MRC London Institute of Medical Sciences
Imperial College - Faculty of Medicine

Histone variants in archaea

Kathryn Mary Stevens

Molecular Systems Group
DNA Replication Group

A thesis submitted to Imperial College London for the degree of Doctor of Philosophy, September 2022.

‘The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.’

For Will. My other half, best friend,
and the only person I could survive lockdown with.

Declarations

The work presented in this thesis is my own, unless specified. Where others have carried out work, this has been acknowledged in the text, figure legends and/or methods as appropriate. Some of the work and figures presented in this thesis has previously been published in PNAS, Genome Biology and Evolution and Seminars in Cell and Developmental Biology (Stevens *et al.* 2020, 2022, Stevens and Warnecke 2022). Publications are included as appendices to the end of the thesis.

Acknowledgements

First and foremost, I want to thank my supervisor Toby, who is just as awesome a person as scientist, and a great source of inspiration. I am really lucky to have had you as a supervisor and grateful for all your support during my PhD (despite the chaos of a global pandemic). Thank you for giving me the freedom to develop ideas and for keeping me on track when I fell down a rabbit hole. Beyond this, thank you for cultivating a group of amazing people in the Molsys lab, past and present.

Thank you, Jake, for being so patient with me even though things always seemed to go wrong on a Friday afternoon. Antoine, this work would not have anywhere near been half as fun nor complete without you. Your passion for science is infectious, and you are a genuine role model. Maria, thank you for all your advice and support, particularly for the many MNase experiments…I'm indebted to all of you for your help and insights throughout my PhD. Sasha, Vas, Val, Shivani, Romain, and Chahrazad, you have been infinitely supportive, and I couldn't have asked for a better group of people to work with.

To my other supervisor, Christian, and members of the DNA replication lab, thank you for the generosity of your time, keen observations, and for making me feel part of the group all the way from CRB.

I also want to thank my amazing family. To my parents, siblings, and my Gran, thank you for your endless encouragement and for reminding me what's important in life. Mum and Dad, any success I have ever had is only because you believed in me and taught me how to be resilient.

Last but not least, to my partner, Will. I can't put into words how much I owe you. In moments where everything has felt impossible, you have always been there. Thank you.

Table of Contents

1. Abstract 13

2. Introduction 14

2.1. Archaea and the Tree of Life 14

2.2. Chromatin architecture in different kingdoms 15

2.3. Histone-based chromatin in eukaryotes 18

2.3.1. Dynamic changes to chromatin structure 19

2.4. Histone variants in eukaryotes 20

2.4.1. The importance of substitutions and more divergent sequence changes 22

2.4.2. The evolution of histone variants in eukaryotes 23

2.5. Histones in archaea: an overview 25

2.6. Structure of archaeal histones 26

2.7. Histones in context 29

2.8. The case for histone variants in archaea 31

2.8.1. Histones which vary in their structural and DNA binding properties 33

2.8.2. Deletion studies 35

2.8.3. Histone paralogs which show differences in expression levels 37

2.8.4. Differential post-translational modification of archaeal histones 38

3. Aims and Objectives 40

4. Dissertation layout 40

5. Predicted structural diversity of archaeal histones 42

5.1. Introduction 42

5.2. Initial testing: *in silico* DNA binding predictions for *M. fervidus* histones HMfA and HMfB 42

5.3. Generating tetrameric histone-DNA complex models for different archaeal species 43

5.4. Histone-DNA complexes exhibit a wide range of tetramerization strengths and DNA binding affinities 46

6. Structural diversity of *Methanosphaera stadtmanae* histones 49

6.1. Introduction 49

6.2. *M. stadtmanae* encodes 7 histones 49

6.3. *M. stadtmanae* histone complexes occupy a broad range of predicted structural properties 51

6.4. Using *in vivo* expression levels of *M. stadtmanae* histones to predict changes to histone-based chromatin in stationary vs exponential phase 54

6.5. Molecular dynamics simulations of *M. stadtmanae* histone homotetramers reveal histones with unique properties 60

6.6. Capstones outside the Methanobacteriales 66

6.7. Conservation of unique properties across capstones from different species....67
6.8. Summary....69
7. Conservation of histone paralogs in Methanobacteriales....70
7.1. Introduction....70
7.2. Methanobacteriales archaea have many histones....70
7.3. There are three ancient histone paralogs in the Methanobacteriales....72
7.4. Some Methanobacteriales species have an increased number of histones....76
7.5. There is little evidence for recombination in Methanobacteriales histones....77
7.6. Phylogenetic analysis of capstone histones in Methanobacteriales....79
7.7. Summary....81
8. Conservation of histone paralogs in Thermococcales....82
8.1. Introduction....82
8.2. Identifying histone paralogs with HTkA-like and HTkB-like properties in Thermococcales species....84
8.3. HTkA-like and HTkB-like histones are almost ubiquitous in Thermococcales....91
8.4. HTkA/B-like histone paralogs do not extend beyond the Thermococcales....96
8.5. Some Thermococcales species encode for histones similar to those in some bacteria....99
8.6. HTkA- and HTkB-like histones have a separate origin to A/B histones in the Methanobacteriales....105
8.7. Summary....106
9. Histone diversification through specific amino acid changes in archaea and eukaryotes....108
9.1. Introduction....108
9.2. *In silico* mutational analysis of the HMfB hexamer complex....109
9.3. Substitutions drive the diversification of archaeal histones....113
9.4. Diversification of eukaryotic histone sequences in comparison to archaeal histones....115
9.5. Summary....118
10. Exploration of recombinant *M. stadtmanae* histone Msp_0769 and capstone Msp_0383....120
10.1. Introduction....120
10.2. Transformation of *M. stadtmanae* histones Msp_0383 and Msp_0769 into *E. coli* and novobiocin phenotype....121
10.3. MNase digestion of *E. coli* chromatin expressing Msp_0769 and Msp_0383....123
10.4. Purification of recombinant *M. stadtmanae* histones Msp_0769 and Msp_0383 from *E. coli*....128
10.5. Electrophoretic mobility shift assay....131
10.6. *In vitro* MNase digestion....133
10.7. Summary....135
11. Discussion....136

11.1. Outstanding questions 137
11.2. The role of histones in archaea in general 137
11.2.1. What drives change in the number of histones? 138
11.2.2. Archaeal histone variants: a global or a local role? 139
11.2.3. The interplay between histone variants and other NAPs 140
12. Methods 142
12.1. *In silico* modelling 142
12.1.1. Single mutation analysis of HMfB 142
12.1.2. Generating tetrameric models of archaeal histone-DNA complexes (fast-mutational scanning) 142
12.1.3. Generating models of *M. fervidus* histones HMfA/B as homodimeric and heterodimeric hexamer complexes, or as forced dimer tetramer complexes 144
12.1.4. Random generation of *M. stadtmanae* histone models in exponential and stationary phase 145
12.1.5. Molecular dynamics simulations 145
12.1.6. AlphaFold structure predictions 146
12.2. Phylogenetic and sequence-based analysis 146
12.2.1. Alignment of protein sequences 146
12.2.2. Phylogenetic tree building and visualisation of phylogenetic trees 146
12.2.3. Archaeal and eukaryotic histone sequence diversity and similarity 147
12.2.4. Recombination detection 148
12.2.5. Identification of single and double HFD proteins across archaea 148
12.2.6. Phylogenetic analysis of histones in Methanobacteriales 149
12.2.7. Identification of histones in Thermococcales 150
12.2.8. Classification of Thermococcales histones into HTkA-like, HTkB-like, and bacteria-type singlets/doublets 150
12.2.9. Pairwise percentage sequence identity for HTkA/B-like histones and bacteria-type histones in Thermococcales 151
12.2.10. Identification of HTkA- and HTkB-like histones in Thermococcales outgroup species, and comparison to *M. fervidus* histones HMfA/B 151
12.2.11. Identification of orthologs and genome order conservation in Thermococcales 152
12.2.12. Expression level of histones 152
12.3. *In vitro* analysis of *M. stadtmanae* histones and heterologous expression in *E. coli* 153
12.3.1. Generation of *E. coli* strains expressing archaeal histones from *M. stadtmanae* 153
12.3.2. Bacterial growth 154
12.3.3. Expression tests 154
12.3.4. Purification of archaeal histones 155
12.3.5. Staining of protein gels with Coomassie blue 156
12.3.6. Measuring protein concentration 157
12.3.7. Electrophoretic Mobility Shift Assay (EMSA) 157

12.3.8. In vitro MNase digestion 158
12.3.9. MNase digestion of *E. coli* chromatin from cells expressing histones from *M. fervidus* or *M. stadtmanae* 158
12.3.10. Novobiocin sensitivity assay 158
13. References 160
14. Appendix 186

List of Figures

Figure 2.1 Tree of life model indicating the relationships between archaea, bacteria and eukaryotes. 15
Figure 2.2 Structural organisation of the eukaryotic nucleosome. 18
Figure 2.3 Histone variants increase the complexity of the nucleosome structure. 21
Figure 2.4 Comparison of archaeal and eukaryotic histones. 27
Figure 2.5 Model of oligomeric histone complexes in archaea. 29
Figure 2.6 Number of histones per species in different lineages of archaea. 32
Figure 2.7 Structure of archaeal histones with duplication of the HFD. 33
Figure 5.1 Length distribution of archaeal HMf-like histones used for modelling. 44
Figure 5.2 Generating models for all combinations of tetrameric histone complexes for each species. 45
Figure 5.3 DNA binding and tetramerisation strength for tetrameric histone-DNA complexes based on the HMfB crystal structure. 47
Figure 6.1 Location and chromosomal context for the 7 histone genes in *M. stadtmanae* DSM 3091. 50
Figure 6.2 Alignment of the 7 histones in *M. stadtmanae*. 50
Figure 6.3 DNA binding and tetramerization strength (interaction strength between dimers) for methanogenic species *M. fervidus*, *M. smithii* and *M. stadtmanae*. 51
Figure 6.4 Modelling results for *M. stadtmanae* histones. 53
Figure 6.5 Expression level of the 7 histones in *M. stadtmanae*. 55
Figure 6.6 Histone expression level in *M. stadtmanae* in exponential phase and stationary phase of growth. 57
Figure 6.7 Histone expression level and tetramerization strength for different archaeal species. 59
Figure 6.8 DNA binding and tetramerization for *M. stadtmanae* histone tetramer models over 100ns compared to the fast-mutational scanning approach 61
Figure 6.9 Trajectory *of M. stadtmanae* homotetrameric histone-DNA models during MD simulations. 63
Figure 6.10 Distance between centre of mass for Msp_0383 reversion mutants. 65
Figure 6.11 Phyletic distribution of capstone histones in archaea. 67
Figure 6.12 Molecular dynamics simulations of histones from species containing putative capstone histones. 68
Figure 7.1 Species tree of the order Methanobacteriales showing the number of histone paralogs in each genome. 71
Figure 7.2 Toy species- and protein-level trees for relationships between histones in archaea. 72
Figure 7.3 Long-term maintenance of histone paralogs in Methanobacteriales 73

Figure 7.4 Comparison of protein-level tree and species tree for Methanobacteriales paralog groups. 75
Figure 7.5 Detection of recombination events in the Methanobacteriales. 78
Figure 7.6 Phylogenetic analysis of potential capstone histones in Methanobacteriales. 80
Figure 8.1 Number of histones in Thermococcales genomes. 85
Figure 8.2 Identification of HTkA-like and HTkB-like histones in Thermococcales. 87
Figure 8.3 Predicted DNA binding and tetramerization strengths for Thermococcales HTkA- and HTkB-like histones. 89
Figure 8.4 Comparison of lysine acetylation in eukaryotic histone H3, and Thermococcales histones. 90
Figure 8.5 HTkA-like and HTkB-like histone paralogs are found in almost all Thermococcales species. 91
Figure 8.6 Protein-level tree of Thermococcales HTkA-like and HTkB-like histones. 93
Figure 8.7 Gene order conservation in Thermococcales. 95
Figure 8.8 Whole genome alignment centred on HTkA/B-like histones. 96
Figure 8.9 Clustering of histones from outgroup Methanofastidiosales, Theionarchaeota and NM3 species. 98
Figure 8.10 Bacteria-type histones in Thermococcales species. 100
Figure 8.11 Protein-level phylogenetic trees of bacteria-type histones showing genomic neighbourhood. 102
Figure 8.12 Species-level phylogenetic trees of bacteria-type histones showing genomic neighbourhood centred on conserved syntenic region. 103
Figure 8.13 Expression level of histones in *T. kodakarensis* and alpha fold prediction of HTkA/B with bacteria-type singlets. 104
Figure 8.14 A/B type histones in Methanobacteriales do not have the same origin as HTkA/B histones in Thermococcales spp. 106
Figure 9.1 Sites sensitive to structural change in archaeal histones. 109
Figure 9.2 Interrogation of the HMfB structure using in silico single mutations. 111
Figure 9.3 In vitro gel shift assay for HMfB single mutants compared to in silico results. 112
Figure 9.4 Sequence diversification of archaeal histones. 114
Figure 9.5 Sequence diversification of archaeal histones without Asgard archaea dramatically reduces diversity at position 54. 115
Figure 9.6 Length filtering of eukaryotic H3 and H4 histones. 116
Figure 9.7 Diversification of eukaryotic histones H3 and H4 in comparison to archaea with multiple histone copies. 117
Figure 10.1 Novobiocin sensitivity assay for *E. coli* strains expressing *M. stadtmanae* histones Msp_0769 and Msp_0383. 122
Figure 10.2 Expression tests for *E. coli* strains expressing histones from *M. stadtmanae*. 124

Figure 10.3 in vivo MNase digestion of *E. coli* K12 strains expressing histones HMfA from *M. fervidus* and Msp_0383, Msp_0769 from *M. stadtmanae*. 125

Figure 10.4 in vivo MNase digestion of *E. coli* K12 strain expressing histone HMfA from *M. fervidus* and *E. coli* strain BL21(DE3) expressing histone Msp_0769 from *M. stadtmanae*. 127

Figure 10.5 Purification of Msp_0383 from K12-msp0383. 129

Figure 10.6 Purification of Msp_0769 from BL21-msp0769. 130

Figure 10.7 Purification of HMfB. 131

Figure 10.8 Electrophoretic mobility shift assay (EMSA) for purified histones HMfB, Msp_0383 and Msp_0769. 132

Figure 10.9 in vitro MNase digestion patterns for HMfB with and without Msp_0383. 134

Figure 14.1 DNA binding strength and tetramerization strength for archaeal tetrameric complexes with species names annotated. 190

Figure 14.2 Distribution of substitutions across histones in archaea. 191

List of Tables

Table 2.1 Differential expression of histones in response to environmental change. 37

Table 7.1 Significant recombination events detected by RDP4 in the Methanobacteriales. 79

Table 12.1 Primer sequences for *M. stadtmanae* recombinant histones. 154

Table 14.1 Genomes with candidate capstones. 186

Table 14.2 HMfB single amino acid mutants characterised in gel shift assays. 187

Table 14.3 HMfB and *S. cerevisiae* H3/H4 sequences and residue numbering for each position in the alignment used for diversity analysis. 189

1. Abstract

Eukaryotic histone variants are involved in a wide range of processes and play a key role in altering nucleosome dynamics to shape the architecture of chromatin. The importance of individual variants has been studied extensively in many eukaryotes. In comparison, we know relatively little about histones in archaea. Despite sequence variation and evidence for potential functional differences between histone paralogs in the same species, whether archaea have histone variants, and therefore the potential for complex histone-based chromatin, has not been comprehensively explored. In this work, I apply structural and sequence-based approaches and present evidence that histone variants exist in archaea. *In silico* modelling suggests that, similarly to some eukaryotic variants, paralogs in archaea can be identified by unique structural properties. In particular, I describe one such variant, a "capstone", that can drastically alter histone-based chromatin by limiting oligomerisation. Other paralogs have less extreme structural properties but are shared between species which separated hundreds of millions of years ago, on par with some eukaryotic histone variants. Although there are shared features between the two, histones in archaea have appear to have explored a different sequence space to eukaryotic histones, evolving separately and in parallel.

2. Introduction

2.1. Archaea and the Tree of Life

For most of the 20th century, living systems were classified as either eukaryotes or prokaryotes. During this time, determining the relationships between organisms was generally reliant comparisons of morphology, embryology and on fossil records (Sapp and Fox 2013). In 1977, Carl Woese and George Fox proposed a novel group of prokaryotes: archaea (Woese and Fox 1977). The discovery of archaea split life into three domains, in which bacteria, eukaryotes and archaea all share a common ancestor (Figure 2.1) (Woese *et al.* 1990). For a long time, this was the accepted view, and archaea were divided into two kingdoms, the Crenarchaeota and Euryarchaeota (Woese *et al.* 1990). Sequencing of more diverse archaeal species revealed novel superphyla, such as the 'TACK' (Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota) superphylum (Brochier-Armanet *et al.* 2008, Elkins *et al.* 2008, Guy and Ettema 2011, Nunoura *et al.* 2011). Interrogation of the TACK archaea uncovered many eukaryotic signature proteins (ESPs) suggesting this lineage represents a close relative of eukaryotes. This suggested a major shift in the tree of life towards a two domain model, where eukaryotes fall within the diversity of archaea as a sister lineage to the TACK superphylum (Guy and Ettema 2011, Williams *et al.* 2013).

In recent years, extensive metagenomic sequencing efforts have identified other archaeal species that are key to understanding the topology of the tree of life. In particular, this includes the Asgard archaea which have genes encoding many unique ESPs (Spang *et al.* 2015, Seitz *et al.* 2016, Zaremba-Niedzwiedzka *et al.* 2017, Liu *et al.* 2021). Re-assessment of the tree of life including these species further reinforces a two-domain view, with eukaryotes branching from within the archaea as a sister clade to the Asgard archaea (Figure 2.1) (Spang *et al.* 2015, Eme *et al.* 2017, Liu *et al.* 2021, Xie *et al.* 2022), although this is occasionally still debated (Nasir *et al.* 2021).

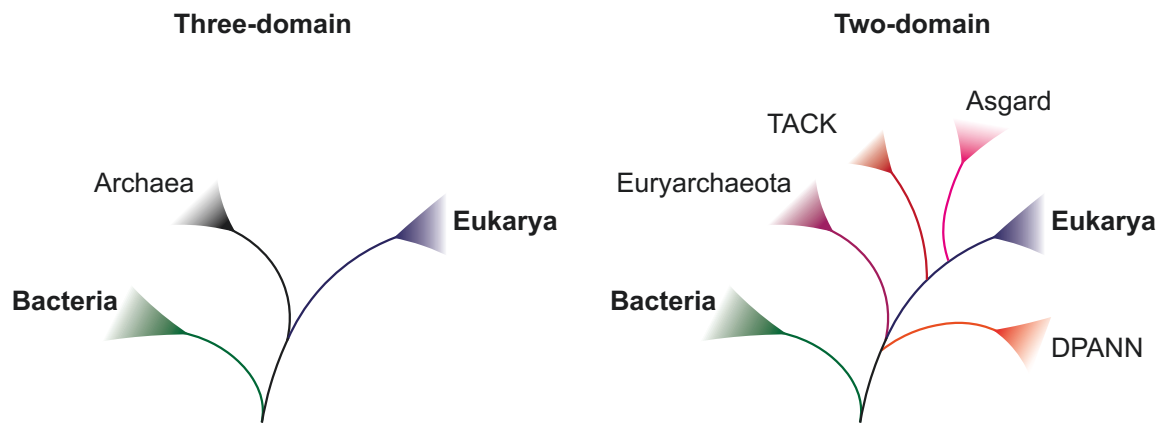


Figure 2.1 Tree of life model indicating the relationships between archaea, bacteria and eukaryotes. Models of both the three-domain (left) and two-domain (right) architectures are shown. The two-domain model includes the main archaeal superphyla/phyla; Euryarchaeota, Asgard, TACK and DPANN. Both trees are rooted at the last universal common ancestor. Tree topology based on (Eme et al., 2017).

The presence of ESPs in the Asgard archaea suggests that the archaeal ancestor of eukaryotes may have already had many of the key components needed for eukaryotic complexity (Guy and Ettema 2011, Williams *et al.* 2013, Spang *et al.* 2015). The universal organisation of eukaryotic chromatin by histone proteins is thought to be one such system. The core building blocks for histone-based chromatin, already present in archaea, may have helped maintain genome stability in high temperature environments and allowed for complex genome organisation in eukaryotes (Pereira *et al.* 1997, Brunk and Martin 2019). Below I will first describe similarities and differences in chromatin architecture between archaea and eukaryotes, before looking at the individual components that mediate these architectures.

