed literature surrounding NCLDV.

Current quandaries around *Marseilleviridae* histones have been compounded by the discovery that the amoeba host system*, A. castellanii,* encodes histone-like doublets, in addition to its canonical singlet histones. Unpublished data of *A. castellanii* proteomic analysis from Dr. Chantal Abergel demonstrates the expression of fused histone-like doublets of proposed H2B and H2A, along with amoeba singlet histones, during viral infection. These putative amoeba doublets are fused in the same order as H2B-H2A MV doublets, yet they share less than 30% amino acid sequence identity with MV doublets and less than 35% sequence identity with the amoeba singlet histones. Amoebae host systems of NCLDV are not well characterized in general, presumably because they are abundant, lower eukaryotes. The discovery of *A. castellanii* histone-like doublets has birthed the necessity of a deeper understanding of *A. castellanii* and reveals a gap in our knowledge about the evolutionary relationship between NCLDV and their eukaryotic hosts.

Several questions arise surrounding the discovery of fused histone-like doublets within the MV *A. castellanii* host. It is not known whether the proposed histone-like doublets of *A.castellanii* behave as histones, even though they likely have the canonical histone fold motif seen in eukaryotic histones[12]. It was also unknown how similar in sequence, structure, and function these putative histone-like doublets are to the recently established doublet containing nucleosome of the MV[8]. Understanding the similarities and differences between structure and function of the known viral doublet histones and proposed histone-like doublets of the amoeba would aid in determining if the *A. castellanii* histone-like doublets maintain the ability to form NLPs. I assembled a multiple sequence alignment of the proposed histone-like doublets in *A. castellanii* against canonical eukaryotic histones and the established doublet histone of

MV (Figure S1). This highlights the presence of the canonical histone fold, sequence similarities, and conservation of DNA binding sites on the proposed histone-like doublets of *A. castellanii*.

Homology models of amoeba histone-like doublets were developed using the canonical *Xenopus laevis (X. laevis)* histones as a reference. These models were then placed into the context of the canonical eukaryotic nucleosome (PDB 1AOI) to determine their ability for accommodation within the nucleosome while providing an initial insight into the orientation and potential behavior of the putative amoeba doublets within nucleosomal DNA. Amoeba doublet models were also compared to *in vitro* and *in-silico* constructs of MV doublet histone NLPs.

Experimentally probing the ability of putative *A. castellanii* histone-like doublets to form NLPs *in vitro* is a critical step in defining the function of *A. castellanii* histone-like proteins and provides the opportunity for comparison against the established viral nucleosome. Given the general similarities of amoeba and viral doublets, it is expected that amoeba doublets will also form NLPs *in vitro.* Through our analyses, we aim to better understand the purpose of fused histone-doublets within a nucleosome context, while contributing to the current body of knowledge surrounding histones and their unfolding diversity.

Presently, no structure of the canonical amoeba nucleosome exists. To date, we have successfully expressed and purified the four singlet amoeba histones (H3, H4, H2A, and H2B) and nucleosome reconstitution experiments are in progress to address this gap in our knowledge. After successful reconstitution of the canonical amoeba nucleosome, we will begin characterizing this structure through negative-stain electron microscopy (neg-stain) and cryo-electron microscopy (cryo-EM) to determine the structure of the *A. castellanii* nucleosome. Experiments to express, purify, and generate nucleosomes using *A. castellanii* proposed histone-like doublets are also currently in progress.

Results and Discussion

Investigating the Transport of Viral Histones: Exploration of the Current Literature

An analysis of published peer-reviewed literature surrounding NCLDV was performed to gain insight into viral-host interactions, and the various roles of viral proteins within the viral life cycle that may allude to a potential shuttling mechanism of MV histone doublets to VFs. The MSRV life cycle begins with entry into the host cell via endocytosis or phagocytosis, with virion digestion occurring 20 min post-infection (PI) within the phagosomes (vesicles formed through invagination of external virions by the amoeba plasma membrane)[10]. *Marseilleviridae* appear to differ from other Acanthamoeba-infecting viruses through release of the virion core into the cytoplasm through simple openings of the phagosome membrane, following virion digestion. This is opposed to the opening of a specific portal of virion capsids followed by a virus-phagosome membrane fusion and release of the core into the cytoplasm[10,13]. Upon release of the core, all contents of the virion, including viral DNA, viral RNA, and viral proteins are available for interaction with host proteins in the cytoplasm[13,14].

Marseilleviridae also exhibit differential gene expression in conjunction with regulation of host protein expression during infectious cycles[15] . Viral gene expression follows early, intermediate, and late patterns with various host proteins being down- and upregulated. As early as 30 minutes PI, MSRV DNA is transcribed, some of the earliest expressed genes being viral transcription factors. Viral DNA polymerase transcription (at 1-2 h PI) is followed by viral histone transcription that begins at 2-4 h PI, continuing throughout the infection cycle[8,15]. Although *Marseilleviridae* encode a complete transcriptional apparatus, their mature virions lack any transcription proteins, and they rely on host nuclear functions to initiate viral DNA transcription. Once viral transcription factors are expressed, both host and viral transcription factors continue to be utilized[10]. Thus, MSRV represent a unique intermediary family demonstrating both nuclear and cytoplasmic phases.

Recent studies of a new member of the *Marseilleviridae* family have revealed the transient recruitment of Acanthamoeba nuclear proteins to the VF beginning 30 minutes PI[10]. The nucleus structure is maintained during host nuclear protein recruitment, and host nuclear factors return to the nucleus approximately two h PI (excluding host Poly-A polymerase as this is not encoded by *Marseilleviridae)*[10]. This has also been demonstrated for MV infection of *A. castellanii*[8]. The preservation of the nucleus is not demonstrated by other giant viruses reliant on host nuclear factors, though the effect of host nucleus preservation on the overall infectious cycle is unclear[10]. Following the packaging of viral proteins and viral nucleic acid in capsids within the VFs, mature virions collect in intracytoplasmic vacuoles for release via cell lysis or exocytosis[10]. These studies indicate an orchestrated crosstalk and sophisticated regulation between the amoeba host and MSRV during infection.

The necessity for refinement of such viral-host interactions is apparent in MV infection of *A. castellanii. A. castellanii* cells transfected with MV histone doublets show the scattering of MV doublets throughout the entire cell, including the nucleus. Upon infection, the majority of MV doublets localize to VFs where they co-localize with viral DNA[8]. Ultimately, MV doublets are found within the mature viral particles alongside MV DNA. In eukaryotes, histone chaperones facilitate the sequential deposition of histones (first H3-H4, then H2A-H2B) onto DNA in the process of nucleosome assembly[16]. If MV doublets do assemble onto viral DNA, they would likely require a chaperone protein.

With the knowledge of mediated host protein expression during MV infection, Dr. Chantal Abergel performed a time-course transcriptomic analysis of *A. castellanii* during MV-infection to identify a putative host histone chaperone in viral histone transport and localization. Results indicated that cellular histone chaperones are either never expressed or have their expression decreased after 2 h PI, as is demonstrated for most *A. castellanii* genes[8]. In addition, no known cellular histone chaperone homologs were detected in the *Marseilleviridae* genome[8], and a BLAST query of the *Marseillevirus* genome against known eukaryotic histone chaperones found no matches (see Supplementary Table 1 for a list of

chaperones within the query). These findings uncover the potential for an uncharacterized protein, viral or cellular, being specialized for viral histone doublets and their transport. They may also imply that viral histone transport and assembly occur through a mechanism other than a chaperone protein altogether.