2.2. Chromatin architecture in different kingdoms

The organisation of genomic DNA into chromatin is a universal feature of all forms of life (Luijsterburg *et al.* 2008, Maruyama *et al.* 2021). The overall structure of chromatin shares many features between bacteria, eukaryotes and archaea, despite major differences in gene organisation and genome size. In some eukaryotes, including many mammals, zebrafish and *Drosophila*, chromatin is folded into topologically associated domains (TADs) which are globular units marking intradomain interactions (Dixon *et al.* 2012, Hou *et al.* 2012, Sexton *et al.* 2012, Kaaij *et al.* 2018). TADs are not universal across eukaryotes, even within animals, demonstrated by the limitation to the X chromosome of *Caenorabditis elegans* (Crane *et al.* 2015). TADs also are scarce in plants, for example they are limited to a few loci in *Arabidopsis thaliana* (Feng *et al.* 2014, Dong *et al.* 2017). TAD-like self interacting domains have been

identified in yeast *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, although these are much smaller (50-100kb in *S. pombe*) and contain few genes (one to five in *S. cerevisiae*) (Mizuguchi *et al.* 2014, Hsieh *et al.* 2015). In the distantly related protist *Tetrahymena thermophila*, TAD-like structures are found in one of the two nuclei (the germline micronucleus) (Luo *et al.* 2020). Additionally, organisation of the genome into gene-rich, active A compartments, and inactive B compartments has been described for some eukaryotes including humans and mice (Lieberman-Aiden *et al.* 2009, Wang *et al.* 2016).

Bacteria contain similarly large self-interacting domains, known as chromosome interaction domains (CIDs) (Le *et al.* 2013, Marbouty *et al.* 2015, Lioy *et al.* 2018). Unlike eukaryotes, A- and B-type compartments have so far not been found in bacteria (Le *et al.* 2013, Marbouty *et al.* 2015, Lioy *et al.* 2018, Takemata and Bell 2020).

Hi-C and chromosome conformation capture (3C) studies have also given us a first glimpse of the overarching chromatin organisation in a few species of archaea. Self-interacting chromosome domains, akin to CIDs have been identified across distantly related organisms – in *Haloferax volcanii*, a member of the Euryarchaea (Cockram *et al.* 2021) and *Sulfolobus acidocaldarius* and *Sulfolobus islandicus*, members of the Crenarchaea (Takemata and Bell 2021). Transcriptionally active and inactive A/B compartments have been found in *Sulfolobus* (Takemata *et al.* 2019), but this feature is not universal across archaea, and appears to be absent in both *Thermococcus kodakarensis* and *Haloferax volcanii* (Cockram *et al.* 2021)).

Despite some similarities in the overall architecture, the composition of the repeating, basic structural units of chromatin varies significantly between domains (Luijsterburg *et al.* 2008, Maruyama *et al.* 2021).

In eukaryotes, genomic DNA is primarily packaged by histone proteins, which assemble DNA into nucleosome complexes (discussed further in section 2.3). This organisation is found in almost all species with a small number of apparent exceptions to this convention. Histones were originally thought to be missing from Dinoflagellates, which compact their DNA using dinoflagellate viral nucleoproteins (DVNPs) (Rizzo and Noodén 1972, Gornik *et al.* 2012), and *Nanochlorum eucaryotum* species of microscopic green algae (Wilhelm *et al.* 1982, Zahn 1984). Subsequent analysis has revealed *Nanochlorum eucaryotum* does in fact encode histones, and MNase digestion reveals a characteristic banding pattern indistinguishable from

that caused by nucleosomes in other eukaryotes (Soo and Warnecke 2021). In dinoflagellates, the situation is more complex. Recent work has shown that divergent histone proteins exist, but these are lowly expressed and their function is unknown (Gornik *et al.* 2012, Marinov and Lynch 2016). Outside these rare exceptions, histone-based chromatin is universal in eukaryotes.

Bacteria do not organise their genome into nucleosomes, but instead encode nucleoid associated proteins (NAPs) which differ across lineages in abundance and variety. NAPs are short DNA-binding proteins which can perform a variety of structural roles – from bending to bridging, and wrapping DNA (Dillon and Dorman 2010, Verma *et al.* 2019). This includes HU (heat-unstable protein) which can assemble into filamentous complexes on DNA at high concentrations (Rouvière-Yaniv and Kjeldgaard 1979, Van Noort *et al.* 2004), and Fis, which induces bending of the DNA (Skoko *et al.* 2006).

Archaea encode a diverse repertoire of DNA binding proteins, the composition of which varies between species (Hocher *et al.* 2021). Primarily, this includes eukaryotic-like histones in the Euryarchaea, as well as some bacteria-like, and archaea-specific, NAPs (Dillon and Dorman 2010, Hocher *et al.* 2021, Laursen *et al.* 2021). The presence of histones extends to the closest archaeal relatives to eukaryotes, the Asgard archaea, which encode for both histone proteins with charged N-terminal tails (Henneman *et al.* 2018) as well as other NAPs such as Alba (Laursen *et al.* 2021). Despite the evolutionary significance of histones in archaea, our understanding of the diversity is very limited, and pales in comparison to that in eukaryotes.

To understand the diversity of archaeal histones, we must first consider histone-based chromatin in eukaryotes which has been studied extensively. Below, I describe the structure of the eukaryotic nucleosome, and the mechanisms employed to alter its properties.

2.3. Histone-based chromatin in eukaryotes

Nucleosomes represent the fundamental, repeating unit of chromatin organisation in eukaryotes (McKnight and Miller 1976, Woodcock *et al.* 1976, Kornberg 1977). The nucleosome particle is an octameric complex comprising of two copies of each of the four core histones; H2A, H2B, H3 and H4 (Eickbush and Moudrianakis 1978). All core histones share a similar histone fold domain (HFD) structure consisting of three alpha helices separated by two loops, and unstructured, positively charged tail regions flanking either side of the HFD (Arents *et al.* 1991, Luger *et al.* 1997) (Figure 2.2).

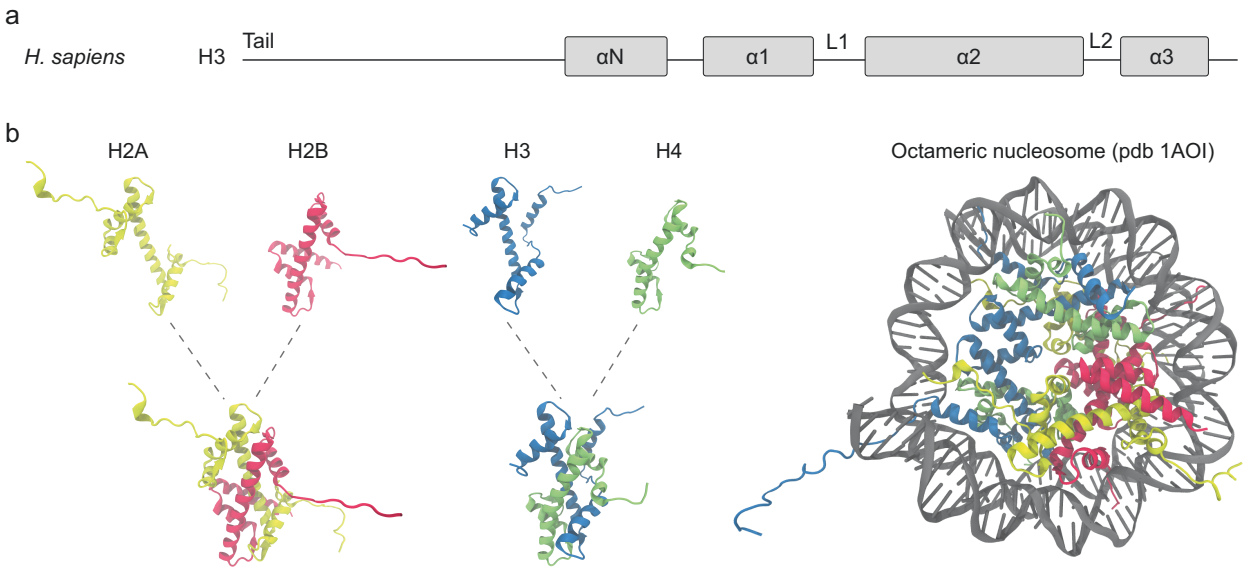


Figure 2.2 Structural organisation of the eukaryotic nucleosome.
a) Example of the secondary structure for eukaryotic histone H3 from Homo sapiens showing the three alpha helices (α1-3), two loop regions (L1 and L2) and tail region with extended N-terminal alpha helix (αN). b) The structure of histone monomers: H3, H4, H2A and H2B which form heterodimers (H2A+H2B, H3+H4). The structure of the octameric nucleosome (pdb 1aoi (Luger *et al.* 1997)) is shown on the right hand side.

Each histone monomer forms an obligate heterodimer (H3 + H4, H2B + H2A) with the corresponding dimer organised in an anti-parallel handshake configuration (Figure 2.2) (Luger *et al.* 1997). Dimerisation occurs through hydrogen bonding between proximal loop regions and hydrophobic interactions in the antiparallel a2 helices, where complementary shapes form between amino acid side chains at the dimer interface (Luger *et al.* 1997). These dimers further assemble into the octameric nucleosome structure, where a 'tetrasome' of two H3+H4 dimers is surrounded by two H2A+H2B dimers, one on either side (Figure 2.2) (Arents *et al.* 1991,

Luger *et al.* 1997). Proper assembly of the nucleosome requires that each histone occupies a distinct position in the complex (Arents *et al.* 1991, Luger *et al.* 1997).

Approximately 147bp DNA is wrapped around the histone octamer in a left-handed orientation (Kornberg 1974, Luger *et al.* 1997) and non-specific DNA binding is mediated primarily through positively charged regions (Arents *et al.* 1991, Luger *et al.* 1997). The formation of nucleosomes on non-specific DNA is consistent across diverse eukaryotes despite significant sequence differences, although the stability and DNA affinity of nucleosomes can vary widely for the same DNA sequence (Patwal *et al.* 2021).

2.3.1. Dynamic changes to chromatin structure

The nucleosome structure described above is static, restricting access for many essential proteins, such as DNA repair enzymes and transcription factors (Li *et al.* 2007). However, in reality, the nucleosome is highly dynamic, and subject to a plethora of dynamic changes which can alter genome accessibility, allowing cells to respond to upstream signals. This ability to modulate chromatin is not exclusive to eukaryotic histone-based chromatin. In bacteria, the composition and architecture of chromatin is also subject to change. For both eukaryotes and bacteria, we have a reasonable understanding of the mechanisms behind the dynamic properties of chromatin. This includes modifications to the protein itself, as well as local or global exchange of DNA-binding proteins with different structural properties

In bacteria, chromatin is constantly remodelled. NAPs with different structural properties are assimilated into or removed from chromatin at varying growth phases or in different conditions, providing a mechanism to regulate gene expression at different stages of growth (Dorman 2013). For example Dps in *E. coli* is almost undetectable in exponential phase, but forms a major constituent of the nucleoid in stationary phase, mediating a compact chromatin state (Ali Azam *et al.* 1999).

In eukaryotes, the primary drivers of change come from post-translational modifications (PTMs), and remodelling events (Becker and Hörz 2002, Clapier *et al.* 2017),

The relative availability of DNA contained in the nucleosome can be altered through a vast array of PTMs (Zhao and Garcia 2015), some of which show species-specificity (Garcia *et al.*

2007) and can be added or removed to all four core histones via histone-modifying enzymes. Often, PTMs occur on the histone tails, but there is also evidence of extensive modification in the globular domain region (Tessarz and Kouzarides 2014). There are many different types of PTMs, such as methylation, acetylation and sumoylation which can alter nucleosome stability, dynamics or DNA binding (Bannister and Kouzarides 2011). The presence of different PTMs provide the possibility for highly tuned regulation of the chromatin architecture. This can occur directly, through the presence of the PTM itself, or indirectly, by promoting differential interactions with other proteins (Bannister and Kouzarides 2011). Adding to this complexity, different modifications at the same site can have divergent effects. For example, methylation of lysine K9 in the N-terminal tail of H3 is associated with compact heterochromatin (Rea *et al.* 2000, Nakayama *et al.* 2001, Rivera *et al.* 2014), whereas H3 with acetylation at this site is involved in transcription elongation, and found predominantly in promoters (Agalioti *et al.* 2002, Wang *et al.* 2008, Gates *et al.* 2017). PTMs therefore represent an important and complex combinatorial method which can alter the properties of the nucleosome.

ATP-dependent chromatin remodellers represent another regulatory mechanism employed to alter eukaryotic nucleosomes, and in turn chromatin. They perform a variety of roles to disassemble, alter, and reassemble the nucleosome. This includes nucleosome sliding and eviction from the DNA, locally rendering the DNA more accessible (Becker and Hörz 2002, Clapier *et al.* 2017). One important remodelling event includes the exchange of replication-dependent histones for histone variants.

2.4. Histone variants in eukaryotes

The addition and removal of histone variants to and from the nucleosome can alter the structural properties of the nucleosome (Figure 2.3b). The incorporation of variants into the nucleosome can serve to increase or decrease chromatin accessibility. Variants can act as a specific interface for the recruitment of distinct chromatin associated proteins, or directly change nucleosome dynamics. This helps regulate DNA-based processes and allows cells to respond to a wide range of environmental changes (Talbert and Henikoff 2014).

Histone variants differ in their primary sequence from canonical histones, and occasionally are fused with additional domains (Figure 2.3a) (Malik and Henikoff 2003). Variants of all four core histones; H2A, H2B, H3 and H4, have been identified, although histone H4 appears to be

considerably more constrained. Indeed, H4 is very highly conserved throughout eukaryotes, with the exception of H4 from *Leishmania* spp., which have highly divergent sequences (Lukeš and Maslov 2000). It was previously thought that H4 had no functionally distinct variants (Thatcher and Gorovsky 1994, Kamakaka and Biggins 2005). However, recent work identified a Hominidae specific variant H4G, implicated in breast cancer (Long *et al.* 2019).

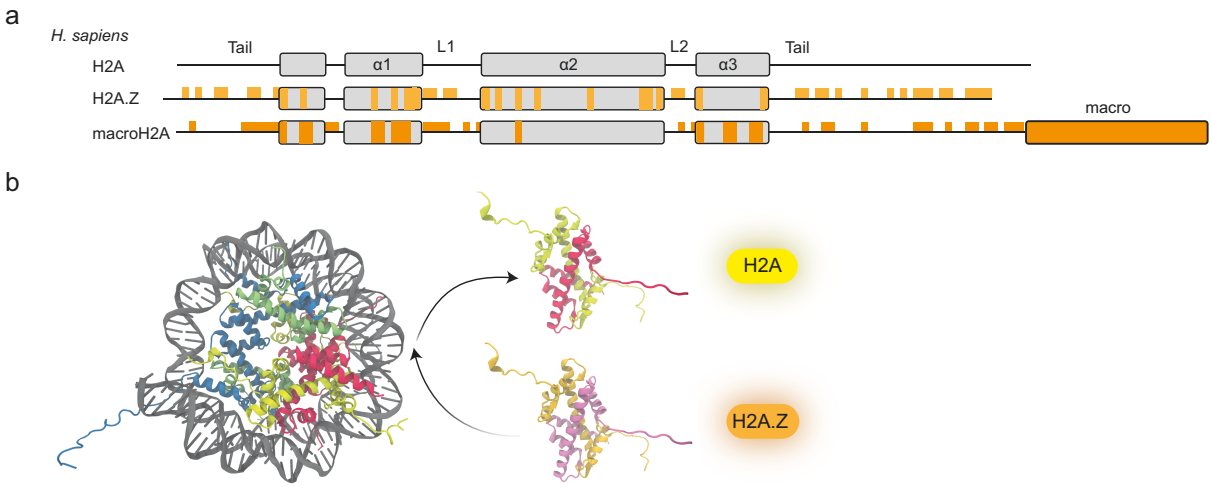


Figure 2.3 Histone variants increase the complexity of the nucleosome structure.
a) Example of the secondary structure for eukaryotic histone H2A from Homo sapiens and variants H2A.Z and macroH2A, showing the three alpha helices (α1-3), two loop regions (L1 and L2) and tail region. For each variant, the positions of substitutions relative to the replication-dependent canonical H2A is highlighted in coloured blocks. The extended macro domain of macroH2A is annotated. b) Dynamic exchange of replication-dependent canonical histone H2A for replication-independent variant H2A.X. PDB 1aoi (Luger *et al.* 1997). H2A = yellow, H2B = red, H3 = blue, H4 = green.

One major difference between core histones and histone variants is their genomic location and the timing of expression. In metazoan, all four core histones are often found clustered together in a tandem array of genes, whereas histone variants are generally transcribed as single genes elsewhere in the genome (Marzluff and Duronio 2002, Malik and Henikoff 2003). The expression of core histone genes is tightly coupled to replication and are distributed evenly, whereas histone variants are typically deposited into nucleosomes independently of replication (Ahmad and Henikoff 2002, Marzluff and Duronio 2002). The relative abundance of each histone variant is not fixed, and can vary, for example over different stages of cell cycle and embryonic development (Martire and Banaszynski 2020).

The deposition of each variant at specific times and to targeted locations on the genome requires dedicated, co-evolved chaperones (Mattiroli *et al.* 2015). Regulation of this process is essential. Deregulation of the expression of histone variants, or changes to their dedicated

machinery frequently contribute to various cancer types, highlighting the importance of tight control of histone variants (Buschbeck and Hake 2017).

2.4.1. The importance of substitutions and more divergent sequence changes

Some histone variants differ from canonical histones by just a few amino acid changes. Despite these small differences, nucleosomes containing these variants exhibit distinct functions. The H3 variant H3.3, for example, contains only four substitutions relative to canonical H3 and replaces nucleosomes evicted during transcription, therefore showing an enrichment in active regions of the genome (Ahmad and Henikoff 2002, Mito *et al.* 2005). H3.3-containing nucleosomes are sensitive to salt dependent disruption, suggesting that less energy may be needed to remove these from the DNA, possibly helping to increase DNA accessibility – or making for more movable roadblocks – in active regions (Jin and Felsenfeld 2007).

The function of specific variants can occasionally be attributed to unique substitutions at PTM sites. Some PTMs occur on a residue only found in a specific variant and play an important part in their role in the cell (Corujo and Buschbeck 2018). For example, H2A.X contains a unique C- terminal SGE/Dϕ motif (where ϕ is a hydrophobic amino acid) which is phosphorylated at the serine in response to double stranded break (Redon *et al.* 2002). This phosphorylation initiates recruitment of a cascade of DNA repair proteins to the site (Scully and Xie 2013). Similarly, one single substitution in the N-terminal tail from an alanine to serine provides a site for phosphorylation in H3.3, stimulating transcription (Armache *et al.* 2020).

Of course, not all amino acid changes need to result in distinct functions. For example, a recent study suggested that the functions of the H2A variant HIS-35 (with a one amino acid substitution) in nematodes may be identical to those of canonical H2A (Sing *et al.* 2022).

In contrast, some histone variants are substantially divergent from their canonical counterparts. This includes variants such as macroH2A, which contains a unique macro domain not found in other H2As (Figure 2.3a). Macro domains bind metabolites of NAD+ (Han *et al.* 2011), a feature seen for one macroH2A isoform (marcoH2A1.1) (Kozlowski *et al.* 2018). The histone domain of macroH2A also shows unique properties. Alone – without the macro domain – the HFD increases salt-dependent stability (Chakravarthy and Luger 2006) as well as reducing

transcription *in vivo* (Nusinow *et al.* 2007). Other histone variants have high sequence diversity, such as centromeric H3 (CENP-A). This H3 variant localises to centromeric regions and establishes centromere identity (Palmer *et al.* 1987, 1991). Compared to canonical H3, nucleosomes containing centromeric H3 can be found as stable tetramers (Dalal *et al.* 2007), and are more susceptible to MNase digestion (Takahashi *et al.* 1992, Krassovsky *et al.* 2012).

The plethora of histone variants with both small and large divergences from core histones highlights the diversity of histone proteins in eukaryotes. The set of histone variants is not equal among species, raising questions about how and when each variant arose.

2.4.2. The evolution of histone variants in eukaryotes

The evolutionary histories of histone variants in eukaryotes can be elucidated using phylogenetic approaches (Molaro and Drinnenberg 2018). Some variants, such as H2A.Z, are highly conserved, evolved once, early on in eukaryotic evolution, and are found in diverse eukaryotic species; from humans to yeast, and even distantly related protists such as *Tetrahymena* (Thatcher and Gorovsky 1994, Jackson and Gorovsky 2000, Redon *et al.* 2002). H2A.Z has been shown to play a variety of roles, both in transcription activation and repression (Zlatanova and Thakar 2008, Giaimo *et al.* 2019). One example is its involvement in the thermosensory response of plants. At increased temperatures, H2A.Z occupancy at genes upregulated by temperature decreases, suggesting a role in limiting expression upon high temperature change (Kumar and Wigge 2010).

Other variants also arose once, but are much more divergent. This includes centromeric H3 with its particularly fast-evolving, diverse tails that vary both in sequence and in length (Malik and Henikoff 2003). This divergent variant is not universal, and is absent in some lineages. In particular, a transition to holocentricity in some insects is correlated with the loss of centromeric H3 from certain lineages (Drinnenberg *et al.* 2014, Senaratne *et al.* 2022).