Overall, a fundamental theme across current literature is a large gap in our knowledge surrounding the true proteomic potential of NCLDV, with genomes that are mostly uncharacterized. MSRV is not excluded from this. Transcriptomic analysis of the prototype member of MSRV (*Marseillevirus T19)*, revealed a total of 457 genes within the 368 kb genome. Of these genes, 69.1% have uncharacterized function, and only four of the 20 most expressed proteins have known function, those being the multiple zinc finger protein, the major capsid protein, histone H3, and zinc finger protein[15]. Similar analysis of MV protein expression reveals four proteins of known function within the 10 most expressed proteins: the H4-H3 histone doublet, major capsid protein, H2B-H2A histone doublet, and a cysteine protease[8]. The abundance of hypothetical (uncharacterized) proteins within the MV genome may actually provide the hidden link to a potential shuttling mechanism of its histone doublets.

As viral-host interactions are obscure, and there is no evidence of host chaperone involvement in viral histone transport, it is plausible that MV has acquired or evolved a unique transport protein in addition to its unique histone doublets. The expanding network of eukaryotic histone chaperones are not well understood due to their structural and functional diversity. Eukaryotic histone chaperones are specialized for histone variants, and this is accomplished through a high degree of structural specificity[16]. It is reasonable to speculate that a viral chaperone protein would exhibit the same degree of specialization, thereby preventing detection of any homologs within the eukaryotic system. Further investigation of these possibilities will help in understanding the potential viral-host interactions behind the mechanism of viral histone shuttling and the ever-evolving intelligence behind viral fitness.

Multiple Sequence Alignment and Secondary Structure Prediction

A. castellanii encodes four putative histone doublets containing H2B and H2A domains that are expressed in addition to the canonical amoeba singlet histones, H2A, H2B, H3, and H4, during viral infection as demonstrated by unpublished data from our collaborator, Dr. Chantal Abergel. The putative doublets are denoted here as Amoeba_H2BH2A_3139 (Am3139), Amoeba_H2BH2A_3162 (Am3162), Amoeba_H2BH2A_0031 (Am0031), and Amoeba_H2BH2A_8385 (Am8385). H2B-H2A histone dimer pairs within Eukarya were aligned against the doublet *A. castellanii* and *Marseilleviridae* histones using HHpred's multiple sequence alignment tool (ClustalΩ). Secondary structure prediction suggests that the proposed *A.castellanii* doublets form histone folds (α1-L1-α2-L2-α3) in each of the putative H2B and H2A domains[12] (Figure 1). Predicted histone fold motifs are a fundamental indication that these proteins are histones, therefore potentially performing the same DNA-packaging function as canonical eukaryotic histones within amoeba.

An immediate observation of sequence alignment (Figure S1) reveals a high degree of amino acid sequence similarity between two of the proposed amoeba doublets (Am3139 and Am3162) while the remaining two diverge significantly (Am0031 and Am8385). Indeed, Am3139 and Am3162 share ~99% sequence similarity and ~98% sequence identity, where they both share only ~40% sequence identity to Am0031 and ~20% sequence identity to Am8385.

As mentioned, MSRV doublets share less than 30% amino acid sequence identity with eukaryotic histones (*X. laevis)*[8]. Intriguingly, proposed *A. castellanii* doublets deviate from both viral doublet and amoeba singlet histones. All four amoeba doublets share ≤ ~30% amino acid sequence identity to MSRV doublets, and Am0031, Am3162, Am3139, and Am8385 have decreasing sequence identity to amoeba singlet histones (H2B and H2A) ranging from ~35% to ~18% (in that order). The sequence identity between amoeba doublets Am3162, Am3139, and Am8385 and *X. laevis* histones H2B and H2A remains in the same

range of ~37 to ~24%. However, Am0031 shares the highest degree of sequence identity with *X. laevis* histones at ~61%. This suggests the possibility of a singular putative histone doublet within *A. castellanii* from which the other putative doublets were derived and also highlights the potential for functional variation between the four putative amoeba doublets. The relatively low sequence identity between the proposed amoeba histone-like doublets, viral doublets, and canonical eukaryotic histones adds another dimension to the evolutionary origin of histones, perhaps signifying an unidentified deviation within well-established eukaryotic histone phylogenetic trees. Although variations in function may parallel variations in histone sequence identities, MV-NLPs are evidence that sequence identity to canonical eukaryotic histones is not the main determinant of the ability of doublet histones to bind DNA and form NLPs.

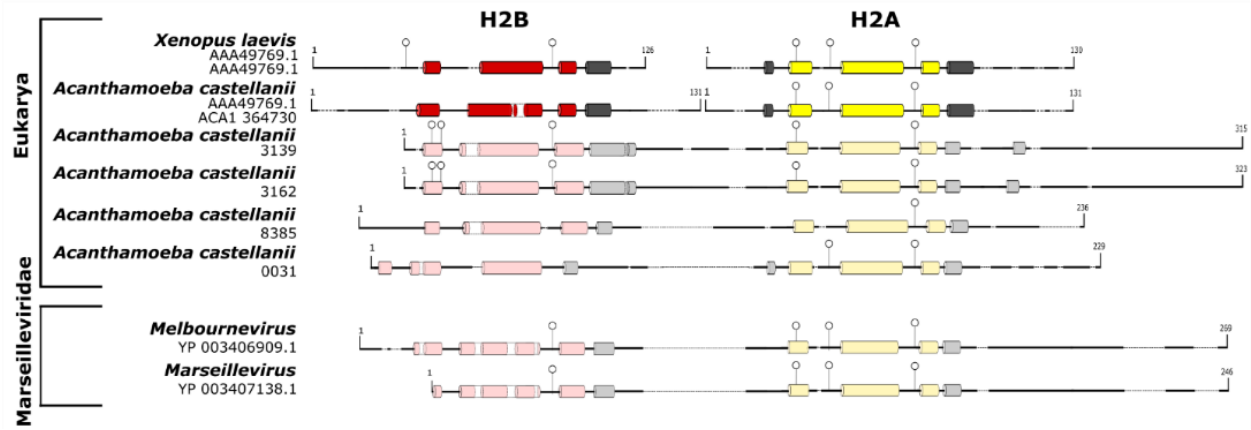


Figure 1. *A. castellanii* putative histone doublets resemble eukaryotic histones

(A) H2B-H2A histone dimer pairs within Eukarya were aligned against the doublet *A. castellanii* and *Marseilleviridae* histones using HHpred's multiple sequence alignment tool (ClustalΩ). Known α helices from the histone fold domain in Eukarya are dark colored tubes (H2B, red; H2A, yellow; and additional helices, gray). "Breaks" within the helices signify an absence of amino acids(s) within the multiple sequence alignment. Predicted α helices in amoeba and MSRV were generated using HHPred's quick 2D prediction web server (shown in lighter coloration) within amoeba and MSRV histone doublets. R-T pairs and conserved DNA-binding residues are marked with lollipops.

Secondary structure prediction also suggests an alpha-helix within the "extended" H2A C-terminal tails of Am3139 and Am3162 that may serve similar structural purposes as the H2A docking domain of canonical eukaryotic histones. The H2A docking domain tethers the H2A-H2B dimer to the H3-H4 tetramer in eukaryotic nucleosomes and represents a critical component of the stability. MV, Am0031, and Am8385 contain longer H2A C-terminal tails than canonical eukaryotic H2A, but these are still shorter than those in Am3139 and Am3162. The H2B C-terminal tails of Am3139 and Am3162 are also longer, and including the proposed H2A domain, their total length exceeds all other histones within the alignment by a minimum of 40 amino acids, surpassing Am0031 and Am8385 each by approximately 100 amino acids.