Yet other histone variants are lineage specific, and show a narrow phyletic pattern with specialised functions. H2A.B, for example is only found in mammalian species (Talbert and Henikoff 2010, Talbert *et al.* 2012), whereas H2A.W is exclusively found in plants (Talbert *et al.* 2012). A systematic study of H2B variants found highly divergent variants throughout

mammalian evolution, with multiple gain/loss events (Raman *et al.* 2022). These variants are subject to purifying selection but evolve faster than canonical H2B (Raman *et al.* 2022).

Finally, some histone variants and their canonical counterparts have diverged from one another in multiple independent events. This includes H2A.X and canonical H2A, which appear to have arisen from one another independently several times by either gain or loss of the H2A.X motif (Thatcher and Gorovsky 1994, Malik and Henikoff 2003). A similar pattern is seen for H3.3 and canonical H3 (Thatcher *et al.* 1994, Malik and Henikoff 2003). H3.3 is not strictly an example of convergent evolution, as the amino acid substitutions are not always conserved, but there are other cases of true convergent evolution in other variants. For example, the KSPK motif convergently evolved in H2A variants macroH2A and H2A.W (Talbert *et al.* 2012). Wild type H2A.W displays greater protection against MNase, and a larger protected fragment compared to H2A.W lacking the motif, indicative of greater interaction with linker DNA (Lei *et al.* 2021). When inserting H2As lacking this motif into *Schizosaccharomyces pombe*, changes to heterochromatin occur suggesting the gain of this motif resulted in a convergent role in heterochromatin in several variants (Lei *et al.* 2021).

The weight of evidence above highlights that histone paralogs in eukaryotes do different things. Different evolutionary patterns of histone variants in eukaryotes suggests retention of distinct paralogs. This is underscored in many instances by distinct functional properties of each variant, which are generally well characterised in eukaryotes.

Comparatively, we know very little about the evolutionary patterns of archaeal histones. Below, I will first outline the structure of histones in archaea, before weighing up the current evidence for histone paralogs playing distinct roles.

2.5. Histones in archaea: an overview

Histone-fold proteins were first identified in *Methanothermus fervidus* by members of the Reeve lab (Sandman *et al.* 1990). Incubation of purified *M. fervidus* histones (HMfA and HMfB) with linearised plasmid DNA showed binding and compaction (Sandman *et al.* 1990). At low concentrations, HMf forms kinks on the DNA, and spherical-like higher order complexes at higher concentrations (Sandman *et al.* 1990). The structure of the HMfB dimer from *M. fervidus*, solved using NMR, showed this was indeed a histone, with similarities in structure to eukaryotic histone H3-H4 and H2A-H2B dimers (Figure 2.4c, d) (Starich *et al.* 1996) AFM imaging demonstrated that *M. fervidus* histones can form beaded fibres along DNA *in vitro*, reminiscent of those formed by eukaryotic histones (Tomschik *et al.* 2001).

Since the discovery of histones in *M. fervidus*, sequencing efforts have identified histone proteins in many archaea, along with other NAPs (Henneman *et al.* 2018, Hocher *et al.* 2021, Laursen *et al.* 2021). In a few clades, such as the Diaforarchaea, histones appear to have been lost (as well as in some members of the Crenarchaeota) and, in some cases, histones may have been functionally replaced by other NAPs (Adam *et al.* 2017, Hocher *et al.* 2021, Laursen *et al.* 2021). For example in the Diaforarchaea species *Thermoplasma acidophilum*, a HU-analog (HTa) has properties reminiscent of histones, including high abundance, protection from MNase digestion, and a preference for GC-rich sequences (Hocher *et al.* 2019).

There is non-uniformity of histone functions in different species of archaea. At least in some archaea, histones are highly expressed (Hocher *et al.* 2021) and found in the nucleoid where they bind to and wrap DNA, constraining DNA in a supercoiled state (Mattiroli *et al.* 2017). In members of the Thermococcales and Methanobacteriales, histones are some of the most abundant proteins and are copious enough to coat the entirety of the genome (Pereira *et al.* 1997, Rojec *et al.* 2019, Sanders, Lammers, *et al.* 2019, Hocher *et al.* 2021). Contrastingly, histones in the Halobacteria are much less abundant and bind to discrete locations in the genome (see section 2.8.1.1). In some systems, particularly those with multiple histones, presence of at least one histone is essential. For example, in *Methanococcus voltae* (Heinicke *et al.* 2004) and *Thermococcus kodakarensis* (Čuboňová *et al.* 2012) deletion of all paralogs has not been possible. However, this is not always the case. For example, the one histone in *Methanosarcina mazei Gö1* does not appear to be essential for cell viability (Weidenbach *et*

al. 2008). Deletion of the histone is not toxic, but reduces growth, increases UV sensitivity and reduces transcription (Weidenbach *et al.* 2008).

No histone chaperones have been identified in archaea, marking a substantial deviation in the potential for regulated deposition of histones from the situation in eukaryotes (Grau-Bové *et al.* 2022).

2.6. Structure of archaeal histones

The structure of archaeal histones is highly similar to that of eukaryotic histones (Starich *et al.* 1996) (Figure 2.4a). The primary structure of archaeal histones consists of a HFD, with three alpha helices separated by loop regions (Arents and Moudrianakis 1995, Decanniere *et al.* 2000) (Figure 2.4a, b). Typically, archaeal histones lack the long tail extensions found in eukaryotic histones, except for in a few species e.g. *Heimdallarchaeota* spp. (Henneman *et al.* 2018) and *Methanocaldococcus jannaschii* (Wilkinson *et al.* 2010) (Figure 2.4b).

Monomeric archaeal histones spontaneously dimerise in a handshake configuration where each interacts in an anti-parallel arrangement, similar in structure to eukaryotic histones (Figure 2.4c, d). At the dimer interface, hydrophobic residues in the alpha helices from each monomer form complementary shapes, and inter-monomer hydrogen bonds form between the contacting loop regions (Arents and Moudrianakis 1995, Decanniere *et al.* 2000).

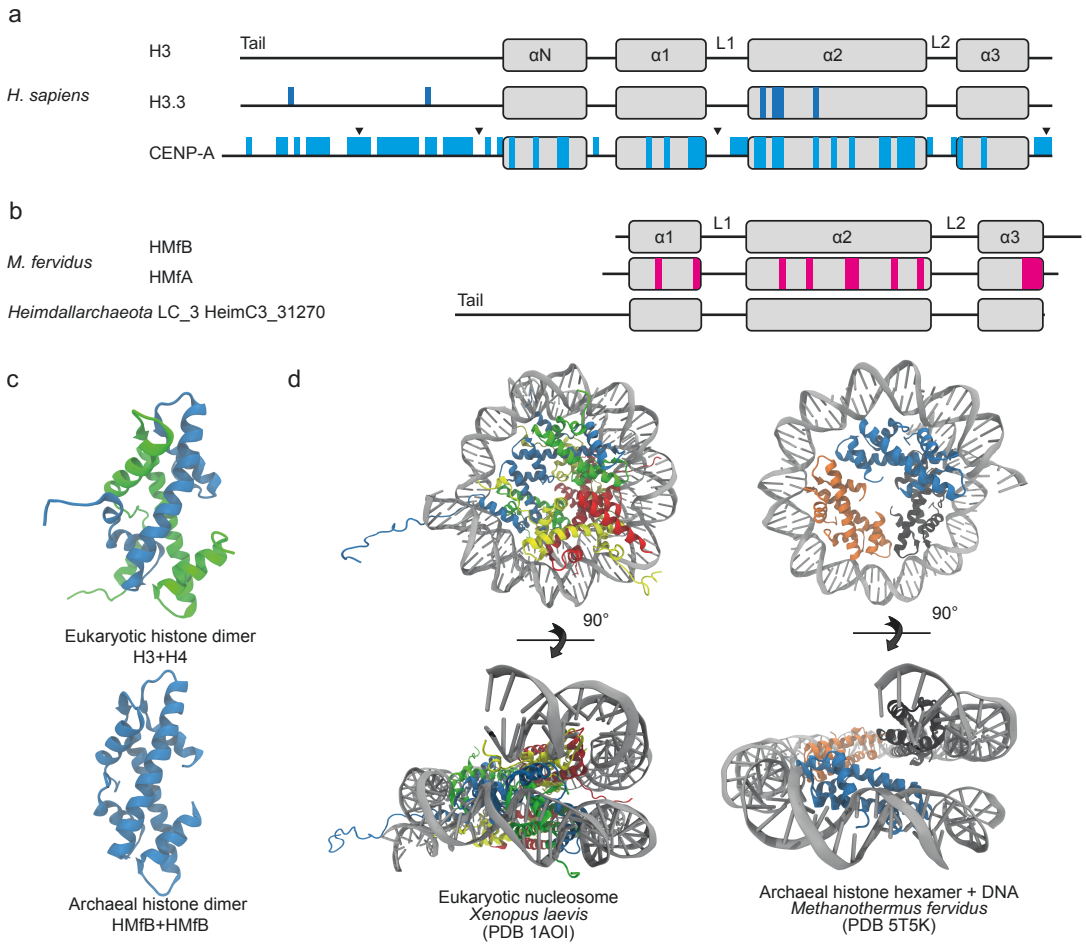


Figure 2.4 Comparison of archaeal and eukaryotic histones.
a) Secondary structure and sequence variation between H. sapiens H3 and histone variants H3.3 and CenH3 from HistoneDB (Draizen *et al.* 2016). Sites which differ from canonical H3 are highlighted as coloured blocks on the secondary structure. Tail regions, alpha helices (αN-α3) and loop regions (L1, L2) are labelled. Insertions are marked with an arrow. b) Comparison of histones HMfA and HMfB from M. fervidus showing sequence differences as in (a). An example of a tailed histone from Heimdallarchaeota LC_3 is shown underneath. c) Structure of the eukaryotic H4+H3 dimer (PDB 1aoi (Luger *et al.* 1997)) and archaeal histone HMfB (PDB=5t5k, (Mattiroli *et al.* 2017)). H2A = yellow, H2B = red, H3 = blue, H4 = green. d) Structure of the eukaryotic octameric nucleosome (left) and the archaeal histone HMfB dimer and hexameric complex with DNA (right). This figure has been adapted from (Stevens and Warnecke 2022).

At least in some species, this dimerisation is more promiscuous than that of eukaryotes, allowing the formation of both homo- and hetero-dimers. For archaea with only one copy of a histone, self-dimerisation reflects a necessity. Direct evidence that homo- and hetero-dimerisation occurs *in vivo* comes from *M. fervidus*, where HMf isolates from cells contain species of both homodimers and HMfA-HMfB heterodimers shown through 2D electrophoresis (Sandman *et al.* 1994).

This flexibility has also been documented extensively for histones from other species *in vitro*. Histones from *M. fervidus* (Sandman *et al.* 1994, Starich *et al.* 1996, Decanniere *et al.* 2000, Mattiroli *et al.* 2017), *T. kodakarensis* (Higashibata *et al.* 1999), *Pyrococcus horikoshii* OT3 (Weng *et al.* 2004) and *Methanobacterium formicicum* (Li *et al.* 1998, Zhu *et al.* 1998) can all form homodimers. Experimental work also showed, through *in vitro* unfolding transitions, that *Pyrococcus* GB-3a (Li *et al.* 1998), *Archaeoglobus fulgidus* and *M. jannaschii* (Li, Sandman, *et al.* 2000) all encode histones which homodimerize. This flexibility is apparent even across species (Li, Sandman, *et al.* 2000). The recombinant *M. jannaschii* histone MJ1647 (which has an extended C-terminal tail) and HMfB from *M. fervidus* form stable dimers (Li, Sandman, *et al.* 2000). Together, this shows that homo- and hetero-dimerisation is likely widespread in archaea, reflecting a divergence from eukaryotic histones.

Histone dimers further assemble into tetramers, which can bind to and wrap DNA and protect from MNase digestion. These tetramers are very similar in structure to the eukaryotic tetrasome (Figure 2.4c, d) (Mattiroli *et al.* 2017). Archaeal histones appear to bind non-specifically, but preferentially, to bendable DNA sequences (Grayling *et al.* 1997, Bailey *et al.* 2000, Nalabothula *et al.* 2013, Rojec *et al.* 2019). Similar to eukaryotic histones, DNA binding is mediated by residues in the loop regions (Soares *et al.* 2000) with each histone dimer binding to and wrapping 30 bp DNA (Grayling *et al.* 1997).

Unlike the eukaryotic octameric structure, some archaeal histones form higher-order extended oligomeric structures (also known as ‘hypernucleosomes’ or ‘archaeosomes’) (Figure 2.5) (Mattiroli *et al.* 2017, Henneman *et al.* 2018, Bowerman *et al.* 2021). The potential for higher order formation was first postulated in 1995 (Arents and Moudrianakis 1995).

This organisation of histone dimers in a spiral stacking arrangement was revealed through the crystal structure of the HMfB hexamer (Mattiroli *et al.* 2017), and was also seen for *T. kodakarensis* histone HTkA in Cryo-EM structures (Bowerman *et al.* 2021). Hypernucleosome formation has been documented *in vivo* in *M. fervidus* (Rojec *et al.* 2019) and *T. kodakarensis* (Maruyama *et al.* 2013, Mattiroli *et al.* 2017, Sanders *et al.* 2021), and they likely also form in other species (Henneman *et al.* 2018). These structures are likely to be dynamic in nature. Cryo-EM structures and molecular dynamics simulations revealed that oligomeric archaeal histone-DNA complexes containing HTkA show breathing motions, where

the stacked dimers open up to expose both the DNA and the stacking interface of the histones (Bowerman *et al.* 2021).

The generation of these higher order oligomers occurs in 30 bp step-wise increments through the continuous addition of dimers (Figure 2.5). Archaeal histones can therefore form complexes of varying sizes - protecting 60 bp DNA as a tetramer, 90 bp as a hexamer and as much as 450 bp DNA as complexes composed of 15 dimers (Nalabothula *et al.* 2013, Mattiroli *et al.* 2017, Sanders *et al.* 2021). Most oligomeric complexes formed *in vivo* are slightly smaller, consisting of around four to ten histones. This is a marked difference from the canonical eukaryotic octameric nucleosome.

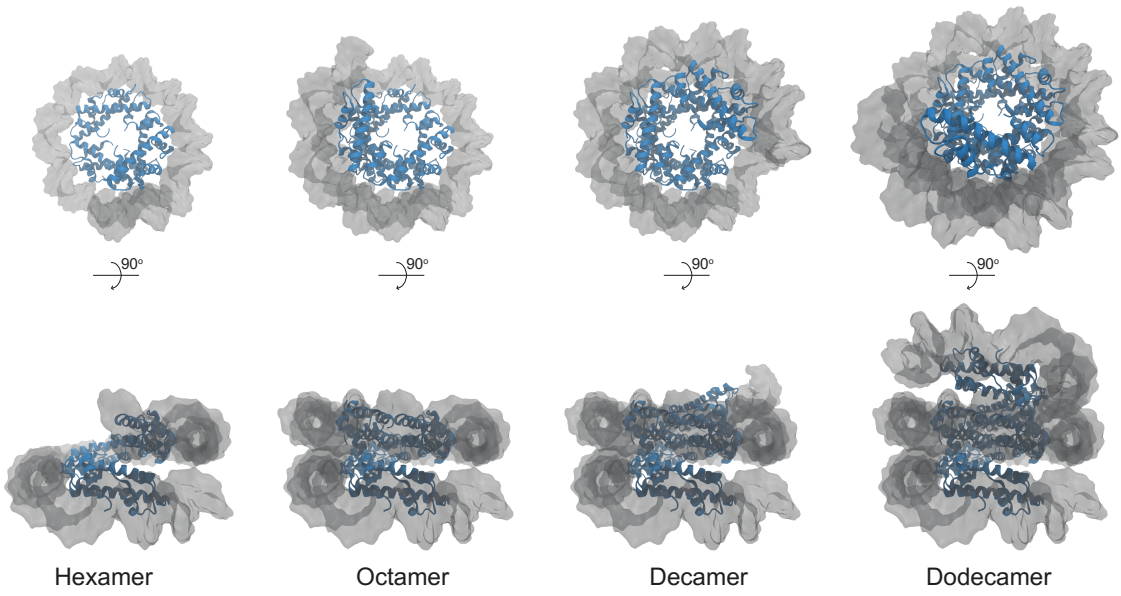


Figure 2.5 Model of oligomeric histone complexes in archaea.
HMfB dimers are sequentially added to a hexameric structure, to form a dodecamer and intermediates. Model is based on the HMfB hexamer (PDB 5t5k (Mattiroli *et al.* 2017)).

2.7. Histones in context

Histones are not in an isolated system with DNA alone, and it is important to consider the context within which archaeal and eukaryotic histones sit. Essential DNA repair proteins, transcription machinery, and other DNA-associated proteins, all require access to the DNA. This provides the potential for interaction between these DNA binding proteins and histone-DNA complexes. What do we know about the impact of histones on transcription and the presence of interacting proteins?

In general, eukaryotic histones pose a barrier to transcription machinery which can be altered by the incorporation of different variants, addition of PTMs, or by remodelling events (see section 2.3-2.4).

The binding of archaeal histones to DNA can also, in theory, pose a barrier to proteins that need to interact with and act on DNA. Archaeal histones can repress transcription (Wilkinson *et al.* 2010), inhibit strand displacement by DNA polymerase (Sun and Huang 2016) and protect DNA from Argonaute-mediated degradation (Zander *et al.* 2017). However, these hurdles appear less steep than they are in eukaryotes; histones from *Methanothermobacter thermautotrophicus* do not appear to pose a complete barrier to transcription after initiation has started in an *in vitro* system, slowing down rather than stopping elongation (Xie and Reeve 2004). Furthermore, transcription factors can outcompete *M. jannaschii* histones for Upstream Activating Sequences (UAS) (Wilkinson *et al.* 2010). MCM helicases can also bypass histones to unwind DNA (Shin *et al.* 2007). Introduction of the archaeal histones from *M. fervidus* to *E. coli* only has mild phenotypic effects but can locally repress transcription (a feature correlated with larger oligomers) (Rojec *et al.* 2019) Although weaker, these barriers are reminiscent of the transcription repression imposed by eukaryotic nucleosomes, which can be altered by different modifications/variants.

The inhibitory effect on transcription *in vitro* for histones from *M. jannaschii* is only observed at very high histone-DNA ratios which may not necessarily be observed in vivo (Wilkinson *et al.* 2010). Similar effects can be seen for complex size. Positioning a high-affinity DNA sequence downstream of the site for transcription initiation, and measuring the relative changes to transcription upon addition of histones (Xie and Reeve 2004). At lower concentrations, *M. thermautotrophicus* histone HMtA2 binds downstream of the promoter and slows down transcription (Xie and Reeve 2004). Increasing the ratio of *M. thermautotrophicus* histone HMtA2 to DNA resulted in longer oligomers which bind over the promoter region, resulting in reduced transcription *in vitro* (Xie and Reeve 2004).

What about interaction with different proteins? Much is known about the interaction of components of the eukaryotic nucleosomes with other proteins, which can be specific for unique variants (such as chaperones, see section 2.4). Contrastingly, there is currently no extensive evidence for proteins which directly interact with archaeal histones. However, it is

safe to assume that archaeal histones are not an isolated system and frequently come into contact with many other proteins. Similar to NAPs in bacteria, archaeal histones are often expressed alongside other chromatin-associated proteins (Luijsterburg *et al.* 2008, Henneman *et al.* 2018, Maruyama *et al.* 2021). Some of these NAPs are widespread and highly abundant, such as Alba, whereas others are specific to certain lineages, such as Sul7 and transcription regulator of the maltose system-like 2 (TrmBL2) (Driessen and Dame 2011, Maruyama *et al.* 2011, Henneman *et al.* 2018). The mechanism of interactions between these different NAPs and histones is largely unknown; however, the interplay between NAPs does seem to alter chromatin structure

In *Sulfolobus solfataricus*, which lacks histones, NAPs Cren7 and Alba mediate different chromatin structures (Alba results in more compact structures), and combinations of both *in vitro* reduce higher-order structures leading to less compact chromatin (Cajili and Prieto 2022). Addition of *T. kodakarensis* histone to Alba and DNA *in vitro* resulted in the formation of different higher-order complexes, with less compact structures observed as the ratio of Alba to histone increased (Maruyama *et al.* 2020).

The combination of different NAPs therefore represents an important potential source of chromatin regulation in archaea which differs among species. In *T. kodakarensis*, histones compete with Alba for DNA binding whereas in *T. acidophilum*, a HU homolog (HTa) facilitates DNA compaction with Alba suggesting different roles of Alba between species (Maruyama *et al.* 2020). But can histones themselves, within species, provide the opportunity for modification of chromatin structure in archaea?