Further comparison between aligned histones indicates the conservation of hallmark DNA-binding sites within the histone folds. Formation of the NCP is reliant on conserved residues within the histone folds of the four histones H3, H4, H2A, and H2B, exhibiting two main types of binding. The primary type of binding occurs between main-chain amide nitrogen atoms of the histones and phosphates of the DNA backbone. Additional binding occurs between the phosphodiester DNA backbone to specific side chains of conserved residues within the histone fold[17]. Many of the specific amino acids are conserved in the proposed *A. castellanii* doublets and established viral doublets including several 'R-T pairs'. This motif consists of arginine sides chains that reach into the DNA minor groove and interact with a threonine from the paired L1 loop[17]. An R-T pair within H2B is conserved among Am3139, Am3162 and MSRV doublets at position 99 (Figure S1), and another R-T pair is conserved amongst all aligned histones within H2A at position 247. Am3139 and Am3162 maintain an additional R-T pair in H2A at position 212 (Figure S1). Additional potential DNA binding sites are shown in Am3139 and Am3162 at positions 49 and 53. Other basic amino acids within the doublet histones could potentially bind DNA that are not demonstrated through sequence alignment but would need to be explored in the context of the amoeba nucleosome to determine their true DNA-binding capacity.

Homology Modeling of Putative Amoeba Histone Doublets

Initial insight into the ability of *A. castellanii* histone-like doublets to form NLPs was gained *in-silico* through construction of homology models[18] of the four unique doublets identified by Dr. Chantal Abergel. Models were constructed through homology to eukaryotic (*X. laevis)* histones where the ~11-39 amino-acid connectors of H2B and H2A were generated via *de novo* methods[19]. This process utilized SWISS Model to build models of each doublet in which the known structures of *X. laevis* histones H2A and H2B were selected as the template based upon sequence homology to amoeba doublet sequences. *De novo* models of the H2B-H2A linker regions were generated and grafted onto the portion of the models constructed from the template *X. laevis* histones to form complete 3-

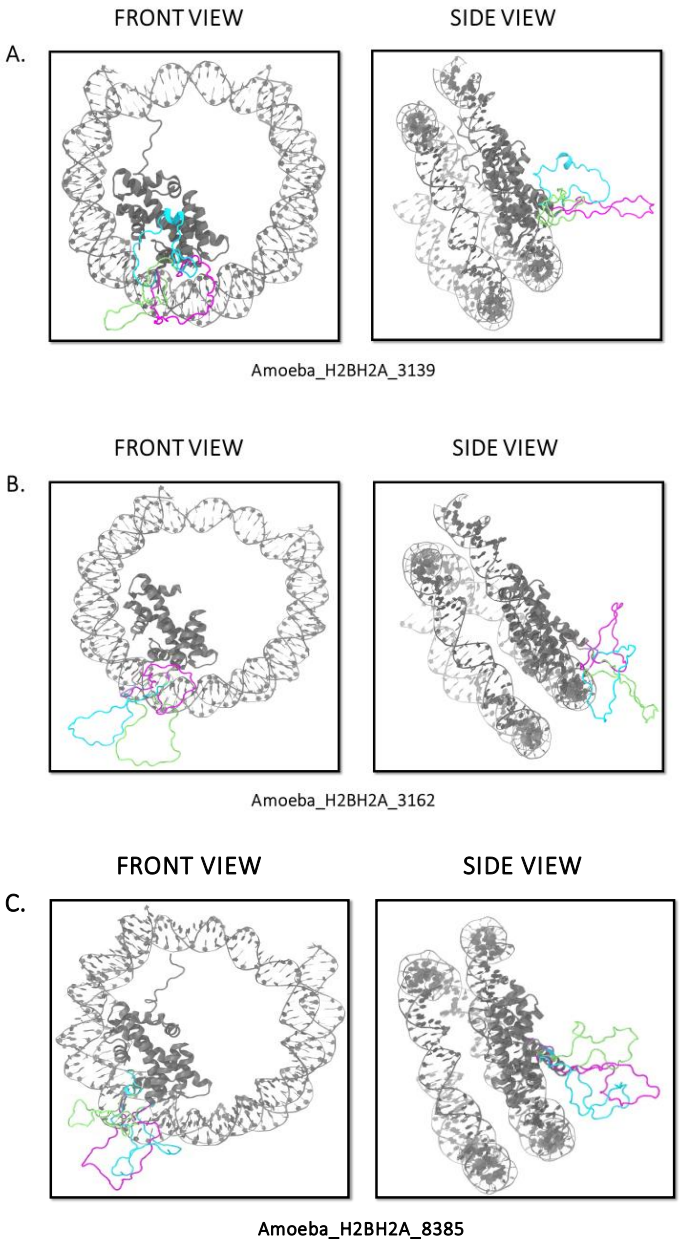


Figure 2. Homology models of *A. castellanii* putative histone doublets

Doublet models illustrate variations in spatial orientations relative to eukaryotic DNA within a canonical nucleosome. (A) Am3139: cyan loop faces the Histone fold; magenta loop orients vertical from the superhelical axis; green loop is directed away from the histone fold, horizontal to the superhelical axis. (B) Am3162: cyan loops sit close to superhelical DNA, horizontal from the axis; magenta loops orient vertical from the superhelical axis; green loops orient horizontal from the axis. (C) Am8385: cyan, magenta, and green loops are all oriented vertical from the superhelical axis.

dimensional doublet models. The main contingency of the ability of the doublets to form an NLP is space accommodation within the core particle. Thus, models were placed into the context of the canonical eukaryotic nucleosome (1AOI) DNA to demonstrate the ability to fit within the space presented by the canonical nucleosome, without steric clashing between the linker region and DNA (Figure 2). The fit of the amoeba doublets was anticipated via comparison to both *in-silico* and *in vitro* MV doublet H2B-H2A in which the linker region is accommodated within the eukaryotic nucleosome and forms an NLP[8]. Importantly, these comparisons do not represent how amoeba doublets would actually bind DNA in a potential amoeba-doublet NLP, and an accurate representation of the behavior of amoeba doublets within an NLP cannot be fully obtained until experimentally testing the binding of amoeba doublets with amoeba DNA *in vitro.* However, the general characteristics of MV-NLPs provide a potential prediction for how amoeba doublets would behave in a nucleosome context due to similarities between viral and amoeba doublets. As such, my results increase confidence that these proposed histone-like doublets can indeed be incorporated into chromatin.

Amoeba doublet models constructed from homology to the canonical *X. laevis* histones H2A-H2B maintain the coordinates of where the canonical histones sit within the eukaryotic nucleosome (PDB 1AOI), and this facilitated their placement within the eukaryotic nucleosome. Modeling software Modeler (v9.20) then generated 10,000 different loops of the linker region connecting the aligned doublets that allowed the accommodation of the doublets within the space of the nucleosome. Modeling software predicted a variety of favorable conformations of the linker region connecting the H2B and H2A domains for three out of the four amoeba doublets after identifying clashing loops using CPPTRAJ of the Amber MD package (v18), where a cutoff distance of 0.8 Å between potentially overlapping atoms was used. Favorable conformations were manually parsed through to further eliminate non-physically relevant loops, such as those intersecting the DNA helix. For three of the four doublets, 3 configurations are shown (Figure 2). Due to limitations of modeling software, non-clashing loops of the linker region of Am0031 with

nucleosomal DNA, or with the proposed H2B and H2A doublet domains were unable to be produced. The Am0031 linker is predicted to be the shortest of the four doublet linkers, at ~11 amino acids, and this may indicate that the connector length (irrespective of its chemical properties) is an important consideration in the amoeba-doublet NLP formation. As of now, it is inconclusive if Am0031 can truly be accommodated within the space of eukaryotic nucleosomal DNA. A sampling size of greater than 10,000 loops would need to be incorporated into the workshopping code to simply increase the number of potential loop configurations.