2.8. The case for histone variants in archaea

The number of histone paralogs varies greatly across species archaea, from zero to 11 proteins with a single HFD, and zero to one doublet HFD proteins (see section 2.8.1.1, Figure 2.6). In some lineages, such as the Asgard archaea and Methanobacteria, histone paralog number has expanded dramatically (Figure 2.6). Within species, histones can differ significantly from one another. For example, two of the seven histones in *Methanosphaera stadtmanae* are only 41% identical at the amino acid level. This variable paralog number provides scope for functional variation.

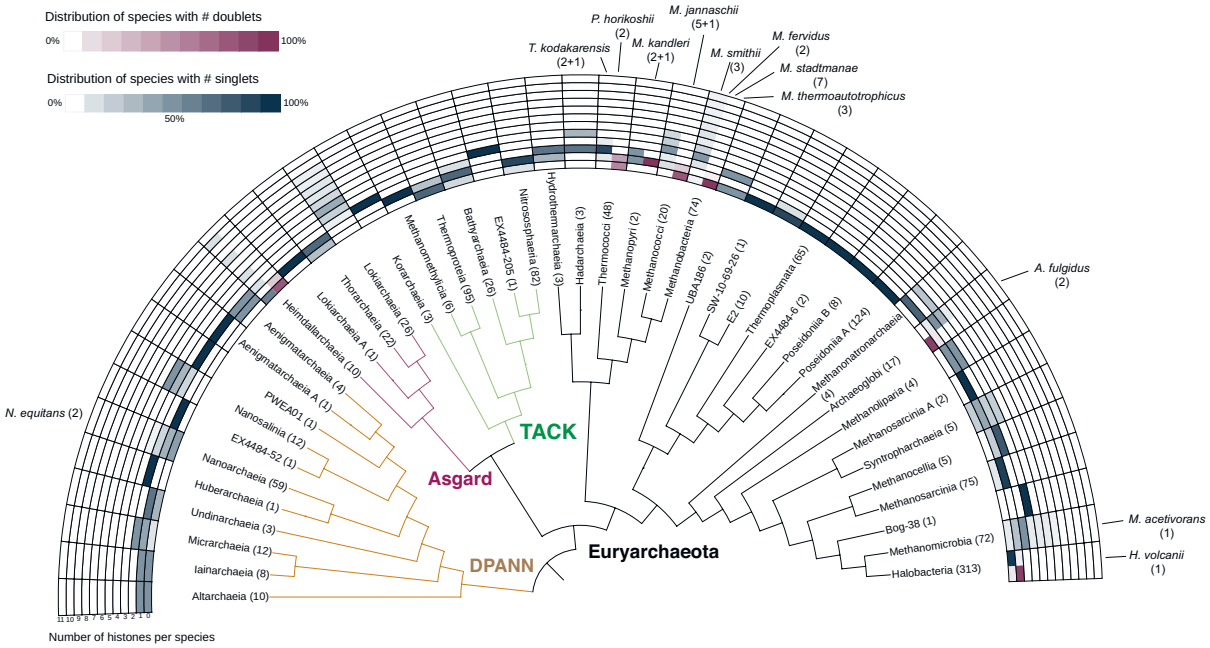


Figure 2.6 Number of histones per species in different lineages of archaea.
A cladogram of archaeal classes from GTDB is shown, rooted at the split between DPANN and Euryarchaeota/Asgard/TACK. Lokiarchaea_A contains Odinarchaeota. Seven classes were removed from the cladogram as they were missing from the dataset. Histones were obtained from primary publication (Hocher *et al.* 2021) and supplemented with histones from additional Asgard genomes form primary sources (Liu *et al.* 2021) and from GenBank (https://www.ncbi.nlm.nih.gov/assembly) using taxid 1935183 on 15-06-21 (see Methods section 10.2.5). The number after the class tip label indicates the number of genomes included in the analysis for that class. For each lineage, the heatmap indicates the proportion of species within that lineage grouped by the number of histones (the number of histones increases from 0 to 11 moving outwards from the tree, and darker colours indicate a higher proportion of species encoding a particular number of histones in that lineage) for doublet (red) and singlet (blue) HFD proteins (see Methods). Species of interest with characterised histones are highlighted. The number after the species name in brackets shows the number of singlets (first number) and doublets (second number) if both are encoded in the genome. In some cases, species with zero histones may reflect missing information (incomplete genome assembly) rather than true absence of histones. This figure has been adapted from (Stevens and Warnecke 2022).

In principle, an expansion in the number of histones could result in homologs conferring different structural properties, and performing specific roles in the cell. In the absence of wide-spread PTMs and long tail extensions, in order for complex histone-based chromatin to exist in archaea, sequence differences between paralogs are likely key for any diversity in histone-based chromatin. There is some evidence to support this hypothesis, much of which comes from species with two histones. Sections 2.8.1 – 2.8.4 review the current evidence for variation between histones within species of archaea.

2.8.1. Histones which vary in their structural and DNA binding properties

2.8.1.1. Structural deviants: Doublet histones

Some archaea encode a 'doublet' histone protein, which consists of two histone fold domains fused together (Figure 2.7a) (Slesarev *et al.* 1998, Ammar *et al.* 2012, Dulmage *et al.* 2015). Doublet histones are found in many species, including *Methanopyrus kandleri*, for which the crystal structure without DNA has been solved (Figure 2.7b) (Fahrner *et al.* 2001), as well as many halophilic archaea (Sakrikar and Schmid 2021).

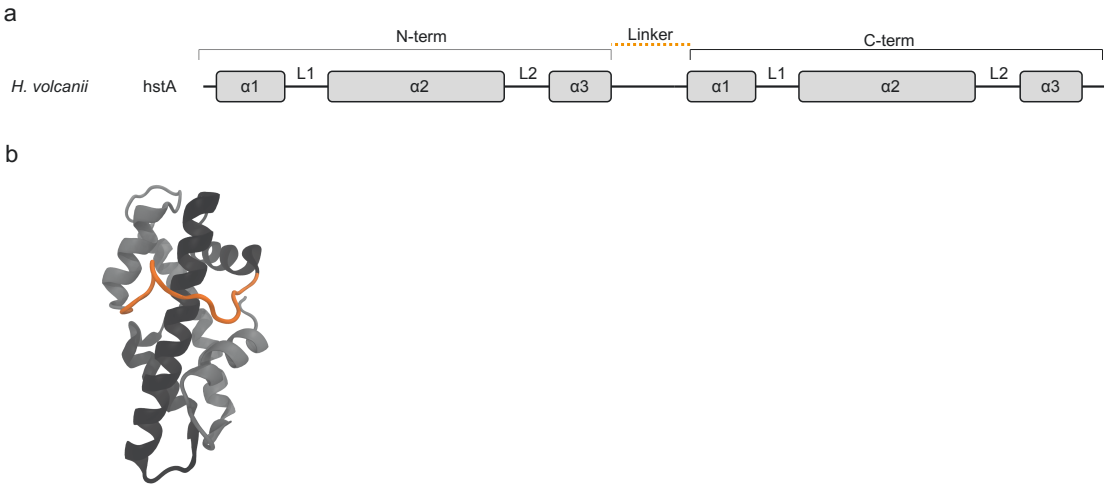


Figure 2.7 Structure of archaeal histones with duplication of the HFD.
a) Secondary structure of the doublet histone hstA from H. volcanii. f) Crystal structure of the doublet histone HMk from M. kandleri (PDB = 1f1e (Fahrner *et al.* 2001)). The N-terminal HFD of the doublet is shown in grey, C-terminal HFD in black, and linker in orange.

Isolation of chromatin from *Halobacterium salinarum* appeared to show nucleosome-like structures, with a higher proportion of these structures in stationary phase compared to exponential phase (Takayanagi *et al.* 1992). In addition, MNase digestion of *H. volcanii* chromatin revealed protected fragments similar to those observed in eukaryotic chromatin, with the doublet histone suggested as the causative agent (Ammar *et al.* 2012). However, efforts to reproduce protected fragments in *H. salinarum* and *H. volcanii* was unsuccessful, and the driver of this pattern – if it exists – remains elusive, as histones are not the only protein that can trigger protection and are not highly expressed (Dulmage 2015, Hocher *et al.* 2021, Sakrikar *et al.* 2022).

Further studies have shown that the histone doublet does not have a major a role in genome compaction in halophiles (Sakrikar and Schmid 2021). In *H. salinarum*, the doublet histone is not essential, and deletion results in subtle changes to cell morphology and gene expression (Dulmage *et al.* 2015, Sakrikar and Schmid 2021). The doublet in this species instead appears to function as a transcription factor to globally activate genes and, in response to low salt conditions, binds to discrete genomic regions to increase ion uptake (Sakrikar and Schmid 2021). Therefore, doublet histones in the halophilic archaea are unlikely to play a key role in modulating chromatin architecture.

2.8.1.2. Structural deviants: obligate heterodimers

In a few cases, histones in archaea appear to have lost the ability to function as a homodimer, and require heterodimerisation for DNA binding. In *Nanoarchaeum equitans*, one of the two histones binds DNA only as a heterodimer (Friedrich-Jahn *et al.* 2009). Similarly, in *Methanobacterium thermautotrophicus* OSU, one out of three histones (HMtB) can only bind DNA when assembled as a heterodimer with HMtA2 and not as a homodimer (Sandman *et al.* 2008). In both cases, these differences are due to small sequence changes: a single amino acid substitution at a DNA binding residue in *M. thermautotrophicus* (Sandman *et al.* 2008), and a small insertion in the L1 region (similar to that seen in eukaryotic H3) in *N. equitans* (Friedrich-Jahn *et al.* 2009). It is worth noting that these apparent cases of heterodimerisation are different from that of eukaryotic histones. Whereas eukaryotic histones form obligate heterodimers, homodimers of these archaeal histone are still able to form, but appear to no longer function as DNA binders, at least *in vitro*.

2.8.1.3. Structural deviants: DNA binding

Another key way in which archaeal histones could differ from one another is through the strength of DNA binding. DNA binding strength (affinity) and effects on DNA compaction of archaeal histones have been investigated for a number of archaeal histones using recombinant preparations. This has often involved the use of electrophoretic mobility shift assays (EMSAs). In this set-up, compaction dictates relative migration and more compact histone-DNA

complexes migrate faster than those with less compact packaging (Sandman *et al.* 1990, Soares *et al.* 2000).

Differences in DNA binding affinity between histones has been shown for several species. Complexes of *M. fervidus* histone HMfB with linear plasmid DNA migrate more quickly than those with HMfA, indicating higher compaction imposed by HMfB (Sandman *et al.* 1994). Tethered particle motion experiments confirm this difference, and show higher stacking energies and higher stability of oligomers containing HMfB compared to HMfA (Henneman *et al.* 2021). *T. kodakarensis* has two histones (HTkA and HTkB) which also show differences in DNA binding affinities (Higashibata *et al.* 1999). Similarly, one histone in *M. jannaschii* (HMjA4) poses less of a barrier to transcription *in vitro*, and has weaker DNA binding than other histones in this species (Wilkinson *et al.* 2010). The differences in DNA binding between histones can occur despite high sequence similarity. HMfA and B, for example are 84% identical at the sequence level (Figure 2.4b).

2.8.2. Deletion studies

Further evidence that the structural differences between histones may be functionally relevant comes from approaches to delete the histone of interest. In some multi-histone species, deletion of either histone results in divergent phenotypes. This is true in *Methanococcus voltae* and *T. kodakarensis* where histones hstA and HTkB, respectively, reduce the growth rate of cells (Heinicke *et al.* 2004) (Čuboňová *et al.* 2012). In *T. kodakarensis*, cells with deletion of the other histone (HTkA) do not show a reduced growth phenotype but instead can no longer be transformed – a result not seen for HTkB-deleted cells (Čuboňová *et al.* 2012).

When there are multiple histones, differential patterns of gene expression can also result from deletion of either histone. Strains with either HTkA or HTkB deleted in *T. kodakarensis* exhibit differences in up- and down-regulated genes, with some overlap between the two (Čuboňová *et al.* 2012). The transcriptional differences for each mutant strain were not linked to the position of histone tetramers based on MNase profiles (Nalabothula *et al.* 2013), however any link to higher order oligomers was not investigated, and differences could also manifest through PTMs (see section 2.8.4). Differences in protein extracts from *M. voltae* with either

hstA or hstB deleted similarly show differences in changes to the transcriptome (Heinicke *et al.* 2004).

It is possible that structural differences between histones themselves may alter transcription. However, it is important to note that dosage and position effects may also play a role. In some cases, deleting one histone may result in overall lower levels of histone protein than deleting the other histone. Deletions can also locally disrupt and alter gene expression. In this regard, it is worth noting that expressing HTkA in *T. kodakarensis* from a plasmid is toxic, both for HTkA deleted strains, and wild type strains (Čuboňová *et al.* 2012).

One key feature of eukaryotic variants is the alteration in their relative abundance response to changes in environmental conditions. Many archaea live in extreme environments where they have to tolerate wide environmental fluctuations such as temperature changes. It is currently unknown whether histones could play a similar role in archaea, but we do have some information from gene expression patterns of archaea grown in different conditions. For example, the expression of two out of six histones in *M. jannaschii* (MJ0168, MJ1258) appears to be distinctly temperature-sensitive, and both are upregulated when exposed to heat shock, one more so than the other (Table 2.1) (Boonyaratanakornkit *et al.* 2007). Upon cold shock, these two histones (plus another – MJ0932) are downregulated (Boonyaratanakornkit *et al.* 2005). Two histones (MJ1258 and plasmid-encoded MJECL17) are upregulated when subjected to pressure change from 7.8 to 500 atm (Boonyaratanakornkit *et al.* 2007). Interestingly, one histone in *M. jannaschii* is downregulated in co-culture with *Thermococcus paralvinellae* vs in monoculture (Liu and Lu 2018). Histones in *T. kodakarensis* show decreased expression upon changes to a lower growth temperature (65^{o}C v 85^{o}C), one more so than the other (Sas-Chen *et al.* 2020).

Table 2.1 Differential expression of histones in response to environmental change.
This table was adapted from (Stevens and Warnecke 2022).

Organism	Condition	Per histone change in expression log2(fold change)	Reference
Methanocaldococcus jannaschii	Heat shock (98oC/88oC) at low pressure (7.8atm)	MJ0168 = 1.1 MJ1258 = 1.5	(Boonyaratanakornkit *et al.* 2007)
	Heat shock (98oC/88oC) at high pressure 500atm	MJ0168 = 1.1 MJ1258 = 1.8	(Boonyaratanakornkit *et al.* 2007)
	1.5h after pressure shock (500atm/7.8atm) at 88ºC	MJ1258 = 1.7 MJECL17 = 1.3	(Boonyaratanakornkit *et al.* 2007)
	Gas substrate limited (600ml/200ml) at 7.8atm	MJ0168 = 1.3 MJ1258 = 1.3	(Boonyaratanakornkit *et al.* 2006)
	Gas substrate limited (600ml at 500atm/200ml at 7.8atm)	MJ0168 = 1.6 MJ1258 = 1.2	(Boonyaratanakornkit *et al.* 2006)
	High H_2 v low H_2	MJ0932 = 1.4	(Topçuoğlu *et al.* 2019)
	Co-culture v Monoculture with *Thermococcus paralvinellae*	MJ0932 = -1.2	(Topçuoğlu *et al.* 2019)
	Temperature shock (65ºC/85ºC)	MJ0168 = -1.7 MJ1258= -1.5 MJ0932= -1.1	(Boonyaratanakornkit *et al.* 2005)
Thermococcus onnurineus	Nutrient change (modified minimal-CO/YPS)	TON_0185 = 1.6 TON_1235 = 0.18	(Cho *et al.* 2017)
	Nutrient change (modified minimal-formate/YPS)	TON_0185 = 0.48 TON_1235 = 0.56	(Cho *et al.* 2017)
Methanocella conradii	Co-culture v Monoculture with *Pelotomaculum thermopropionicum*	Mtc_0089 = 0.17 Mtc_0839=-1.1	(Liu and Lu 2018)
Thermococcus kodakarensis	Temperature change (65ºC/85ºC)	HTkA = -1.1 HTkB = -0.51	(Sas-Chen *et al.* 2020)

The physiological relevance of the differences shown above is unknown, but the evidence above hints that not all histones are playing equal roles.

2.8.3. Histone paralogs which show differences in expression levels

In vivo studies suggest that, in cells expressing multiple histones, not all paralogs are expressed to the same level. This includes the methanogenic archaea *M. voltae* (Heinicke *et al.* 2004) and *M. fervidus* (Sandman *et al.* 1994). Importantly, the differences between homologs can also vary at different stages of growth. *M. fervidus* histone HMfA is much more highly expressed than HMfB in exponential phase, constituting around 80% of histones isolated from cells

(Sandman *et al.* 1994). As cells approach stationary phase, the relative level of HMfB increases to reach a ratio of ~1:1 HMfA: HMfB (Sandman *et al.* 1994). Given the differences in DNA compaction levels generated by each *M. fervidus* histone (see section 2.8.1.3), this suggests that cells in stationary phase are likely characterised by more compact histone-based chromatin (Sandman *et al.* 1994). A similar situation can be seen in *M. voltae*, where one histone (hstA) decreases as cells move towards stationary phase but the other (hstB) increases (Heinicke *et al.* 2004).

2.8.4. Differential post-translational modification of archaeal histones

In eukaryotes, some PTMs are common to core histones and their variants, but others are specific to each variant, owing to specific amino acid substitutions (Corujo and Buschbeck 2018). Homologs of writers and erasers of histone-specific PTMs can also be found in archaea and bacteria, suggesting the potential for modification to the core HFD, similar to eukaryotic histones (Tessarz and Kouzarides 2014, Grau-Bové *et al.* 2022). In comparison to eukaryotes, the distribution of these modifying enzymes is much more sparce and some classes are absent entirely (such as KAT11) (Grau-Bové *et al.* 2022). In archaea, this primarily includes writers and erasers specific to histone acetylation and methylation, and both these PTMs have been identified for archaeal histones (Grau-Bové *et al.* 2022). However, the presence of histone modifying enzymes does not necessarily determine that histones will be modified, as demonstrated by the lack of PTMs on histones from *M. jannaschii* and *Methanosarcina acetivorans* (Forbes *et al.* 2004). This suggests that PTMs are unlikely to be the main mechanism employed to regulate histone-based chromatin in archaea.

Although PTMs are not found on all archaeal histones, they have been documented in select species. Some of these may be differentially post-translationally modified at unique sites. Evidence from *T. kodakarensis* and *Thermococcus gammatolerans*, which both have two histones, indicates that each histone exhibits a different pattern of lysine acetylation (Alpha-Bazin *et al.* 2021). These acetylated lysines are at different positions in the HFD from those of eukaryotic histones (Tessarz and Kouzarides 2014), suggesting they may result in distinct properties. At sites where the lysine is unique to only one histone, this can be acetylated, suggesting the pattern of modification may be specific to each homolog (Alpha-Bazin *et al.* 2021). Differential modification of histones was also observed in *Methanobrevibacter*

cuticularis, with lysine PTMs on two out of four histones, the pattern of which varied for each (Grau-Bové *et al.* 2022). The position of some histone PTMs are shared between species, such as *T. kodakarensis* and *T. gammatolerans* (Alpha-Bazin *et al.* 2021). Equally, shared patterns of lysine methylation and acetylation are seen in *Methanospirillum stamsii* and *M. cuticularis* (Grau-Bové *et al.* 2022).

3. Aims and Objectives

Whether the apparent differences between archaeal histones allow them to play adaptive roles in dynamically shaping chromatin architecture, similar to eukaryotic histone variants, and how important these differences are, remains an open question. Attempting to directly apply the same definitions of histone variants from eukaryotes to archaea is difficult. Unlike eukaryotic histones, there is positional ambiguity in relation to higher order structures. Archaeal histones are not tied to a defined position in the nucleosome and therefore variants are not likely swapped in and out of one specific space. Most archaeal histones are also not found in gene clusters, there are no known dedicated chaperones or other interactors, and we have no evidence for expression coupled to cell cycle phases. Instead of using the above definitions to identify variant histones in archaea, we can look for paralogs which display differences in protein sequences which are conserved across species (similar to H2A.Z), or those conferring unique structural properties (akin to H2A.X) (Talbert and Henikoff 2010, Henikoff and Smith 2015).

In this thesis, I use structural modelling and evolutionary approaches to try and understand whether paralogs play distinct roles in archaea. I show that histone paralogs with conserved properties are shared across species, indicative of maintenance over hundreds of millions of years, suggestive of the importance of each paralog, and with similar timescales to some eukaryotic histones. I also describe some variants, which we have come to call capstones, that may alter chromatin architecture more dramatically.

4. Dissertation layout

The results are divided into five chapters. The first (Chapter 5) concerns the *in silico* modelling of tetrameric histone complexes across archaeal species, and the resulting changes to predicted structural properties. In Chapter 6, I use models of *Methanosphaera stadtmanae* histones as a case-study to show divergence in structural properties within species, and identify one histone variant, a capstone, which may dramatically alter local chromatin architecture. In Chapters 7 and 8, I use phylogenetic approaches to identify conserved paralogs in Methanobacteriales and Thermococcales, respectively. Chapter 9 details the results of non-phylogenetic approaches comparing sequence diversification in archaeal and eukaryotic histones. Finally, in Chapter 10,

I outline the approaches I and Antoine Hocher took to try to characterise the capstone from *M. stadtmanae* using *in vitro* techniques and heterologous expression in *E. coli*. Chapter 11 discusses the implications of these results, and open questions which remain key to understanding the importance of histone variants in archaea.

5. Predicted structural diversity of archaeal histones

5.1. Introduction

As discussed in the Introduction, there is some evidence for functional differences between histones in archaea (see section 2.8). To date, this evidence is restricted to a few, mostly hyperthermophilic, species of archaea and has been mainly limited to those with just one or two histone genes, in particular *M. fervidus* and *T. kodakarensis*. In principle, this demonstrates that histone proteins in the same species can have differing structural properties which may translate into functional differences *in vivo*, akin to eukaryotic histone variants. But evidence of differences more widely in archaea is lacking.

Many archaea encoding multiple histone genes are difficult to culture and/or have no genetic tools, making characterisation of histones *in vivo* a daunting task for many species. I therefore used *in silico* structural modelling to probe the potential structural properties of archaeal histones, including those from species which are genetically inaccessible.

The work in this chapter was done under the day-to-day supervision of Jacob Swadling, who designed the pipeline to interrogate structural properties.

5.2. Initial testing: *in silico* DNA binding predictions for *M. fervidus* histones HMfA and HMfB

To see whether we could use *in silico* approaches to discriminate archaeal histone structural properties, I first modelled what are arguably the best characterised archaeal histones: HMfA and HMfB from *M. fervidus*. Models of each histone as a homotetramer were created using a fast-mutational scanning approach (see section 5.3) based on the HMfB crystal structure (PDB=5t5k, (Mattiroli *et al.* 2017)). This structure contains 6 HMfB monomers with DNA, allowing us to probe the DNA binding properties of each. Models were energy minimised, and the predicted DNA binding of each was calculated (see Methods). Gibbs free energy change (ΔG) at the interface between the protein and DNA was calculated as a readout of DNA binding affinity, and change in Gibbs free energy ($\Delta\Delta G$) determined relative to the HMfB structure.

HMfB hexamers are predicted to have stronger DNA binding than HMfA, consistent with trends from experimental *in vitro* work (ΔΔG HMfA hexamer = 17.456 kcal/mol relative to HMfB) (Sandman *et al.* 1994, Bailey *et al.* 2002). I also generated hexameric structures of HMfAB heterodimers, resulting in intermediate DNA binding between HMfA and HMfB (ΔΔG HMfAB = 7.68 kcal/mol relative to HMfB).

Investigating the DNA binding properties of heterodimers *in vitro* has required the use of fused histone dimers to force two histones to dimerise (Marc *et al.* 2002, Sandman *et al.* 2008). This strategy makes use of the short linker sequence to connect the N-terminal and C-terminal HFDs. Prior studies have used fused histone dimers with a linker sequence between each monomer to interrogate DNA binding differences between HMfA and HMfB, as well as heterodimeric mixtures of the two (Marc *et al.* 2002). These studies show the same trend in DNA binding as above, with HMfB binding DNA more tightly than HMfA. Complexes containing the forced HMfAB heterodimers have intermediate DNA binding strength between the two, albeit more similar to HMfB than HMfA (Marc *et al.* 2002). To see whether I could also recreate these results using our modelling technique, we created tetrameric models using the same linker amino acid sequences as Marc *et al*. (2002). As with the hexameric complexes, the overall pattern of predicted DNA binding strength (HMfB > HMfAB > HMfA) was conserved (ΔΔG relative to HMfB: ΔΔG HMfAB = 2.088 kcal/mol, ΔΔG HMfAA = 3.0233 kcal/mol).

This suggests that *in silico* modelling of histones onto the HMfB structure and subsequent analysis can replicate differences in DNA affinity that have been observed *in vitro*.

5.3. Generating tetrameric histone-DNA complex models for different archaeal species

To explore the potential diversity of histones more broadly across archaea, I generated further histone-DNA complexes for a large dataset of singlet archaeal histones. The structural diversity of archaeal histones containing a single HFD, and not doublet histones was explored for two main reasons. Firstly, the structure of doublet histones does not have DNA, precluding DNA binding calculations (Fahrner *et al.* 2001). Secondly, we know from prior work that the doublet histones in halophilic archaea are unlikely to play a major role in chromatin organisation (see section 2.8.1.1, (Dulmage *et al.* 2015, Sakrikar and Schmid 2021)). The doublets are also

restricted to fewer clades (Figure 2.6), whereas archaeal histones consisting of a single HFD can be found in a much wider range of archaeal species (Stevens and Warnecke 2022).

Protein sequences for histones from 282 diverse archaea were kindly curated by Simonetta Gribaldo (Institut Pasteur) (Adam *et al.* 2017) and filtered for lengths between 60-80 amino acids to obtain a comprehensive set of monomeric histones (Figure 5.1). These sequences were aligned and a list of substitutions needed to generate a model of each histone based on the HMfB structure was generated.

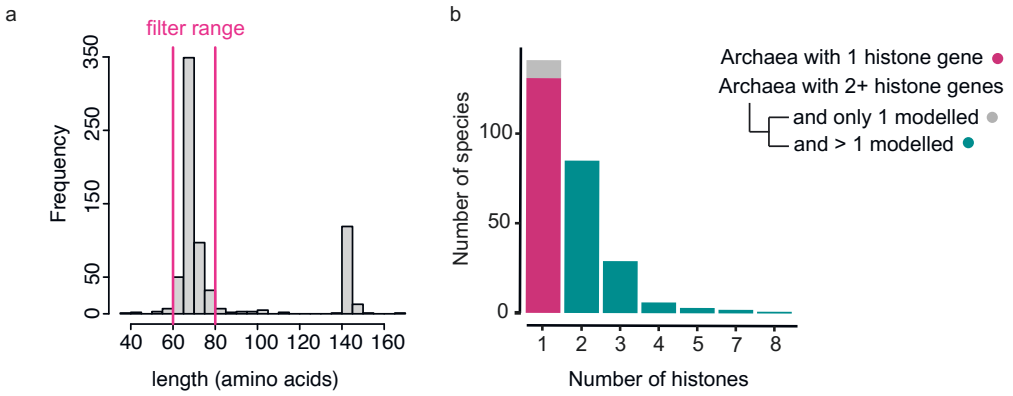


Figure 5.1 Length distribution of archaeal HMf-like histones used for modelling.
a) pink lines show the length distribution of histone sequences, from 60-80 amino acids. b) Number of species with each number of histones in the final modelling dataset after filtering. Only one histone was modelled in some species encoding 2+ histones, owing to the other being filtered out (see Methods). This figure was adapted from (Stevens *et al.* 2020).

For each species, models of the histone complex as a tetramer with DNA were created (Figure 5.2). The hexameric structure of HMfB from *M. fervidus* DNA (pdb = 5t5k, (Mattiroli *et al.* 2017)) was used as the initial starting structure to generate the models. The DNA in this structure was SELEX-selected for HMfB (Bailey *et al.* 2000, Mattiroli *et al.* 2017). This complex was reduced in size to a tetramer (four monomers, two dimers) and used as a template to build each model, using the BuildModel function of FoldX (Schymkowitz *et al.* 2005) to map sequences onto HMfB (see Methods). A tetramer model was used as this allows calculation of both DNA binding and tetramerization without assuming that larger oligomers are formed. Other than the hyperthermophilic archaea *M. fervidus* and *T. kodakarensis*, it is unknown whether histone polymerisation is a universal trait across archaea, and the propensity for this structure to form likely varies across species based on sequence differences (Henneman

et al. 2018). *In vivo*, archaeal histones can bind DNA as tetramers in multiple species, and represent the smallest DNA wrapping unit (Marc *et al.* 2002, Reeve *et al.* 2004, Mattiroli *et al.* 2017, Rojec *et al.* 2019).

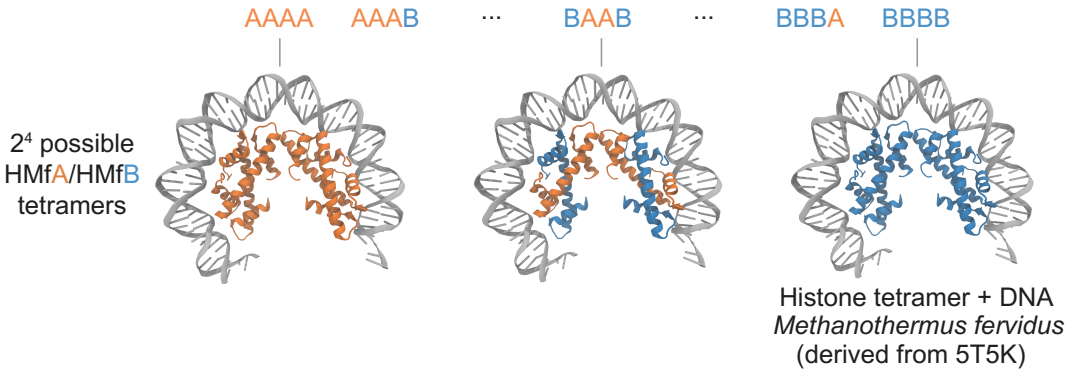


Figure 5.2 Generating models for all combinations of tetrameric histone complexes for each species. Models were generated for all possible combinations of histone tetramers (homotetramers and heterotetramers) for each species. For archaea encoding multiple histones, each monomeric histone was modelled into each chain –the example of models generated for HMfA (orange) and HMfB (blue) from *M. fervidus* is shown. This figure was adapted from (Stevens *et al.* 2020).

For archaea encoding multiple histone genes, all possible combinations of histone monomers in the tetrameric structure were generated, including both homotetramers and heterotetramers (Figure 5.2). To generate heterotetrameric models, each monomeric histone was modelled into each histone chain (one of each of the four monomers) in the tetramer structure, resulting in n^4 models per species, where n indicates the number of histone paralogs present in the genome (Figure 5.2).

Histones which had large insertions, deletions in the core histone fold, or long N-terminal extensions were excluded from any further analysis as the results from these are likely less accurately represented by homology modelling. This left a final set of 349 homo-oligomeric and 15905 hetero-oligomeric histone complexes (Figure 5.1b).

The models were energy minimised to reduce high energy clashes. The Gibbs free energy change (ΔG) at the interface between the histone tetramer and DNA (DNA binding) and at the interface between dimers (a proxy for tetramer stability/interaction strength) was then calculated for each model. As this approach allows for an interrogation of structural properties in computationally tractable timeframes, for the purposes of comparison, this is termed ‘fast-mutational scanning’.

5.4. Histone-DNA complexes exhibit a wide range of tetramerization strengths and DNA binding affinities

Predicted DNA binding and tetramerisation strength for histone-DNA complexes varies greatly across the archaeal species sampled (Figure 5.3a). Most models have slightly weaker energies than HMfB (more positive ΔΔG values, Figure 5.3b – darker areas indicate higher density of models with that energy) with two main DNA binding peaks. Outside of this cluster, some histones appear to exhibit more extreme changes to DNA binding and/or tetramerisation strength.

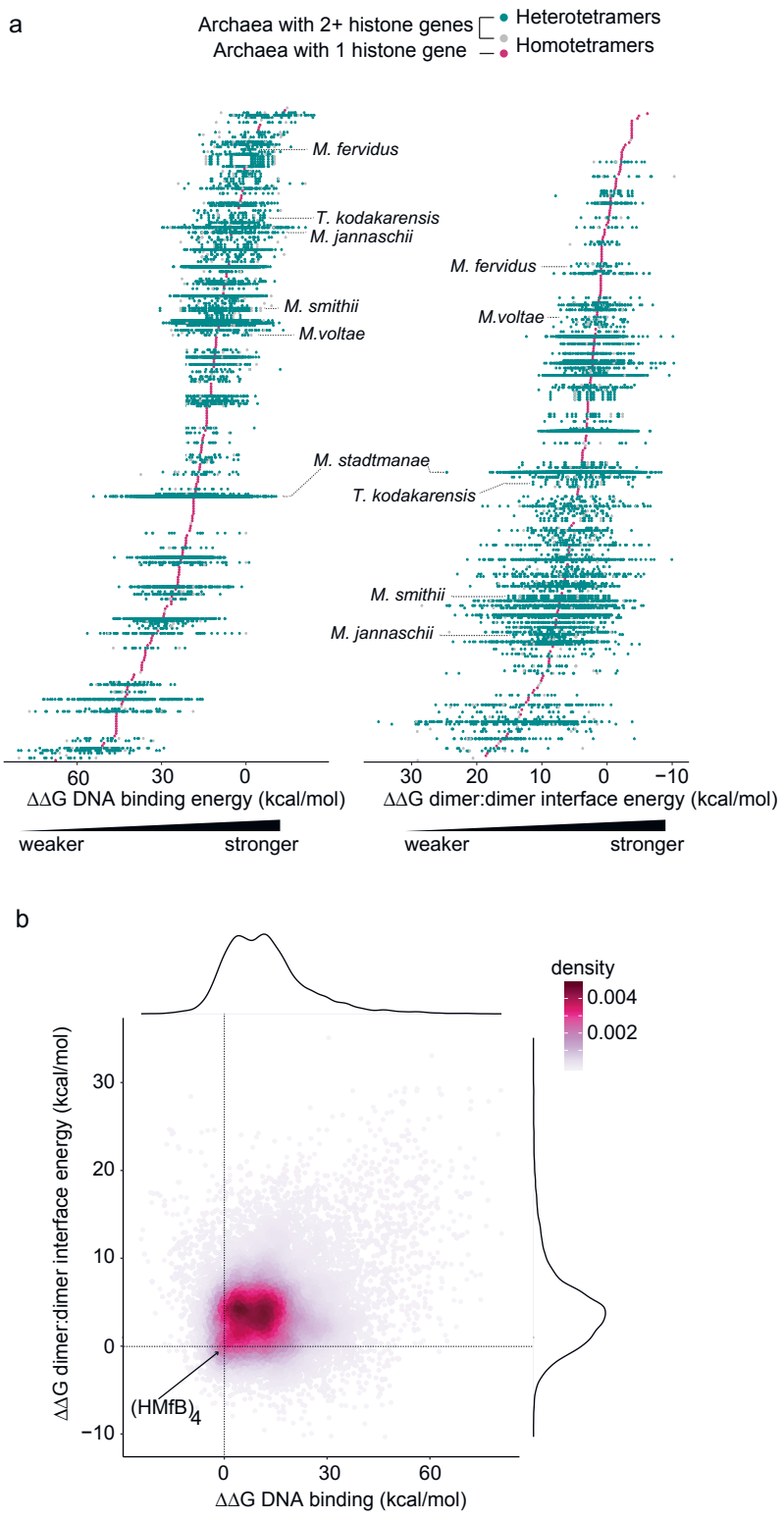


Figure 5.3 DNA binding and tetramerisation strength for tetrameric histone-DNA complexes based on the HMfB crystal structure.
a) DNA binding (top) and interaction strength between dimers (tetramerization interface energy - bottom) for all histone tetramer models for the archaea in the dataset. Results for each combination of histones in a tetrameric model are shown as an individual point, and are grouped by species on the y axis. Species are ordered by the mean interaction energy in each case. b) Density of energies for all models. DNA binding and tetramerisation are slightly weaker than HMfB for most histones. ΔΔG was calculated relative to HMfB from *M. fervidus*. This figure was adapted from (Stevens *et al.* 2020).

Some of the extreme cases are halophiles. For example, the histones with the weakest DNA binding and tetramerisation strength include many halophilic archaea, predominantly three Nanohaloarchaea species; *Nanohaloarchaea archaeon PLBr10_U2g19/21, Candidatus Nanosalina sp. J07AB43, Candidatus Haloredivivus sp. G17* (see Appendix Figure 14.1). Histones in these species often have substitutions at either residue 19, 53 or both, which are all known to be important for DNA binding (Soares *et al.* 2000). Aside from the DNA binding residues mentioned here, other residues may also contribute to DNA binding affinity. For example, mutating positions 18 and 51 in HMfB to the amino acids found at the same position in eukaryotic histones (E18K and G51K, respectively) increases DNA affinity (Soares *et al.* 2003). Some of these have been characterised using *in silico* mutation of the HMfB structure in Chapter 9.

The presence of halophiles at the extreme end of the data highlights that making meaningful comparisons across species is complicated by changes in intracellular and growth environments across archaea. Many different factors can influence the interactions between protein subunits and between proteins and DNA *in vivo*, including concentration of solutes in the cell, growth temperature, and intracellular pH (Martin *et al.* 1999). For example, the intracellular salt concentration is much higher in the halophilic archaea which have apparent weak DNA binding and tetramerization, as many halophilic archaea have much higher K^+ concentration than other archaea (Jensen *et al.* 2015, Thombre *et al.* 2016). This is a clear potential confounder when trying to understand effective differences in DNA binding behaviour as all calculations were performed under the same conditions *in silico*. Full characterisation of all archaeal species in this regard is missing, preventing controlling for intracellular or external changes in a systematic manner. Another confounding factor is that most of these histones have not been experimentally characterised and it is possible that, in some instances, they may have diversified to different roles outside general chromatinisation, such as involvement in transcriptional regulation more similar to the halophilic doublets (Sakrikar and Schmid 2021, Sakrikar *et al.* 2022). In the absence of complete information, comparison of histone-DNA complexes across species is therefore difficult. We can, however, make comparisons *between* histones *within* species, where intracellular conditions will be more consistent for different histones.

6. Structural diversity of *Methanosphaera stadtmanae* histones

6.1. Introduction

Which system provides a good background in which to explore histone structural diversity within species? From the structural modelling results, *M. stadtmanae* stands out. It has seven histones, and displays the largest range in DNA binding, and fourth largest range of tetramer stability based on our modelling results (Figure 5.3a). The high number of histones and wide range of predicted energies for *M. stadtmanae* histones positions this species as a good model from which to probe further the apparent differences between histones.

Methanosphaera stadtmanae (originally named *Methanosphaera stadmaniae* (Miller and Wolin 1985)) is a methanogenic archaeon belonging to the Methanobacteriales (Liu and Whitman 2008, Borrel *et al.* 2020) and was isolated from the human gut (Miller and Wolin 1985). It can grow on methanol and H_2, at an optimum temperature of 36-40^{o}C, and a pH optimum of 6.5 to 6.9 (Miller and Wolin 1985). The 1.7 Mbp circular genome of the strain *M. stadtmanae* DSM3091 has a low GC content (28%) (Fricke *et al.* 2006).

M. stadtmanae has possible implications for human health, and may induce a pro-inflammatory response. Unlike the methanogen *Methanobrevibacter smithii*, it has been implicated in inflammatory bowel disease and has potential clinical relevance (Blais Lecours *et al.* 2014). *M. stadtmanae* is more prevalent in the stool of patients suffering with Irritable Bowel Disease (IBD) compared to control samples (Blais Lecours *et al.* 2014).

6.2. *M. stadtmanae* encodes 7 histones

Methanosphaera stadtmanae DSM3091 has seven genes encoding histones, a relatively high number compared to other archaea (Figure 2.6, Figure 6.1a). Of these, six histones were previously identified in the original annotated genome sequence (Fricke *et al.* 2006) with one additional histone (Msp_0924.5) between genes annotated as Msp_0924 and Msp_0925.

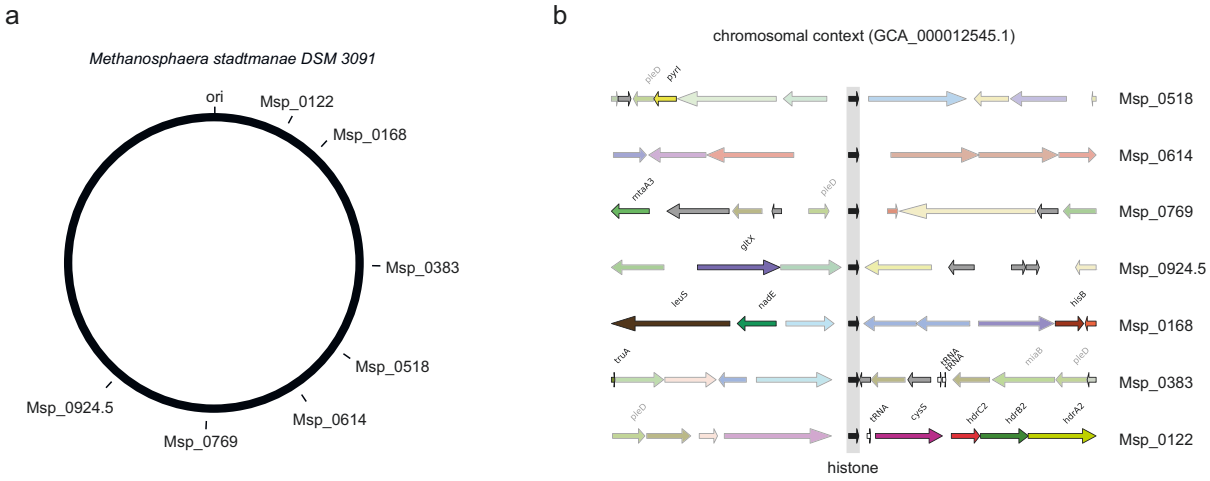


Figure 6.1 Location and chromosomal context for the 7 histone genes in *M. stadtmanae* DSM 3091.
a) Chromosomal map showing the location of each histone, including the extra identified histone Msp_0924.5. b) Genomic context for the 7 histones plotted using GeneSpy (Garcia *et al.* 2019) plotted by Antoine Hocher. This figure has been published in (Stevens *et al.* 2020).