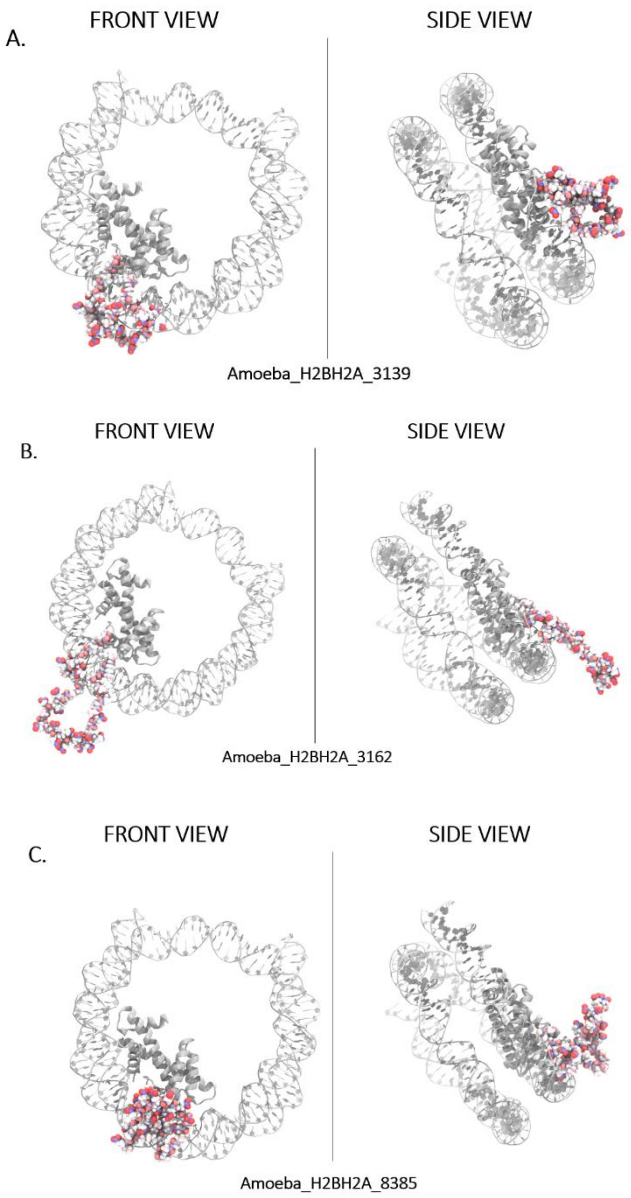


Figure 3. Charge Representation of Modeled *A. castellanii* Putative Histone Doublets

Acidic and basic residues of the H2B-H2A linker region are shown. Red spheres represent negatively charged atoms of charged residues; blue spheres are positively charged atoms of charged residues. (A) Am3139 (B) Am3162 (C) Am8385.

Amongst the three successfully accommodated doublets, three main characteristics of the linker region are demonstrated: non-clashing with the DNA double-helix, various agreeable proximities to the histone folds, and a "free-arm" nature. The first two characteristics are indicative of the histone fold domains being the main structural element constituting formation of the core particle, meaning that the linker region does not interfere with the ability of the H2B and H2A domains to interact with helical DNA and potentially H3,

H4. The third trait of the linkers lying outside of the core particle may offer the potential for interactions with other molecules within the cell.

Canonical eukaryotic histones associate with chaperones and chaperone complexes that are important for their folding, post-translational modification, nuclear import, stability, assembly, and genomic localization[16]. Generally, nucleosomes must be highly dynamic to modulate restriction and accessibility of DNA within the broader scope of regulated gene expression. If the proposed histone-like doublet linkers are not tightly associated with nucleosomal DNA, they may be energetically and spatially accessible to such proteins and protein complexes. The chemical nature, particularly the charge, of the linkers must be considered with such interactions, and it is important to note that histone chaperone mechanisms in this context remain poorly understood. However, this does not exclude the potential for an interaction with another protein through the arm of the linker between the H2B and H2A of the proposed histone doublets of the *A. castellanii*.

Following sequence alignment trends, Am3139 and Am3162 contain linker regions with isoelectric points (pIs) of ~3.5, differing from Am0031 and Am8385 (pI ~8.5). The isoelectric point is the pH at which an amino acid, or protein, has a neutral charge, while fluctuations of this specific pH result in net positive or negative charges. The linker region of Am3139 contains 12 negatively charged amino acids (D and E), 3162 linker contains 16, and Am0031 and Am8385 linkers contain one and four, respectively (Figure 3). The highly negative Am3139 and Am3162 linkers would likely not permit interactions with the phosphodiester backbone of nucleosomal DNA, where this may be possible for Am0031 and Am8385. The difference in the linker region charge between Am3139, Am3162 and Am0031, Am8385 would naturally govern different interactions between classes of histone chaperones, and also between the nucleosome core particle (NCP). Interactions (or lack thereof) of the linker region with DNA and/or the histone fold likely plays a role in overall NCP structure. The significant difference of amoeba doublet linker charges may suggest their evolvement for unique interactions within the cell.

MV histone doublets form defined, yet unstable NLPs *in vitro*[8]. Several features of viral doublets are thought to contribute to the instability of MV-NLPs, including a lower pI of the doublet histone core (octamer of two H2B-H2A and two H4-H3 doublets)[8]. The eukaryotic octamer is comprised of two H2A-H2B dimers and one $(H3\text{-}H4)_2$ tetramer[12] and maintains a pI of ~11, where the MV doublet core is ~9.5[8]. The basic amino acids of canonical histone octamers facilitate a defined pathway for DNA, leading to the idea that a lower pI of the MV-doublet core is partly responsible for a less stable MV-NLP[8]. Proposed Am3139 and Am3162 histone cores (octamer of two H2B-H2A doublets and one amoeba singlet $(H3\text{-}H4)_2$ tetramer) are in a similar range with pIs of 9.94 and 9.87, respectively, and Am0031 and Am8385 proposed cores are more positively charged at pI 10.61 and 11.03, respectively. Charge comparisons of the viral and amoeba doublets supports the idea that the amoeba doublets have the capacity to bind DNA in a nucleosome context and suggests that Am3139 and Am3162 may behave more similarly to MV histones within the context of the nucleosome.

The potential effect of the *A. castellanii* linker length on NLP formation is more difficult to determine. One interpretation from amoeba doublet models is simply that length somehow contributes to NLP formation, as the shorter Am0031 connector region was unable to produce favorable configurations *in-silico*. This does not negate the possibility that Am0031 forms an NLP *in vitro* or *in-vivo.* How length affects the accommodation of amoeba doublet histones within a canonical NCP is uncertain. The fact that the three modeled amoeba doublets exhibit the potential to interact with both helical DNA, histone folds, and intracellular molecules suggests that linker length plays a key role in their spatial distribution which facilitates these potential interactions. The exact parameters of an acceptable linker length are also unclear as the MV H2B-H2A contains a linker of ~20 amino acids[8] while the three modeled amoeba doublets contain linkers that are predicted to be a minimum of 25 residues (range from 25 – 39 residues).

Together, the relative similarity between Am3139 and Am3162 and proposed viral doublets in charge and the accommodation of three of the modeled linkers within canonical eukaryotic nucleosomal DNA support

the idea that amoeba doublets maintain the ability to form NLPs. The differences in charge and length of the individual amoeba doublet linkers perhaps offers further support for the idea that *A. castellanii* doublets were evolved to fulfill distinct and separate functions within the nucleosome context, from both one another and from canonical amoeba singlet histones. These functions may serve as a novel modulator in DNA accessibility and chromatin dynamics.

Acanthamoeba castellanii Histone Expression and Purification

Expression of *A. castellanii* histones was tested in BL21 (DE3), BL21 (DE3) pLysS, and Rosetta(DE3) pLysS *E. coli* cell lines; all histones were present in the insoluble fractions of cell lysates, except for histone H4 which did not appear to express in BL21 (DE3) pLysS cells (Figure 4). Migration of histones through SDS-gels is dictated by their relative molecular weight (MW). Including the 6-histidine tag from the pET15-dTOPO plasmid vector, histone H2A has a MW of ~18 kDa, H2B ~21.5 kDa, H3 ~19 kDa, and H4~16 kDa. Rosetta cells exhibited over-expression of all four proteins and was selected as the best strain for protein expression and stock preparation.