All seven histone genes are surrounded by intergenic regions, both upstream and downstream, suggesting they are transcribed as single genes (Figure 6.1b), further confirmed by operon-mapper (Taboada *et al.* 2018). This single operon feature appears to be a common characteristic, seen more broadly for different NAPs (Hocher *et al.* 2021).

M. stadtmanae histones display high variability in their protein sequence (Figure 6.2). The sequence identity ranges from the most distant histones, with 41% sequence identity (Msp_0614 and Msp_0924.5), to the most similar with 66% identity (Msp_0518 and Msp_0168). Residues directly involved in DNA binding in *M. fervidus* HMfB (positions 10, 19, 53, 54 in HMfB, 11, 20, 54, 55 in the alignment) are conserved across all seven histones (Figure 6.2, see Chapter 9) (Soares *et al.* 2000).

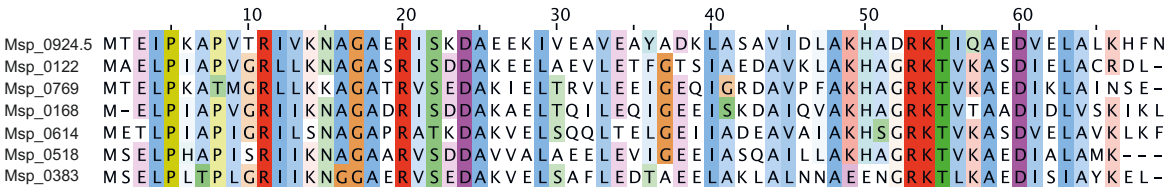


Figure 6.2 Alignment of the 7 histones in *M. stadtmanae*.
This figure was generated using Jalview (Waterhouse *et al.* 2009). This figure has been published in (Stevens *et al.* 2020).

6.3. *M. stadtmanae* histone complexes occupy a broad range of predicted structural properties

Tetrameric complexes built from combinations of the seven *M. stadtmanae* histones (=2401, see Chapter 5) had a wide range of energies (Figure 5.3a). Predicted DNA binding affinities ranged from ΔΔG of -10.47 to 54.39 kcal/mol, and predicted tetramer stabilities ranged from ΔΔG of -9.33 to 23.63 kcal/mol relative to HMfB.

In comparison to other methanogens, *M. stadtmanae* has a much higher number of histones. How do the predicted structural properties of *M. stadtmanae* histones compare to methanogens with fewer histones? The tetrameric modelling includes results for some of these methanogenic species, including *M. fervidus* and *M. smithii*, which have 2 and 3 histones, respectively. *M. smithii* and *M. stadtmanae* are found in similar environments; both are mesophilic methanogenic species which were isolated from human faeces and form part of the human intestinal microbiota (Miller *et al.* 1982, Miller and Wolin 1985). *M. fervidus* was isolated from more extreme environments, with a much higher optimum growth temperature of 83ºC (Stetter *et al.* 1981). Compared to both *M. fervidus* and *M. smithii*, *M. stadtmanae* histone-DNA complexes display a much denser and broader range of both DNA binding and tetramerisation energies (Figure 6.3). Even with just three histones, *M. smithii* histone complexes are able to sample a wide range of energies, a point returned to later on (see Chapter 7).

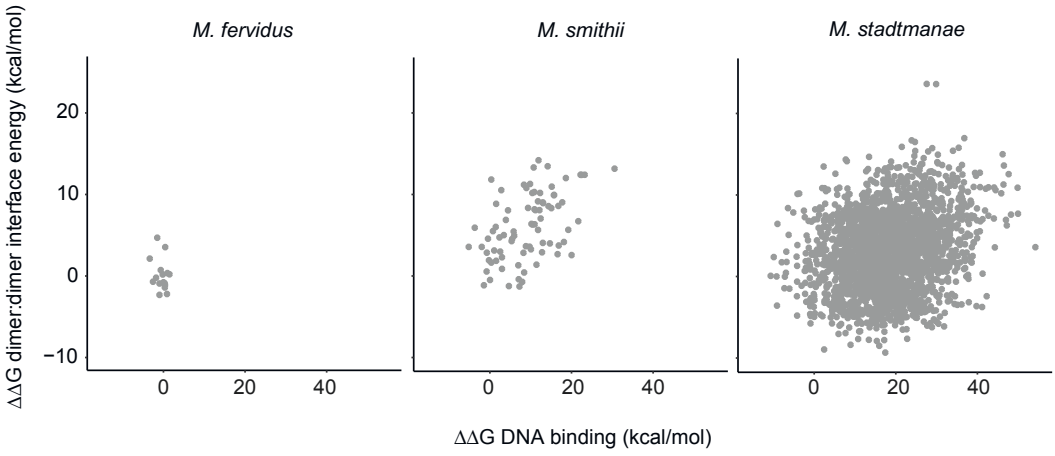


Figure 6.3 DNA binding and tetramerization strength (interaction strength between dimers) for methanogenic species *M. fervidus*, *M. smithii* and *M. stadtmanae*.
ΔΔG values are shown relative to HMfB. This figure has been published in (Stevens et al., 2020).

Does *M. stadtmanae*, with its high number of histones, display the potential to generate different chromatin states through exchange of histones? For *M. stadtmanae* to be able to tune its chromatin architecture by the differential assembly of paralogs, each must confer at least partially distinct structural changes to the histone-based chromatin. To determine whether each paralog has defined predicted structural properties, the identity of the dominant histone was mapped onto the results for all tetrameric complex models containing two different histones. The state space occupied by complexes containing different combinations of histones is not random. Overlapping but reasonably distinct areas are occupied by complexes dominated by a particular histone (Figure 6.4a). Homotetramers of each paralog occupy discrete DNA binding and tetramer stability energies (Figure 6.4a). Complexes containing predominantly one particular histone (where three out of four of the histones are of one type) occupy the state space near to the homotetramer for that histone (Figure 6.4a). The space between these extremes is heavily populated with other complexes.

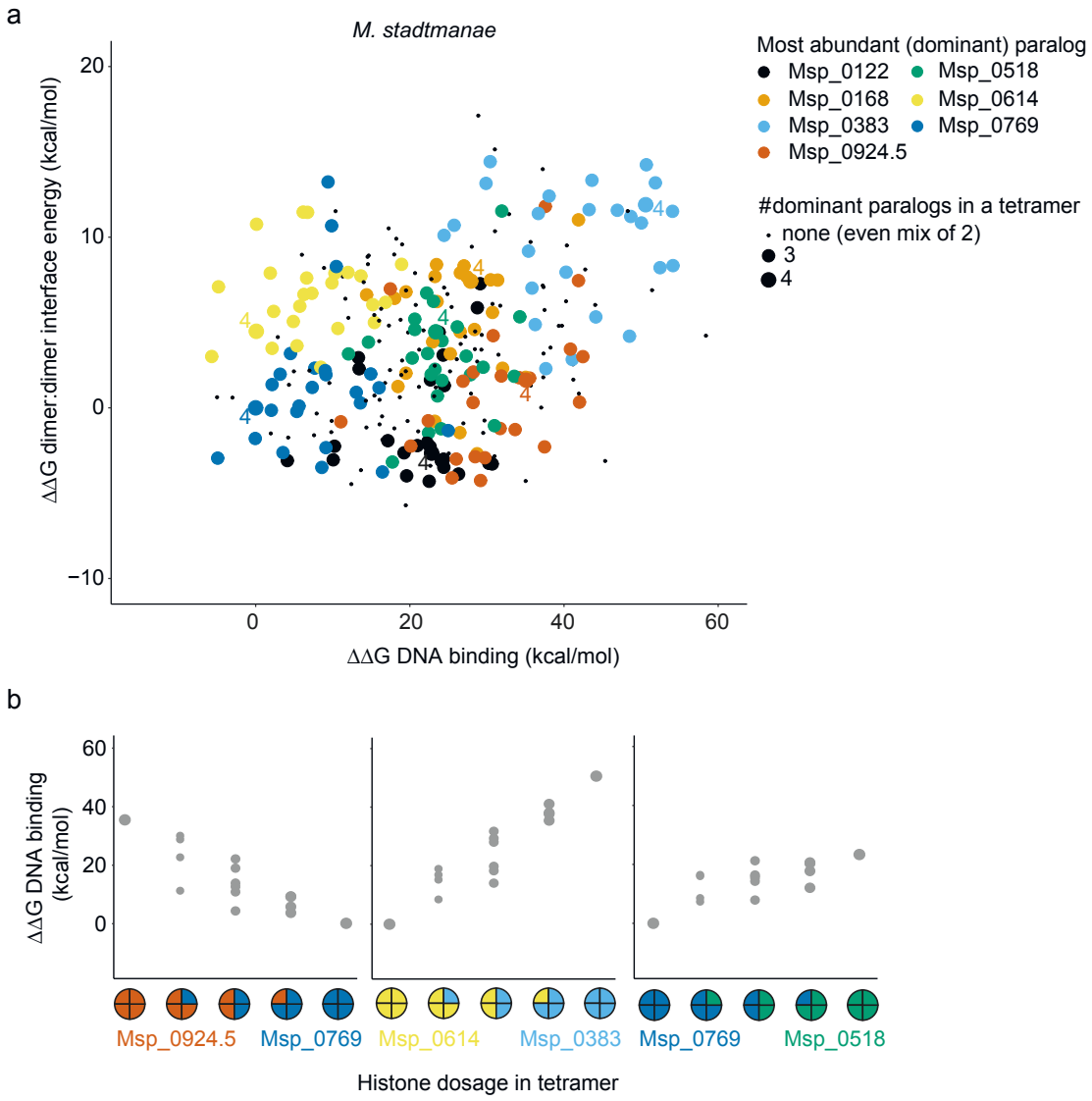


Figure 6.4 Modelling results for *M. stadtmanae* histones.
a) DNA binding and tetramerization strength (interface energy between dimers) for models containing two different paralogs. Each point represents an individual histone model. Points are coloured by the dominant histone in the complex (i.e., if three monomers in the tetramer are Msp_0769, the point is coloured dark blue). b) DNA binding results for complexes containing two paralogs. Each point represents an individual model, and points are grouped by the number of a particular histone paralog in the tetramer. DNA binding results are shown in increasing paralog dosage. ΔΔG values are shown relative to Msp_0769. This figure has been published in (Stevens *et al.* 2020).

For some paralog pairs, DNA binding scales with increasing the occupancy of a particular paralog in incremental steps from one homotetramer to another (Figure 6.4b). For example, Msp_0924.5 has a relatively weak predicted DNA binding affinity whereas Msp_0769 is predicted to have a strong DNA binding strength (Figure 6.4a, b). Swapping one Msp_0924.5 monomer for Msp_0769 to a homotetramer of Msp_0924.5 increases predicted DNA binding. This increase in binding strength continues as further monomers of Msp_0769 are added into

the tetramer complex (Figure 6.4b). This scaling is not seen for tetramerisation. In contrast to DNA binding, where every monomer contacts the DNA and individually contributes to DNA binding, only two monomers are at the interface between dimers, and therefore the identity of these drives tetramerization strength.

6.4. Using *in vivo* expression levels of *M. stadtmanae* histones to predict changes to histone-based chromatin in stationary vs exponential phase

The modelling results revealed the possibility of graded chromatin control in *M. stadtmanae* through iteratively changing the components of histone oligomeric complexes. Whether this occurs *in vivo*, and what energies of the histone complex landscape are used, remains unknown. If incorporation of histones into chromatin – in the absence of dedicated chaperones or sequence specificity – is random, the composition of histone-based chromatin will likely be defined by changes in the expression level of histones (see Discussion) (Stevens and Warnecke 2022).

Ruth Schmitz and Corrina Bang (our collaborators at the University of Kiel in Germany) provided cell-free protein extracts for *M. stadtmanae* cultures in exponential and stationary phase and used RT-qPCR to track histone expression level. Label-free mass spectrometry was also used to determine the abundance of each of the 7 histones in *M. stadtmanae* at the LMS (London Institute of Medical Sciences) (Figure 6.5a). Cells were grown to exponential and stationary phase in growth medium as previously described (Bang *et al.* 2012).

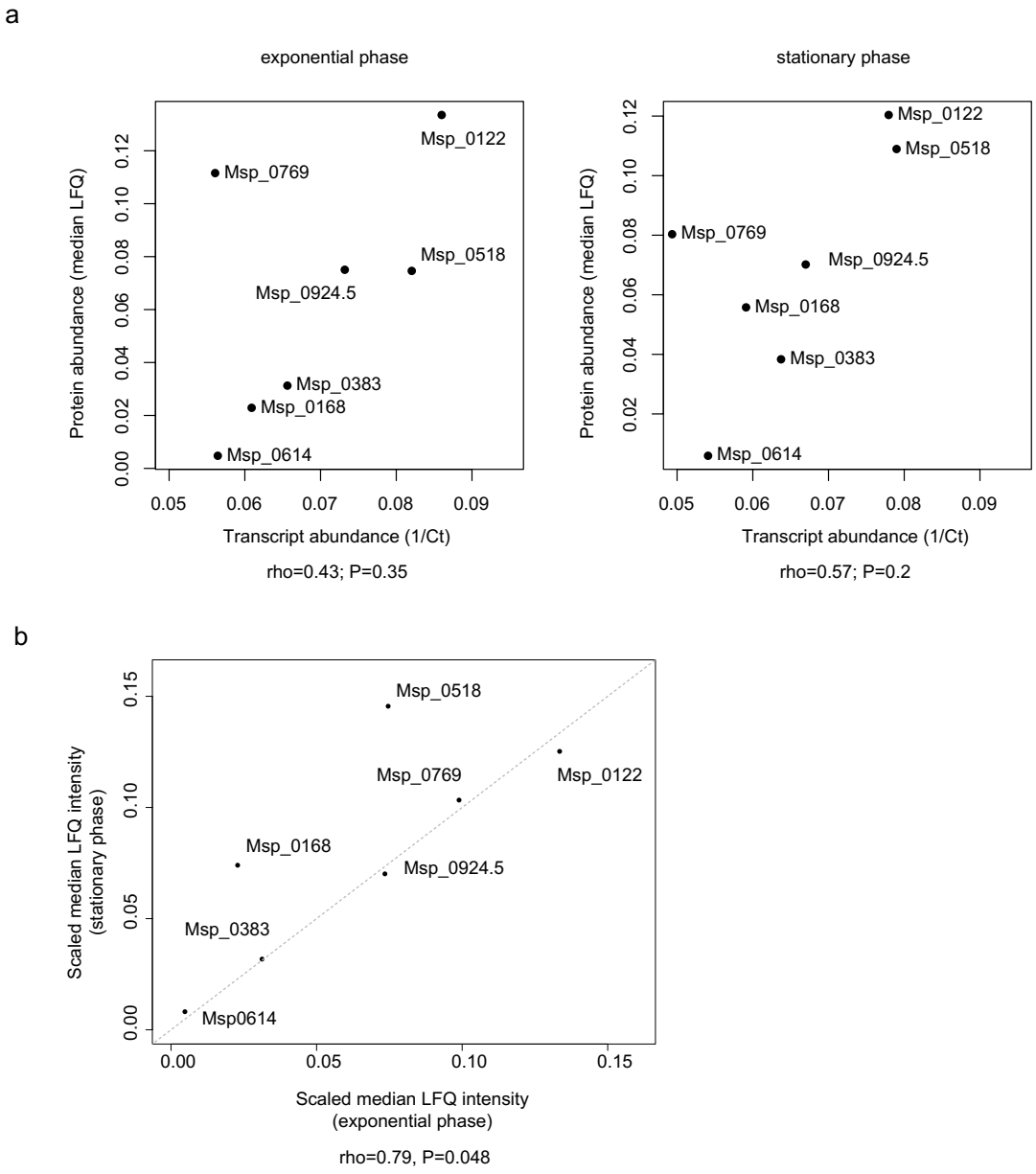


Figure 6.5 Expression level of the 7 histones in *M. stadtmanae*.
a) Transcript abundance and Protein abundancy (as median LFQ intensity) in exponential and stationary phase. b) Expression level in stationary phase and exponential phase as median protein abundancy. The expression data used here were generated by Ruth Schmitz and Corrina Bang from the University of Kiel. This figure has been published in (Stevens *et al.* 2020) and was plotted by Tobias Warnecke.

Overall relative expression levels in exponential and stationary phases are well correlated (Figure 6.5b). Msp_0122 has consistently the highest median protein abundance (Figure 6.5a, b) and paralogs Msp_0769, Msp_0518 and Msp_0924.5 are also relatively highly abundant. Msp_0614 is consistently the least abundant paralog in both growth phases (Figure 6.5a).

Histones Msp_0168 and Msp_0518 show a noticeable increase in median protein abundance in stationary compared to exponential phase (Figure 6.5a).

For *M. stadtmanae* cells in exponential phase, there is a strong correlation between paralog abundance and tetramer stability for each histone (Figure 6.6a, rho = -0.82, P = 0.034). The histones which are the most highly abundant (Msp_0122 and Msp_0769) are predicted to have the strongest tetramer stability, and predicted stability decreases in paralogs with lower abundance.

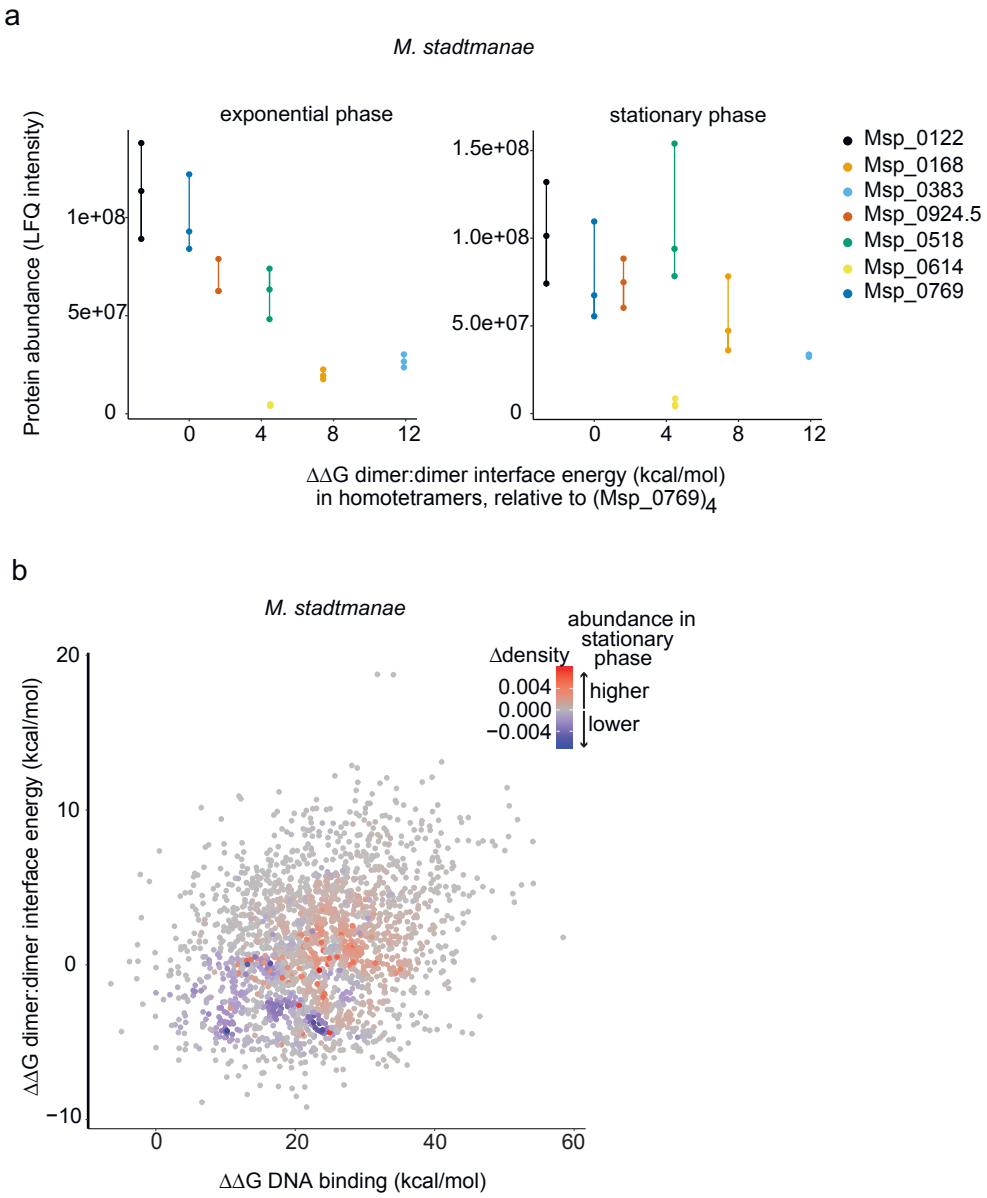


Figure 6.6 Histone expression level in *M. stadtmanae* in exponential phase and stationary phase of growth.
a) Interface energy between dimers (tetramerization strength) and protein abundancy (replicates of LFQ intensity) for each histone. b) Random sampling of 100,000 histone tetramers weighted by the relative abundancies of each histone (mean LFQ intensity) in stationary and exponential phase. Relative change in predicted complex abundancy is shown. Red = increased abundancy in stationary phase, blue = decreased abundancy. ΔΔG values are shown relative to Msp_0769. This figure has been published in (Stevens *et al.* 2020).