The purification protocol was optimized in multiple iterations, such as an increase of the pH of the nickel-NTA column and TSK-SP columns equilibration and elution buffers from 5.2 to 7.5, allowing for improved binding of the histones to the column. The addition of DNase enzymes to the cell lysates has helped to eliminate DNA contamination, and the implementation of a TSK-SP column increased overall purity to improve histone refolding during octamer formation. SDS-gels demonstrate the presence (or absence, in the case of H4) of the histone proteins after loading and running samples on nickel-NTA and TSK-SP columns. The removal of impurities within eluted nickel-NTA column fractions through use of the TSK-SP column is shown via comparison of SDS-gels during purification (Figure 5). Histones H2A, H2B, and H3 were successfully purified, while attempts to purify H4 were unsuccessful. The first assessment in determining the cause of product loss within H4 purification is evaluating its initial expression. As

indicated in Figure 4 D, H4 was successfully over-expressed in Rosetta(DE3) pLysS cells. It is possible the cell pellets were not entirely lysed during preparation of cellular inclusion bodies and H4 was lost during washing of cell pellets. The on-put fraction of the nickel-NTA column potentially contains H4, as the bands within the corresponding lane on the SDS-gel are not clearly distinguishable (Figure 5 D). Other lanes within the SDS-gel demonstrate bands that are in the range of H4 MW (~16 kDa). The third lane from right of the ladder shows a band very close to 16 kDa. The presence or absence of H4 within the nickel-NTA column fractions could not be concluded without further protein verification measures, thus several fractions were combined and loaded onto the TSK-SP column. Fractions from the TSK-SP column also demonstrate proteins in a range of 16 kDa, demonstrated by the SDS-gel. These fractions were sent for mass spectroscopic analysis (MS). MS results did not indicate the MW of H4 within any of the samples, and all fractions were discarded. It is possible that H4 did not elute from either the nickel-NTA column or the TSK-SP column as a result of overly tight binding. Further alterations to the protocol could have been made in attempts to obtain a purified product. Due to time limitations, histone H4 was sent to a large-scale purification lab, providing a lyophilized stock of all four histones.

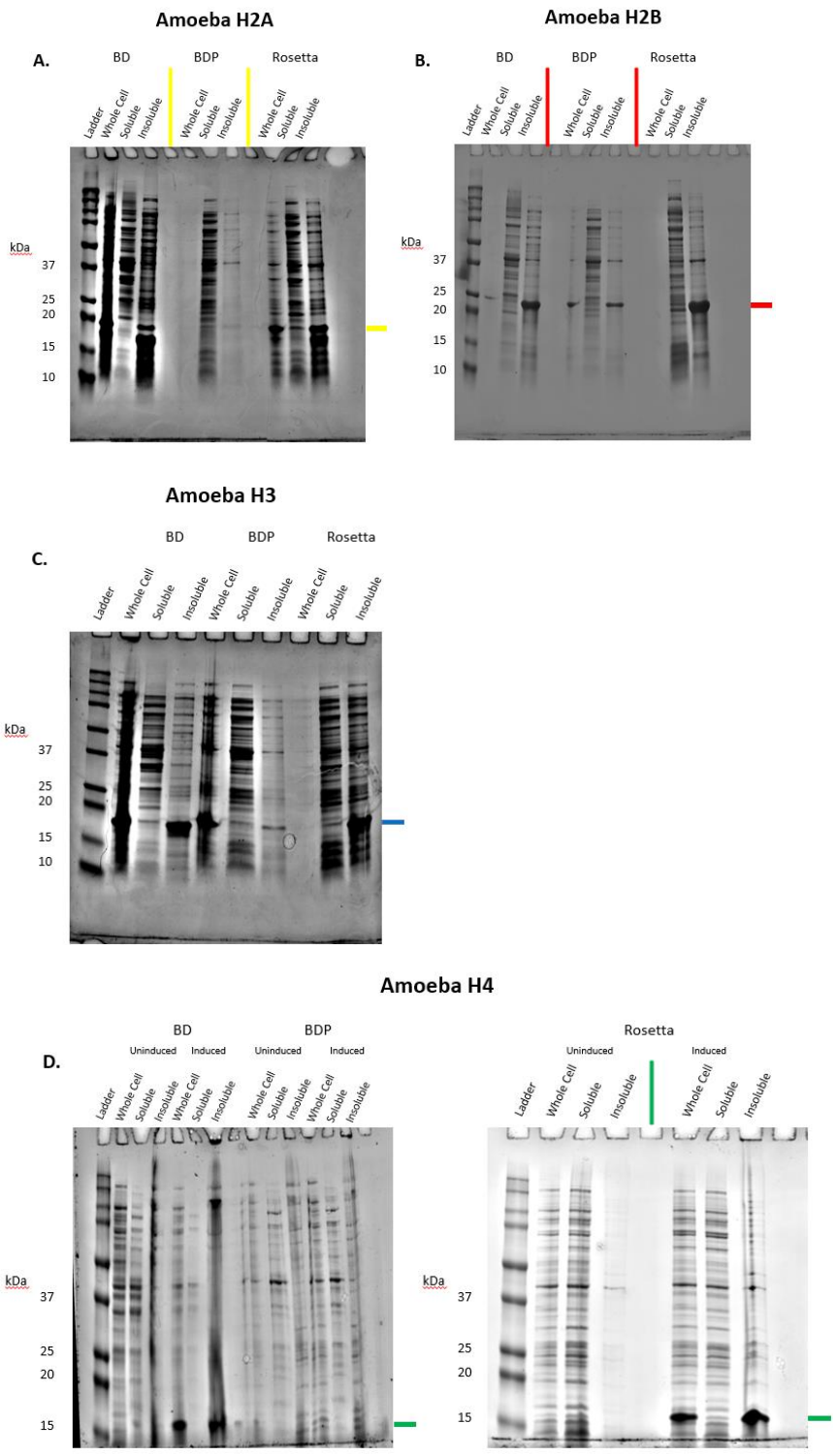


Figure 4. Expression tests of *A. castellanii* in BL21 (DE3) (BD), BL21 (DE3) pLysS (BDP), and Rosetta(DE3) pLysS (Rosetta) *E. coli* cell lines containing whole cell, soluble, and insoluble fractions of cell lysates. Cell lysates were analyzed by SDS PAGE; uncropped gels are shown. (A) Amoeba H2A. (B) Amoeba H2B. (C) Amoeba H3. (D). Amoeba H4. H4 expression test was carried out with uninduced and induced cells as an additional diagnostic for protein expression. H4 did not express in the BDP cell line.