Does this correlation between abundance and tetramerization extend to other species? To answer this questions, the previous modelling results were compared to transcript and protein abundance data processed by Tobias Warnecke from GEO and primary publications (Jäger *et al.* 2014, Cho *et al.* 2017, Müller *et al.* 2020). Correlation between abundance and tetramerisation appears also to occur in at least one other Methanobacteriales species, *M.*

smithii. Histones Msm_0213 and Msm_0844 are both highly expressed, and have a predicted stronger tetramerization strength than the lesser expressed Msm_1260 (Figure 6.7). Stronger tetramerization is also observed in the most abundant of the two histones in *Ferroglobus placidus*, *Archaeoglobus fulgidus*, and *Archaeoglobus profundus* (Figure 6.7). However, this pattern is not observed in the hyperthermophilic archaeon *T. kodakarensis*, nor *T. onnurineus* (Figure 6.7). The difference between species suggests there is no global relationship across archaea between stability and abundance of histones, suggesting there may be differences in the type of composition of histone-based chromatin. As highlighted earlier, comparison across species can be confounded by intracellular and environmental conditions. Integrating this information could help identify if global trends occur.

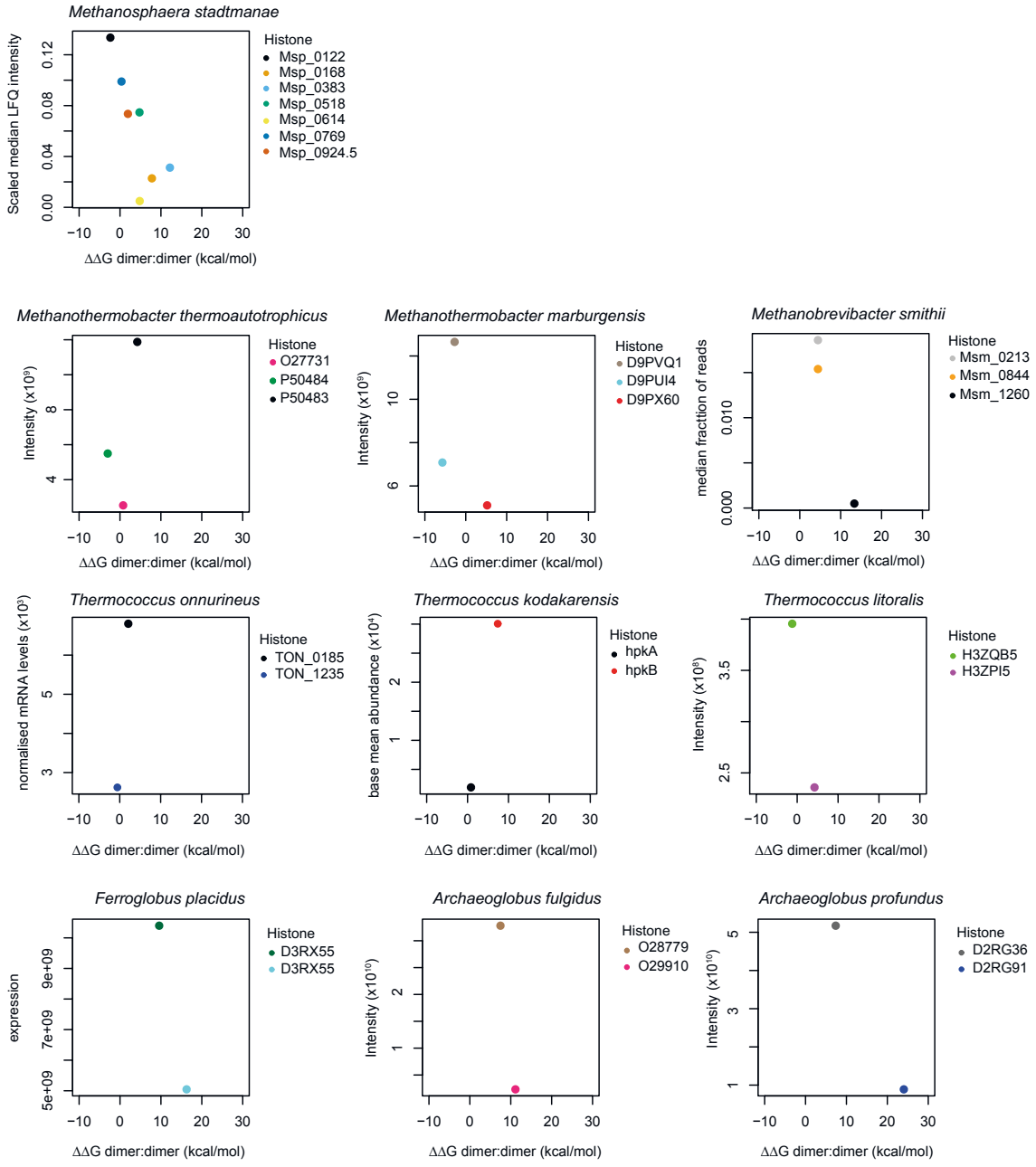


Figure 6.7 Histone expression level and tetramerization strength for different archaeal species. Tetramerisation strength shown as interface energy between dimers. Expression data for species other than *M. stadtmanae* was obtained from NCBI's GEO and primary publications by Tobias Warnecke (see Methods for details). This figure has been published in (Stevens *et al.* 2020).

The abundance of histones in *M. stadtmanae* can be used to predict changes to the relative abundance of histone complexes at different stages of growth. To model predicted histone complexes in the cell, 100,000 tetramer complexes were sampled with the chances of each histone being incorporated to each complex weighted by their relative protein abundance.

Using this method, histones with lower abundancy have a relatively lower chance of being picked for complex formation than histones with a high abundancy.

From exponential to stationary phase, there is a shift in the centre of gravity of chromatin state space from more stable tetramers with higher DNA binding, to weaker tetramers with decreased predicted DNA affinity (Figure 6.6b). This is primarily due to the upregulation of Msp_0168 and Msp_0518 (Figure 6.6). This reduced tetramer stability may act to overall decrease the abundancy of stable larger oligomeric complexes in stationary phase. Based on this shift between growth phases, *M. stadtmanae* cells in stationary phase are predicted to be characterised by more accessible histone-based chromatin. This is in contrast to what has been observed for *M. fervidus* where, in stationary phase, HMfB – which has increased DNA affinity and compaction levels – increases relative to HMfA, suggesting a more compact chromatin state in stationary phase (Sandman *et al.* 1994). However, it is important to note that this analysis is a purely histone-centric result. Here, only histone abundance is considered, and not the interactions, if they occur, with other protein components nor how these change across conditions. The NAP Alba is also present in *M. stadtmanae* (Hocher *et al.* 2021). We know nothing about how histones in *M. stadtmanae* interact with Alba and other chromatin associated proteins, which may further modulate chromatin state (see Discussion).

6.5. Molecular dynamics simulations of *M. stadtmanae* histone homotetramers reveal histones with unique properties

To gain a deeper understanding of the extremes of histone-based chromatin in *M. stadtmanae*, molecular dynamics simulations were performed for homotetrameric models of each histone. As noted previously, our initial fast-mutational scanning approach to modelling considered the core histone fold domain, and did not include small insertions or any N/C-terminal extensions. We therefore built 'full' atomistic models for each tetramer, by adding or removing residues where necessary (Figure 6.2).

To prepare the models for simulation, the overall negative charge of the DNA backbone was neutralised with Na^+ ions and the model was solvated before being subjected to energy minimisation and a short heating step. After heating, 100 ns simulations of each model were performed using classical molecular dynamics. DNA binding and interaction strength between

dimers (tetramer stability) were calculated after the models reached equilibrium, from frames 1,500 to 5,000 (Figure 6.9a).

The DNA binding affinities from the simulations correlate well with those calculated using static models (fast mutational scanning in Chapter 5) (rho=0.96, P<0.001, Figure 6.8a). Tetramerisation energies, in comparison, do not correlate as well (rho=0.5, P>0.001, Figure 6.8b), suggesting deviations from the crystal structure and dynamic behaviour revealed during the simulations.

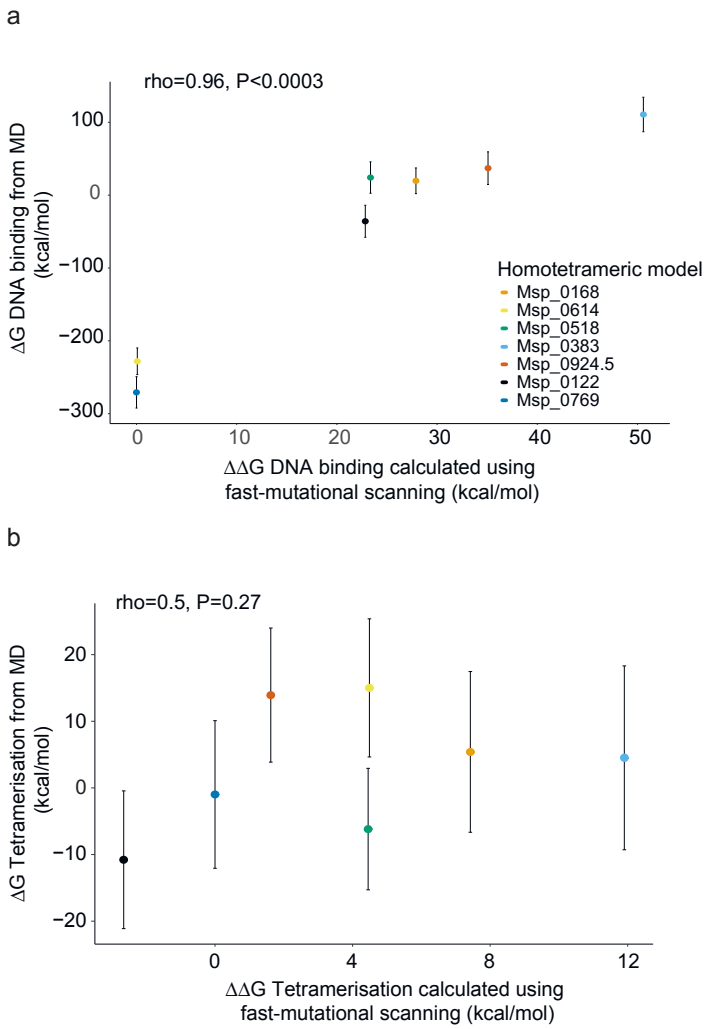


Figure 6.8 DNA binding and tetramerization for *M. stadtmanae* histone tetramer models over 100ns compared to the fast-mutational scanning approach
a) DNA binding and b) tetramerization for molecular dynamics (MD, y axis) and original tetrameric models (fast-mutational scanning, x axis – Chapter 5). This figure has been published in (Stevens *et al.* 2020).

What could be driving these structural differences? Root mean squared deviation (RMSD) shows the overall movement of atoms in a structure over a simulation, and can be used to monitor change in the overall trajectory from the initial starting structure. Some movement away from the initial starting structure is to be expected – crystal structures are usually high energy and, during the simulation, the atoms move to lower energy positions.

In our simulations, there is a movement of around 4 Å from the starting structure for all models, with the exception of Msp_0383 which displays a much more extreme deviation (up to ~14 Å) (Figure 6.9a). Based on our initial models, this histone had the weakest predicted DNA binding and tetramerization strength (Figure 6.4a).

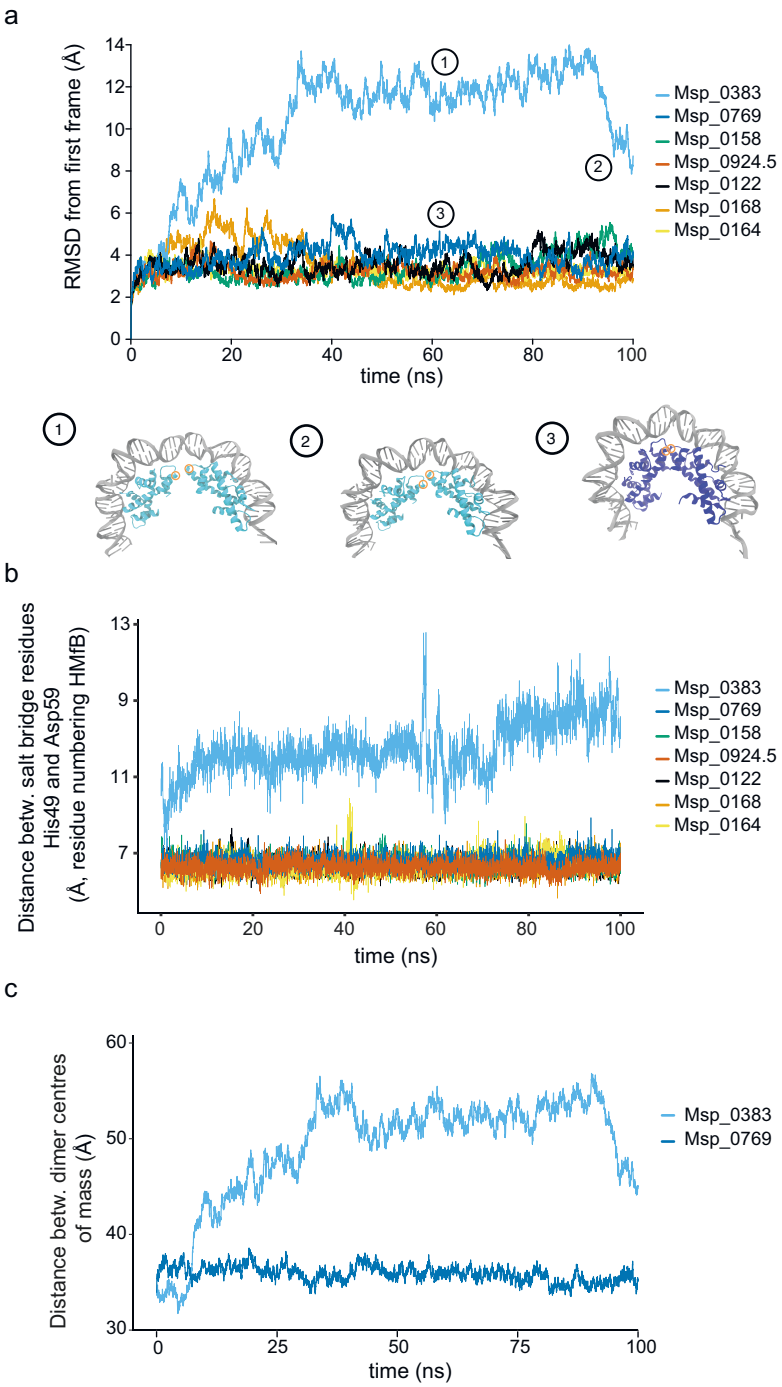


Figure 6.9 Trajectory *of M. stadtmanae* homotetrameric histone-DNA models during MD simulations. a) Root mean squared deviation (RMSD) from the starting structure for each model over the 100 ns simulation. Each homotetrameric complex is coloured by the histone paralog. Three snapshots of the trajectory are shown for 1) and 2) Msp_0383, 3) Msp_0769, and the relative time points indicated on the RMSD plot underneath. b) Distance between the Histidine (H) at position 49 relative to HMfB and Aspartic acid (D) at position 59 relative to HMfB sequence over time for each homotetramer. c) Distance between the centre of mass for each dimer for homotetrameric models containing Msp_0383 and Msp_0769. This figure has been adapted from (Stevens *et al.* 2020).

Analysing the trajectory, we find that Msp_0383 exhibits an unstable interface between the two dimers, unlike the other 6 histone models. Over the simulation, the dimers move apart, whilst remaining bound to the DNA (Figure 6.9a, see RHS panel, Figure 6.9b, c). This binding mode is particularly divergent from the other histones and suggests that Msp_0383 cannot form stable tetramers.

What is driving this divergent behaviour at the amino acid level? Msp_0383 diverges in sequence from the other histone paralogs in *M. stadtmanae* (Figure 6.2). Notably, it has a substitution from a positively charged (H, histidine) amino acid to negatively charged (E, glutamic acid) at a position which is conserved in the other *M. stadtmanae* histones, and indeed more widely across archaea (Position 50 in Figure 6.2, position 49 relative to the HMfB protein sequence). This residue is known to form a salt bridge with D59 (aspartic acid, HMfB residue numbering) in HMfB, and is necessary for tetramerisation (Marc *et al.* 2002, Higashibata *et al.* 2003). In *Thermococcus kodakarensis* histone HTkB, mutation to an alanine, valine or glycine at this position results in reduced DNA compaction (Higashibata *et al.* 2003). This site is particularly sensitive for HMfB *in vitro*, and mutants can alter the DNA binding and supercoiling state of HMf histones (Marc *et al.* 2002). HMfB with the H49E substitution lost the ability to flexibly constrain DNA in both positive and negative supercoiling, and plasmid DNA was only found in the positively supercoiled state (Marc *et al.* 2002). This mutant incubated with linear plasmid DNA resulted in less compaction, and was unable to form tetramer complexes with short 80bp DNA fragments unless very concentrations of the histone were added to DNA with a high affinity for HMfB (Marc *et al.* 2002).

Is the instability of Msp_0383 tetramers due to this single amino acid change? *In silico* substitution at this position back to the commonly found H reduces the distance between dimers dramatically in simulations (Figure 6.10a), almost back to levels seen for the other histones.

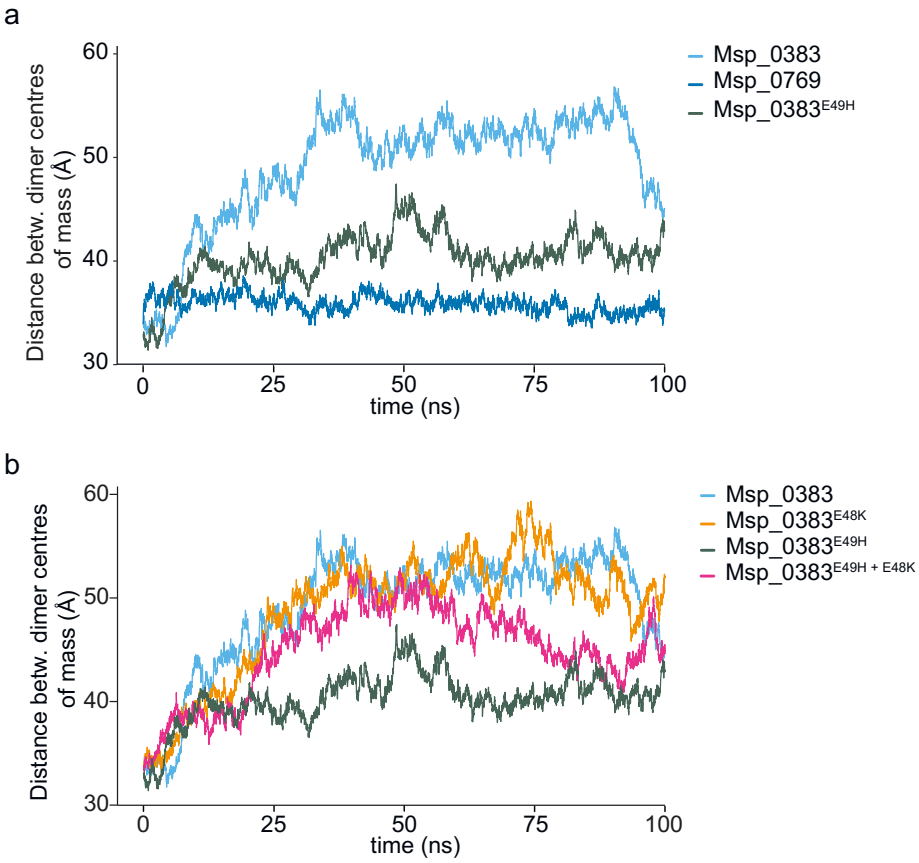


Figure 6.10 Distance between centre of mass for Msp_0383 reversion mutants.
a) Distance between the centre of mass for dimers in homotetrameric models of a) Msp_0383, Msp_0769 and Msp_0383 with E49H substitution, and b) Msp_0383, Msp_0383 E49H, Msp_0383 E48K and Msp_0383 E49H + E48K. This figure has been adapted from (Stevens *et al.* 2020).