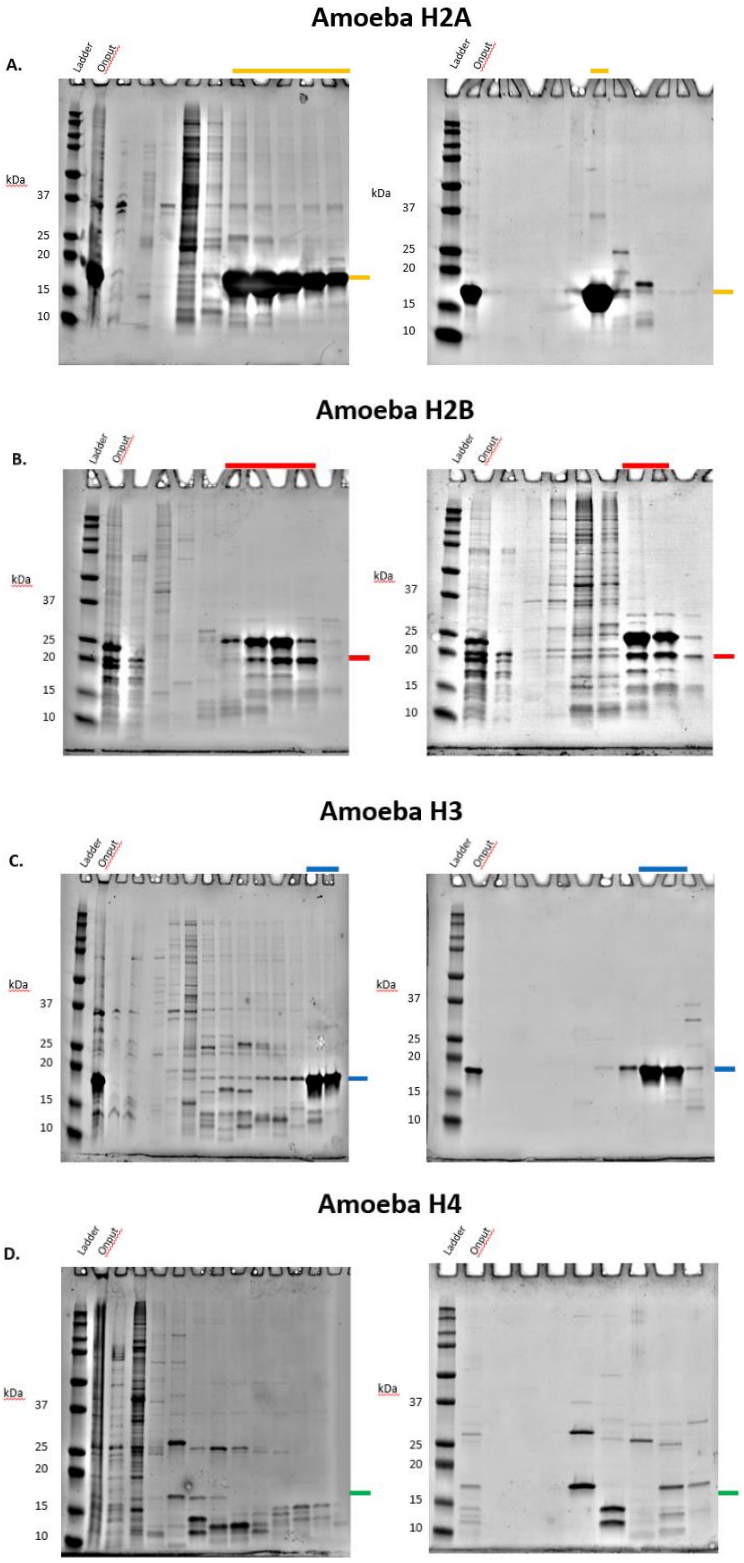


Figure 5. Fractions pooled and combined from nickel-NTA (left) and TSK-SP columns (right) based on increases in absorbance at 280 nm for purification of individual *A. castellanii histones*. (A) Amoeba H2A. (B) Amoeba H2B. (C) Amoeba H3. (D). Amoeba H4.

Acanthamoeba castellanii Octamer Refolding

The first step in nucleosome, or nucleosomal array, reconstitution is preparation of core histone octamers from lyophilized core histones (H3, H4, H2A, and H2B). Histones are combined in equal molar amounts and assembled into octamers by dialyzing the samples out of a denaturing buffer into refolding buffer[20]. The process of octamer refolding generates secondary histone complexes such as aggregates, H3H4 tetramers and H2AH2B dimers, and these are purified away from the octamers using size exclusion chromatography (SEC)[21]. SEC is a chromatographic method that separates molecules by size and shape. During SEC purification, the sample is eluted from the column in refolding buffer and absorbance is measured at 280 nm[22]. Secondary refolding products are detectable at different elution volumes and A280 nm absorbance values with different species eluting in order of decreasing size. Canonical eukaryotic histone aggregates, dimers, and octamers elute at established volumes and provide a guideline for eukaryotic histone octamer purification[21].

The amoeba histone octamer was refolded and purified with eukaryotic (*X. laevis)* octamer as a control. Octamers were refolded using the protocol outlined in the methods. S200 peak fractions of *X. laevis* assumed aggregates, dimers and octamers were run on a BOLT 4-12% Bis-Tris denaturing gel. However initial attempts to refold and purify both amoeba and *X. laevis* histone octamers (as controls) were unsuccessful and refolding attempts remain on-going.

Conclusions and Future Directions

The discovery of encoded histone-like proteins in various members of the *Nucleocytoviricota* has broadened our understanding of viral complexity while raising many questions regarding their implication within the viral life cycle. MV is the first giant virus to demonstrate the formation of nucleosome-like particles with its fused H2B-H2A and H4-H3 histone doublets *in vitro* and to exhibit reliance on these doublets for fitness and infectivity[8,27]. Yet, of the other identified histone-like sequences in NCLDV

genomes, it is unclear if they all maintain the ability to form NLPs. In addition, the viral-host interactions, including those that may facilitate the intracellular transport of viral histones, remain elusive. Further research in this field would aid in the functional elucidation of histone-like proteins in giant viruses, and perhaps reveal why histone-like proteins are not universal amongst NCLDV.

Data from Dr. Chantal Abergel uncovering histone-like doublet proteins, fused in the same arrangement as MV histone doublets, solidified the need for a better understanding of the NCLDV host systems. Secondary structure prediction demonstrating the resemblance of amoeba doublets to canonical eukaryotic histones, and the *in-silico* accommodation of the 3-dimensional doublet homology models within the eukaryotic nucleosome support the idea that amoeba histone-like doublets have the potential to form nucleosome-like particles *in vitro.*

Further sequence analysis to identify other potential hallmark traits of eukaryotic histones, including conserved post-translational modification sites, and the H2A "acidic patch" and "R-D clamps" that support nucleosome stability[28,29] is warranted. In addition, a sequential step in the process of modeling the amoeba doublets is their placement within a canonical eukaryotic nucleosome containing the *A. castellanii* $(H3\text{-}H4)_2$ tetramer as well as a homology models of the structured regions of H2A and H2B, to generate an entire nucleosome particle containing amoeba doublets and canonical amoeba H3-H4. This would offer a broader view of the predicted behavior of amoeba doublets within a nucleosome and how they may behave *in vitro*, beginning fundamentally with the potential clashing of the amoeba doublets with H3-H4. Importantly, despite *de novo* generation of linker structures providing multiple feasible configurations within nucleosomal DNA, they are not sufficient to allow modeling through real-space refinement. Preceding *in vitro* characterization of the *A. castellanii* doublets, the derivation of minimization models of the doublets in the context of a nucleosome would provide initial insight into where and how the doublet amino acid side chains lie in space, and therefore how they would interact

with nucleosomal DNA. However, minimization models would need multiple iterations to obtain the most accurate representation of potential energy levels of the residues in space.

Additional context for how amoeba doublets may behave and function within a nucleosome comes from comparison to viral doublets. The similar pIs of the amoeba and viral doublets (specifically Am3139 and Am3162) might suggest that amoeba doublets would exhibit less tight DNA-binding and formation of a similarly unstable NLP. Notably, *A. castellanii* doublets share less than ~30% sequence identity with MV doublets and less than ~35% sequence identity to eukaryotic *X. laevis* histones, excluding Am0031 (at ~61%). Between the amoeba doublets themselves, Am3139 and Am3162 are nearly identical, diverging from Am0031 and Am8385 (at ~40% and ~20% sequence identity, respectively). This may imply the diversification of *A. castellanii* and MV doublets for distinct functions within the cell. Overall, the characterization of potential amoeba-doublet NLPs *in vitro* would establish a foundation for understanding the potential protein-DNA and protein-protein interactions in an *in-vivo* setting.

The immediate next step in the process of experimentally probing the formation of an amoeba doublet NLP is solving the structure of the canonical amoeba singlet nucleosome, either by x-ray crystallography or by single-particle cryo-EM, to provide the necessary comparison. Identifying potential structural differences between the canonical amoeba nucleosome and an amoeba-doublet NLP would aid in elucidating the function of amoeba doublet-containing nucleosomes and may also help in determining the likelihood of their formation *in-vivo*. Following from this, comparison of a successfully characterized amoeba-doublet NLP to the established MV-doublet NLP is another means of elucidating potential differences in the function of histone doublets within Eukarya and their viral counterparts.

The presence of MV histone doublets within the host nucleus during infection also presents the intriguing possibility of the formation of “hybrid nucleosomes” containing amoeba and viral histones that warrants exploration. Experiments to probe the formation of amoeba singlet, amoeba doublet and viral doublet

histone hybrid nucleosomes can be performed whilst attempting to characterize a putative amoeba doublet nucleosome-like particle. The investigation of potential hybrid nucleosomes would bring greater clarity to our understanding of complex viral-host interactions while perhaps providing clues to the mysterious evolutionary origins of histone doublets.

Methods

Protein BLAST of *Marseilleviridae* genome against eukaryotic histone chaperones

The Basic Local Alignment Search Tool (BLAST)[23] was used to perform a protein query of known eukaryotic, human histone chaperones[16] with the *Marseillevirus* (Taxonomic id: 15134580) protein database. Refer to Supplementary Table 1 for the comprehensive list of human histone chaperones within the query.

Multiple sequence alignment and secondary structure prediction

Predicted *A. castellanii* and *Marseilleviridae* histone doublet proteins were aligned with eukaryotic histone proteins with HHpred's Multiple Alignment using Fast Fourier Transform (MAFFT) with a 1.53 gap open penalty. 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 *A. castellanii, Marseilleviridae* and *Eukarya*[24].

Homology modeling

Homology models of *A. castellanii* doublets Amoeba_H2BH2A_3139 (Am3139), Amoeba_H2BH2A_3162 (Am3162), Amoeba_H2BH2A_0031 (Am0031), and Amoeba_H2BH2A_8385 (Am8385) were constructed using SWISSMODEL, and the histone structures from the *Xenopus laevis* nucleosome (PDB 1AOI) were used as reference[18,25]. Using Modeler (v9.20), 10,000 loops of each linker region of the doublets (3139 residue 93 to 118, 3162 residue 93 to 125, 0031 residue 98 to 109, 8385 residue 92 to 131) were generated. 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 overlapping atoms was used[26]. Any loops identified by CPPTRAJ without a clash were manually parsed though to eliminate loops that were not sterically overlapping but also not physically relevant, such as those forming "knots" in the histone folds and directly intersecting the double helix. Multiple sterically and physically agreeing configurations of loop models were selected via manual parsing. Figures were rendered using Chimera and VMD.

Expression of *Acanthamoeba castellanii* histones

pET15-dTOPO plasmid DNA of each *A. castellanii* histone was provided by VectorBuilder and transformed into BL21 (DE3), BL21 (DE3) pLysS, and Rosetta(DE3) pLysS *Escherichia coli (E. coli)* cell lines via heat shock as above. Each histone had its expression tested at various conditions, considering the soluble and insoluble fraction of each (Figure 4, A-D). After confirming optimal expression, each individual histone was grown up in 6 liters of 2XTY and induced with 1 mL IPTG when optimal density reached the range of 0.4-0.6. Cells were growth for 2 h post induction and stored at -20 C for purification.

Inclusion body preparation

Preceding harvesting of cells, cellular inclusion bodies of each histone were isolated. Stored cells were resuspended in wash buffer (50 mM Tris pH 7.5, 100 mM NaCl, 1mM 2-Mercaptoethanol (BME), 1mM Benzamide) and blended using the Tekmar Tissuemizer. To remove free DNA, DNase was added to the blended cells and allowed to incubate at 25 C for 30 minutes. Following incubation of DNase, inclusion bodies were isolated using the protocol outlined in[20].

Purification *Acanthamoeba castellanii* histones

Purification protocols were initially varied and eventually optimized for protein purity and yield. Following optimization, the below protocol was implemented for each histone.

Proteins were eluted from a His Trap HP 5 mL column at RT, using elution buffer (50 mM NaCl, 20 mM Tris pH 7.5, 7M Urea, 500 mM Imidazole). Eluted fractions were run on 15% Tris-glycine SDS for 35 minutes, 200V. Selected fractions were combined and eluted on a 55 mL TSK-SP column, using TSK-SP-elution buffer (1M NaCl, 20 mM Tris pH 7.5, 7M Urea, 5 mM BME, 1 mM EDTA) overnight. Eluted fractions were run on 15% Tris-glycine SDS for 35 minutes, 200V. Protein fractions were combined and dialyzed against 1mM BME and water for approx. 18 h.

Dialyzed proteins were lyophilized and stored at 4°C.

Acanthamoeba castellanii histone refolding

Refer to the protocol within reference[20].

Acknowledgements

I want to thank the Biological Sciences Initiative of CU Boulder for funding a portion of the time spent conducting research and for providing guidance and support throughout the entire project timeline. I also want to thank the Undergraduate Research Opportunities Program of CU Boulder for providing grant funding for a portion of my project. Lastly, I want to acknowledge my graduate student mentor, Chelsea Toner, who provided countless hours of guidance and instruction throughout the entirety of my project. Her dedication as a mentor ultimately allowed my undergraduate honors thesis to come to fruition.

References

1. Yutin, N. *et al.* Eukaryotic large nucleo-cytoplasmic DNA viruses: Clusters of orthologous genes and reconstruction of viral genome evolution. *Virol. J.* **6**, 167-202 (2009).

2. Koonin, E.V., Yutin, N. Evolution of the Large Nucleocytoplasmic DNA Viruses of Eukaryotes and Convergent Origins of Viral Gigantism. *Adv. Virus Res.* **103**, 167-202 (2019).

3. Complete genomes: Herpesviridae. *National Center for Biotechnology Information.* https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10292.

4. Walker, P. J. *et al.* Changes to virus taxonomy and to the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses. *Arch. Virol.* **166**, 2633–2648 (2021).

5. 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* **7**, 000649 (2021).

6. Koonin, E.V. *et al*. Global Organization and Proposed Megataxonomy of the Virus World. *Microbiol. Mol. Biol. Rev*. **84**, e00061-19 (2020).

7. Colson, P., Scola, B. L., Raoult, D. Giant Viruses of Amoebae: A Journey Through Innovative Research and Paradigm Changes. *Annu. Rev. Virol.* **4**, 61-85 (2017).

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

9. Boyer, M. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl. Acad. Sci. USA.* **106**, 21848-53 (2009).

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

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

12. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260 (1997).

13. Doutre, G., Philippe, N., Abergel, C., Claverie, J.M. Genome Analysis of the First Marseilleviridae Representative from Australia Indicates that Most of Its Genes Contribute to Virus Fitness. *J. Virol.* **88**, 14340-14349 (2014).

14. Arantes, T. *et al.* The Large Marseillevirus Explores Different Entry Pathways by Forming Giant Infectious Vesicles. *J. Virol.* **90**, JVI.00177-16 (2016).

15. 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).

16. 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).

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

18. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**, 296–303 (2018).

19. Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinforma.* **15**, 5.6.1-5.6.30 (2006).

20. Luger, K., Rechsteiner, T. J., Richmond, T. J. Expression and Purification of Recombinant Histones and Nucleosome Reconstitution. *Chromatin Protoc.* **119**, 1-16 (1999).

21. Rogge, R. A. *et al.* Assembly of Nucleosomal Arrays from Recombinant Core Histones and Nucleosome Positioning DNA. *J. Vis. Exp. JoVE* **79**, 50354 (2013).

22. Hong, P., Koza, S., Bouvier, E.S. Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *J Liq Chromatogr Relat Technol*. **35**, 2923-2950 (2012).

23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).

24. 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).

25. Guex, N., Peitsch, M. C., Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **30 Suppl 1**, S162-173 (2009).

26. 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).

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

28. Mattiroli, F. *et al.* Structure of Histone-based Chromatin in Archaea. *Science* **357**, 609-612 (2017).

29. McGinty, R. K., Song, T. Recognition of the nucleosome by chromatin factors and enzymes. *Curr. Opin. Struct. Biol.* **37**, 54-61 (2016).

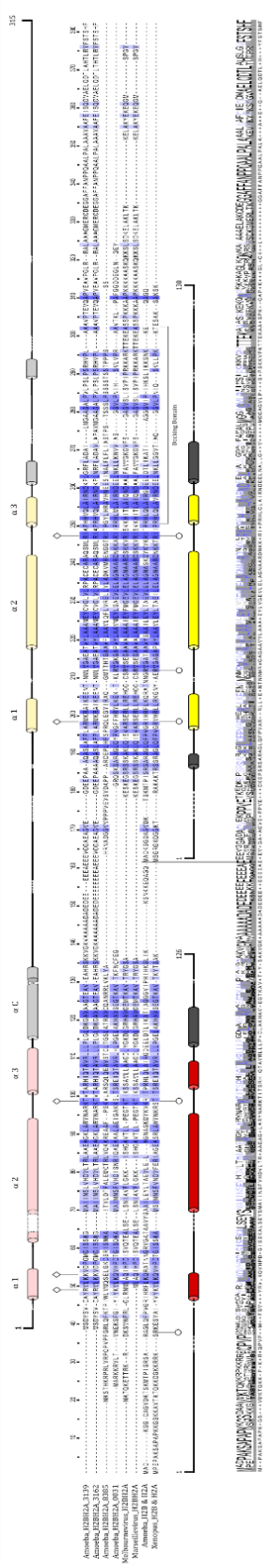
Supplemental

Supplementary Table 1. A comprehensive list of known human histone chaperones used in the protein query of a BLAST search with the *Marseillevirus* protein database; ND: not defined. Table adapted from reference[15]. No homologues of human chaperones were identified within the *Marseillevirus* genome.

Histone chaperone	Histone preference	Complex (or complexes)	Domains	Uniprot Accession #
Anti-silencing function 1A/B (ASF1A/B)	H3.1–, H3.2–, H3.3–H4	Multiple	ASF1, Ig-like, B-domain	Q69DB9
Minichromosome maintenance protein 2 (MCM2)	CENP-A–, H3.1–, H3.2–, H3.3–H4	MCM2–7 complex	AAA+ ATPase, HBD	Q7T0M6
Tonsuku Like (TONSL)	H3–H4	MMS22–TONSL complex	ARD, TPR, LRR	Q91883
Retinoblastoma associated protein 46 (RBAP46)	H3–H4	HAT, HDAC, NuRF, NuRD, PRC2	WD40	Q6AZV1
Heat shock protein 90A/B (HSP90A and HSP90B)	H1, H2A, H2B, H3, H4	Heat shock family	HSP90, ATPase-like, TPR binding motif	P41154
Heat shock cognate 70 (HSC70)	H1, H2A, H2B, H3, H4	Heat shock family	HSC70, ATPase domain, peptide binding domain	A0A1L8GFM0
Somatic nuclear autoantigenic sperm protein (sNASP)	H3.1–, H3.2–, H3.3–H4, H1	H AT	TPR	Q2VPH0
Importin 4 (IPO4)	H3.1–, H3.2–, H3.3–H4	Nuclear import receptor	HEAT domain, Armadillo-type-fold	Q4U0Y4
Suppressor of Ty 2 (SPT2)	H3–H4	ND	SPT2, HMG box	Q7ZY81
Suppressor of Ty 6 (SPT6)	H3–H4	ND	SPT6 core domain, SH2	P07222
Acidic-leucine-rich nuclear phosphoprotein 32E (ANP32E)	H2A.Z–H2B	P400–TIP60	LRR, ZID domain	P20397
Protein YL1 (YL1), vacuolar protein sorting 72 homologue (VPS72)	H2A.Z-H2B	SRCAP/SWR-C, P400–TIP60	ZID domain	Q06459
Holliday junction recognition protein (HJURP)	CENP-A-H4 (Cse4-H4)	Centromere assembly	Scm3/HJURP	B7ZR98
Patient SE translocation (SET)	H3–H4	INHAT, Vps75–Rtt109	NAP1-like	B7ZR96
Nucleosome assembly protein 1-like (NAP1L1-6)	H2A–, H2A.Z–H2B, H3–H4, H1	Nuclear import importin 9, Kap114	NAP1-like	Q98TA5
Nucleophosmin (NPM1)	H3–H4, CENP-A–H4, H1	SWAP	Nucleoplasmin	A0PCL5

(*Continued on next page)*

Nucleoplasmin 2 (NPM2)	H2A–H2B	ND	Nucleoplasmin	Q07CZ6
Nucleoplasmin 3 (NPM3)	ND	NPM1	Nucleoplasmin, NPM1 binding	A0JMT0
Nucleolin (NCL)	H2A–H2B, H1	SWAP	RNA binding, DNA helicase/ATPase	Q8QFR2
Chromatin assembly factor 1, subunit A (CHAF1A or p150)	H3.1–H4	CAF1	PIP-box, p60 binding, HP1-binding, WH domain	A0A1L8HQE8
Chromatin assembly factor 1, subunit B (CHAF1B or p60)	H3.1–H4	CAF1	WD40, B-domain	Q69DB9
Retinoblastoma associated protein 48 (RBAP48)	H3.1–, H3.2–, H3.3–H4	CAF1, HDAC, NuRF, NuRD, and PRC2	WD40	Q90W61
Histone regulation A (HIRA)	ND	HIRA/HIR	WD40, B-domain	A0A1L8EYU1
Ubinuclein 1 (UBN1)	H3.3–H4	HIRA/HIR	UBN1, HRD	A0A1L8I0F0
Calcineurin-binding protein cabin-1 (CABIN1)	ND	HIRA/HIR	TPR	A0A1L8I0E3
Suppressor of Ty 16 (SPT16)	H2A–H2B, H3–H4	FACT	Peptidase-like pleckstrin homology	Q9W603
Structure-specific recognition protein 1 (SSRP1)	H2A–H2B, H3–H4	FACT	Pleckstrin homology, HMG box	Q9W602
Death domain-associated protein 6 (DAXX)	H3.3–H4	DAXX–ATRX	HBD, four-helix bundle, SIM, PML targeting	A0A1L8HZH1
Alpha-thalassaemia/mental retardation syndrome X-linked (ATRX)	ND	DAXX–ATRX	Swi/Snf2-related helicase, ADD, HP1 binding, MeCP2 binding	B7ZQW8



(*figure legend on next page)*

S1. Complete Multiple Sequence Alignment of *A. castellanii* histone-like doublets

Complete sequence alignment of proposed H2B-H2A histone doublets of *A. castellanii,* H2B-H2A histone doublets of *Marseilleviridae,* canonical H2B-H2A histones of *A. castellanii,* and canonical H2B-H2A of *X. laevis.* Predicted α helices of H2B-H2A (light red and yellow) of *A. castellanii* proposed histone doublets were generated using HHPRED's Quick 2D prediction web server. Canonical histone dimers H2B-H2A of *A. castellanii* and *X. laevis* were each aligned to *Marseilleviridae* and *A. castellanii,*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-T pairs and DNA-binding residues are indicated in the *X. laevis* histone dimer (bottom) with their conservation within *A. castellanii* proposed histone doublet, Am3139 (top). Known α helices form the histone fold domain in *X. laevis* are shown as dark colored tubes; H2B helices are red, H2A are yellow, and additional helices are in gray. Logo plot demonstrating residue conservation among the alignments provided by ClustalΩ tool is shown below.