There is also a substitution at position 48, again from a positively charged (K, lysine) to negatively charged (E, glutamic acid) amino acid. Could this position also drive the weak tetramer interface in Msp_0383? Reversion mutants of Msp_0383 *in silico* with E48K and with the double mutant E49H + E48K were also created. Unlike E49H, there is not a dramatic reduction in the distance between dimers (Figure 6.10b). The single E48K mutant has almost the same refraction to tetramerisation as the wild type (WT) Msp_0383 (Figure 6.10b). The double mutant has slightly reduced distance between dimers compared to WT over time, but slightly greater than the single E49H mutant (Figure 6.10b).

The dynamic structural properties of this histone can therefore be traced back to a single amino acid change. The H49E substitution alone is sufficient to reduce the distance between dimers. Due to its divergent structure, we named this variant ‘capstone’. When incorporated into a

growing oligomeric complex, this capstone may limit further oligomerization because of the unstable tetramer interface.

6.6. Capstones outside the Methanobacteriales

Are there examples of potential capstones in other archaea or is this variant unique to *M. stadtmanae*? Given my modelling results, an argument can be made that the capstone histone can be minimally defined by a substitution away from a positively charged amino acid at position 49, relative to the HMfB sequence. By searching for histones with a substitution at position 49 to a negative or hydrophobic amino acid, I find other putative capstones in a number of archaeal lineages (Figure 6.11, Appendix Table 14.1), including Hadesarchaea, MSBL1 and Nanohaloarchaea, although most lineages do not encode capstones.

Archaeal histones have been predominantly studied in hyperthermophilic species (Grayling *et al.* 1996, Fahrner *et al.* 2001, Friedrich-Jahn *et al.* 2009, Čuboňová *et al.* 2012). It is therefore noteworthy that species with capstones tend not to be thermophiles or hyperthermophiles, although we found no significant trends in environment or GC content between capstone and non-capstone archaea.

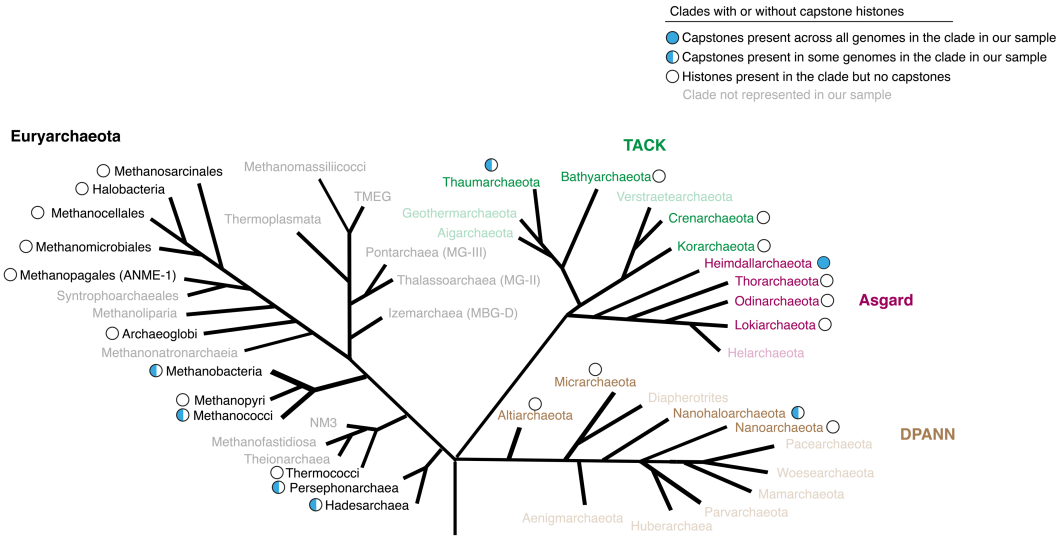


Figure 6.11 Phyletic distribution of capstone histones in archaea.
For the diverse archaeal dataset used for modelling, the presence of potential capstones (defined by substitution to a hydrophobic or negatively charged amino acid at position 49 relative to HMfB) is shown. For each lineage, blank circles denote no capstones, blue circles show capstones present in all genomes, and circles with half blue show histones present in some genomes. Lineages not present in our dataset are shown in reduced opacity. The tree was adapted from (Borrel *et al.* 2020), and has been published in (Stevens *et al.* 2020).

6.7. Conservation of unique properties across capstones from different species

On the basis of sequence alone, potential capstone histones can be found across many species of archaea. But do these putative capstones retain the structural properties found in *M. stadtmanae*? To see whether these putative capstones also exhibit the refraction to tetramerization observed for capstone Msp_0383 in *M. stadtmanae*, I performed further short molecular dynamics simulations. A subset of the tetramer histone models from species with a capstone and noncapstone histone (*M. smithii*, *Hadesarchaea archaeon* DG-33-1, and *Candidatus Nanosalina* sp. J07AB43) were subjected to the same simulation protocol as performed for the *M. stadtmanae* histone homotetramer models.

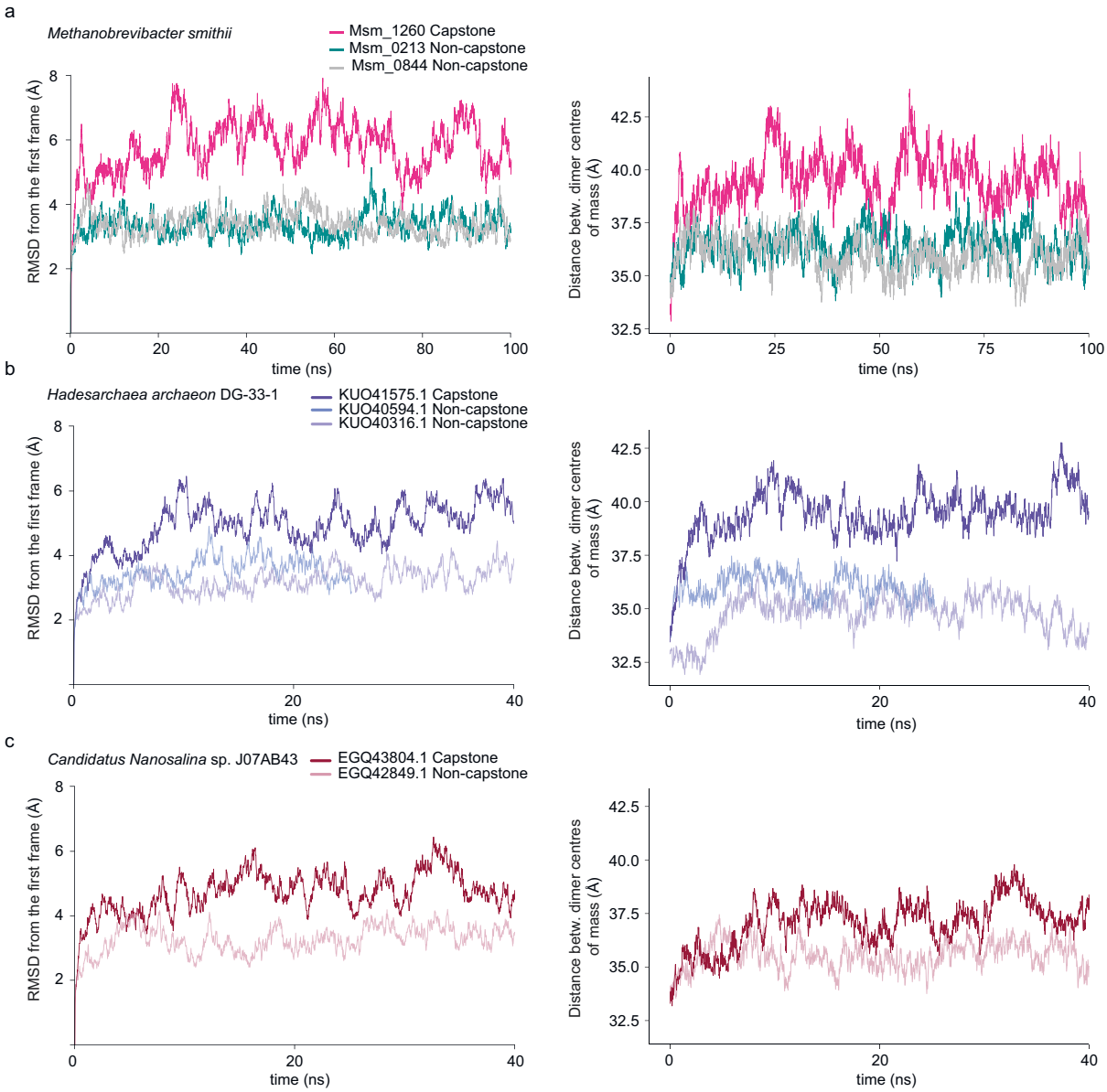


Figure 6.12 Molecular dynamics simulations of histones from species containing putative capstone histones.
Tetrameric models of capstone and noncapstone histones from a) *Methanobrevibacter smithii*, b) *Hadesarchaea archaeon* DG-33-1 c) *Candidatus Nanosalina* sp. J07AB43 were simulated using the same approach as for *M. stadtmanae* histones. Panels on the left-hand side show RMSD from the starting frame over time. The right-hand side plots show the distance between the centre of mass of each dimer over the simulation. This figure has been published in (Stevens *et al.* 2020).

As is the case for the *M. stadtmanae* capstone Msp_0383, there is a reduction in the distance between dimers over the simulation for all putative capstone histones compared to the noncapstone in each species (Figure 6.12). This trend is weaker for the capstone in *Candidatus Nanosalina* sp. J07AB43 (Figure 6.12c).

This suggests that the functional properties of capstone histone variants are conserved, and may have arisen convergently multiple times throughout archaea through a single substitution at a conserved position in the HFD.

6.8. Summary

The *in silico* modelling of *M. stadtmanae* histones highlights the potential for diversity of histone-based chromatin in this species, as well as the ability to regulate chromatin structural properties through incorporation of different paralogs (Stevens *et al.* 2020). One histone variant in particular, Msp_0383, may be able to impose a more radical change to chromatin architecture by preventing extension of histone oligomeric complexes. These findings indicate that some archaeal histones have evolved distinct structural properties.

Looking at the protein sequence of histones across species of archaea, the H49E substitution in Msp_0383 does not appear to be an isolated case and there are other archaeal histones with substitutions at this site. The presence of capstones outside the Methanobacteriales and phyletic pattern of the capstone across archaea also suggests that histones with capstone functionality have evolved multiple times independently in separate lineages (Stevens *et al.* 2020). The capstone is likely easy to evolve, requiring just a single amino acid change at position 49, as revealed by MD simulations for histones from different species. It is worthwhile noting here that there are other sites at the tetramerisation interface which may similarly alter the ability of histones to form stable tetramer interfaces and there may therefore be other sites for substitutions which can generate histones with capstone-like functionality. Archaea without capstones may also disrupt or regulate histone oligomerisation through other NAPs which may outcompete histones for the DNA.

7. Conservation of histone paralogs in Methanobacteriales

7.1. Introduction

Aside from functionally defining variants in a given species, sequence-based analysis can enable the definition of variants in a phylogenetic context. The relationship between, and evolution of, histone paralogs can be studied because the histone fold is amenable to diversification and their function is evolvable. In eukaryotes, phylogenetic analysis has been used extensively to identify likely scenarios for how histone variants evolved (Molaro and Drinnenberg 2018). Protein-level trees built from multiple sequence alignments containing sequences from different organisms can be used to infer the relationship between paralogs (Molaro and Drinnenberg 2018). Using this approach, some variants appear to have a single, ancient origin, conserved functions and dedicated chaperones (e.g. H2A.Z) whereas others are polyphyletic, arising multiple times independently (e.g. H2A.X, see section 2.4.2) (Talbert and Henikoff 2010).

The presence of more than one histone in many genomes of archaea (Figure 2.6) raises the question of paralogy. Is there any evidence for paralogous groups of histones that have an ancient origin, or is diversification of variants more recent and lineage specific?

The trees in this Chapter were built by A. Hocher, and were analysed as a joint effort.

7.2. Methanobacteriales archaea have many histones

The order Methanobacteriales (which belongs to the Methanobacteria Class) is a particularly interesting case study as species have some of the highest numbers of histones, with many more sequenced representative species than in other groups (Figure 2.6). Applying phylogenetic approaches to histones in related species in this Order (which includes the genera Methanobacterium, Methanosphaera and Methanobrevibacter and species *M. stadtmanae*, *M. smithii* and *M. fervidus*) allows us to infer relationships between histones.

All genomes in our Methanobacteriales dataset encode at least three histone paralogs with the exception of *M. fervidus* which has two (Figure 7.1). Based on this, it is parsimonious to think

that the common ancestor of the Methanobacteriales also had three histones, and that the three paralogs in extant species are ancient and shared between genera.

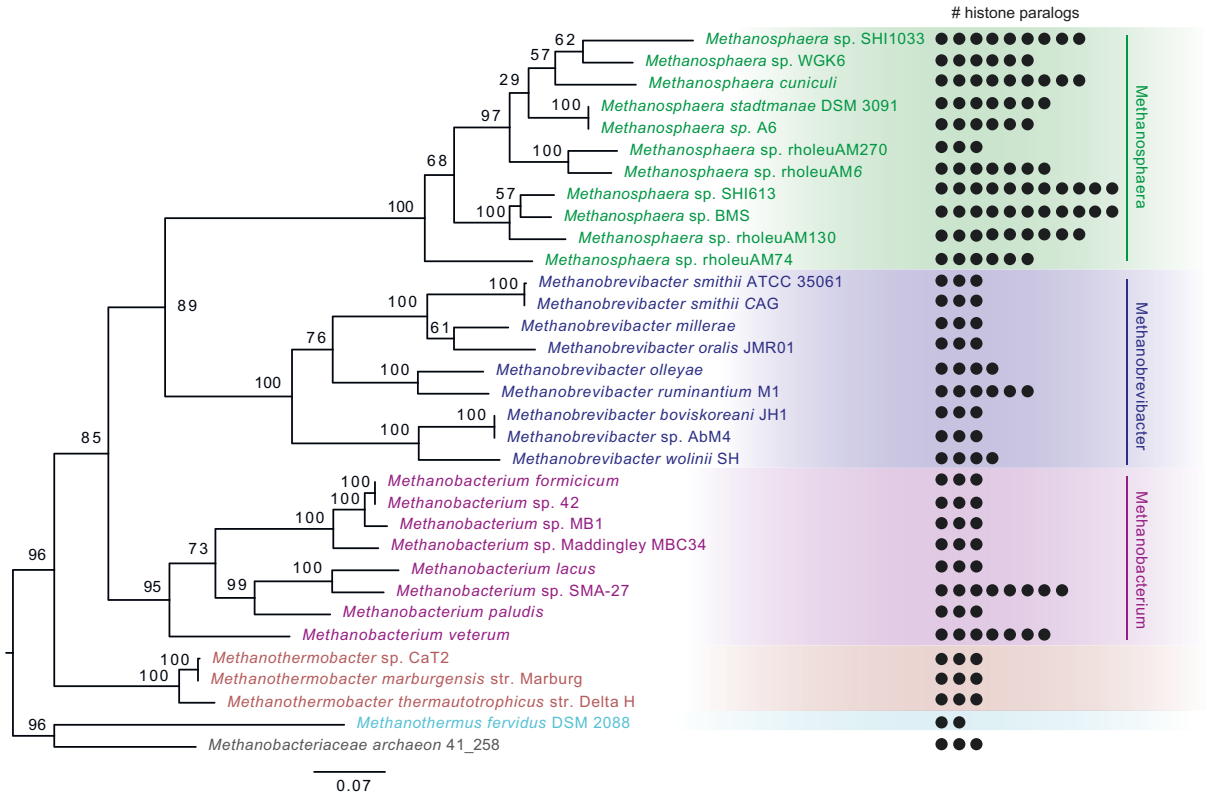


Figure 7.1 Species tree of the order Methanobacteriales showing the number of histone paralogs in each genome.
The maximum likelihood reference phylogeny using IF-2a is shown. 200 non-parametric bootstraps were run, and values are shown as a percentage. *M. fervidus* was used as an outgroup to root the tree. For each genome, the number of histone paralogs is shown on the right-hand side as circles where each circle represents a gene encoding a histone. This figure was produced by A. Hocher and has been published in (Stevens *et al.* 2020)

Does the hypothesis for ancient paralogy for three paralogs hold true? This can be looked at in a number of ways, using both protein-level trees as well as probing synteny.

If paralogs are shared between species with an ancient origin, these shared, ancient paralogs from different species would be expected to branch as paralogous groups as opposed to by species (Figure 7.2a). In reality, deciphering the shared history different proteins using trees alone can be complex. This is particularly true for archaeal histones, which are short (~60-80 amino acids in length, Figure 5.1a) often resulting in poor bootstrap results. Another parallel way to look at this problem is through synteny. Traces of the evolutionary history of shared paralogs can often be found in the neighbourhood of surrounding genes. In eukaryotes, genes found in the syntenic regions of some histone variants are identical, helping to reinforce

phylogenetic relationships (Molaro and Drinnenberg 2018). Mammalian short H2A variants are one such example of this where orthologs of H2A.B and H2A.L could be identified in many mammalian genomes by their synteny (Molaro *et al.* 2018). In this case, despite their shared history, over time, orthologous short H2As within species have been subject to many gene conversion events resulting in orthologs clustering into species-specific branches on protein-level trees (Molaro *et al.* 2018).

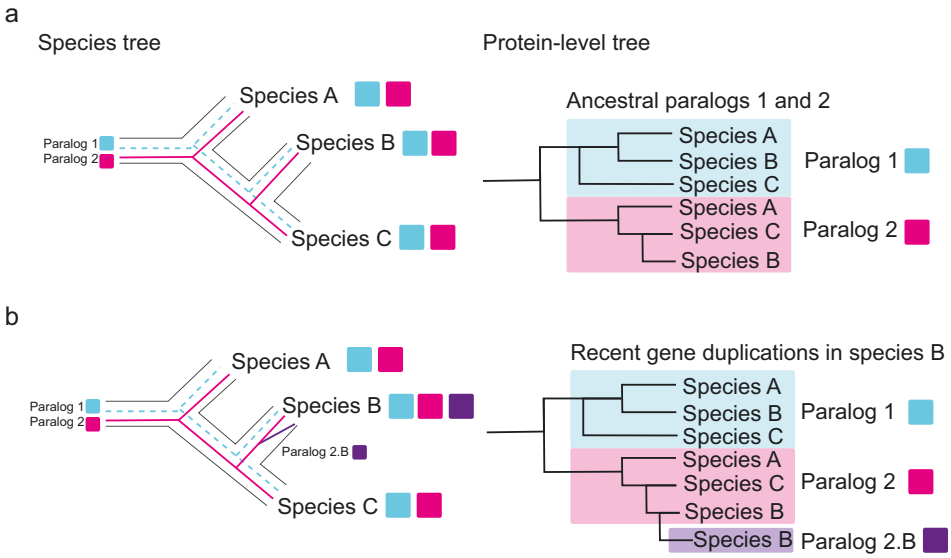


Figure 7.2 Toy species- and protein-level trees for relationships between histones in archaea.
In related species with multiple histone paralogs, two scenarios are shown. Species A, B and C all have a) two paralogs (1 and 2) which are ancestral, shown by paralogs grouping by paralog number, rather than species in protein level tree. b) In species B, a third paralog (2.B) branches with paralog 2 from the same species (and is not present in species A or C), suggestive of a more recent duplication of paralog 2 in the branch leading to species B.

7.3. There are three ancient histone paralogs in the Methanobacteriales

Protein-level trees containing histones from Methanobacteriales species were built by A. Hocher and analysed in conjunction with local synteny to explore whether the apparent ancient paralogy holds true. Three groups of paralogs could be identified based on shared local synteny and phyletic pattern (Figure 7.3a, b). In paralogous groups 1 and 3, there is shared synteny across Methanobacterium and Methanobrevibacter (Figure 7.3a, b). Even deeper paralogy for some groups is suggested by looking at the local region around histones from the more distantly related *M. fervidus*. Upstream of HMfB is *trpS*, which is also found in group 1 (Figure 7.3a). Downstream of HMfA, is *radB* similar to paralog group 2 (Figure 7.3a). Clear evidence for paralog group 2 based on synteny and branching patterns could not be found for Methanobrevibacter spp. (Figure 7.3b).

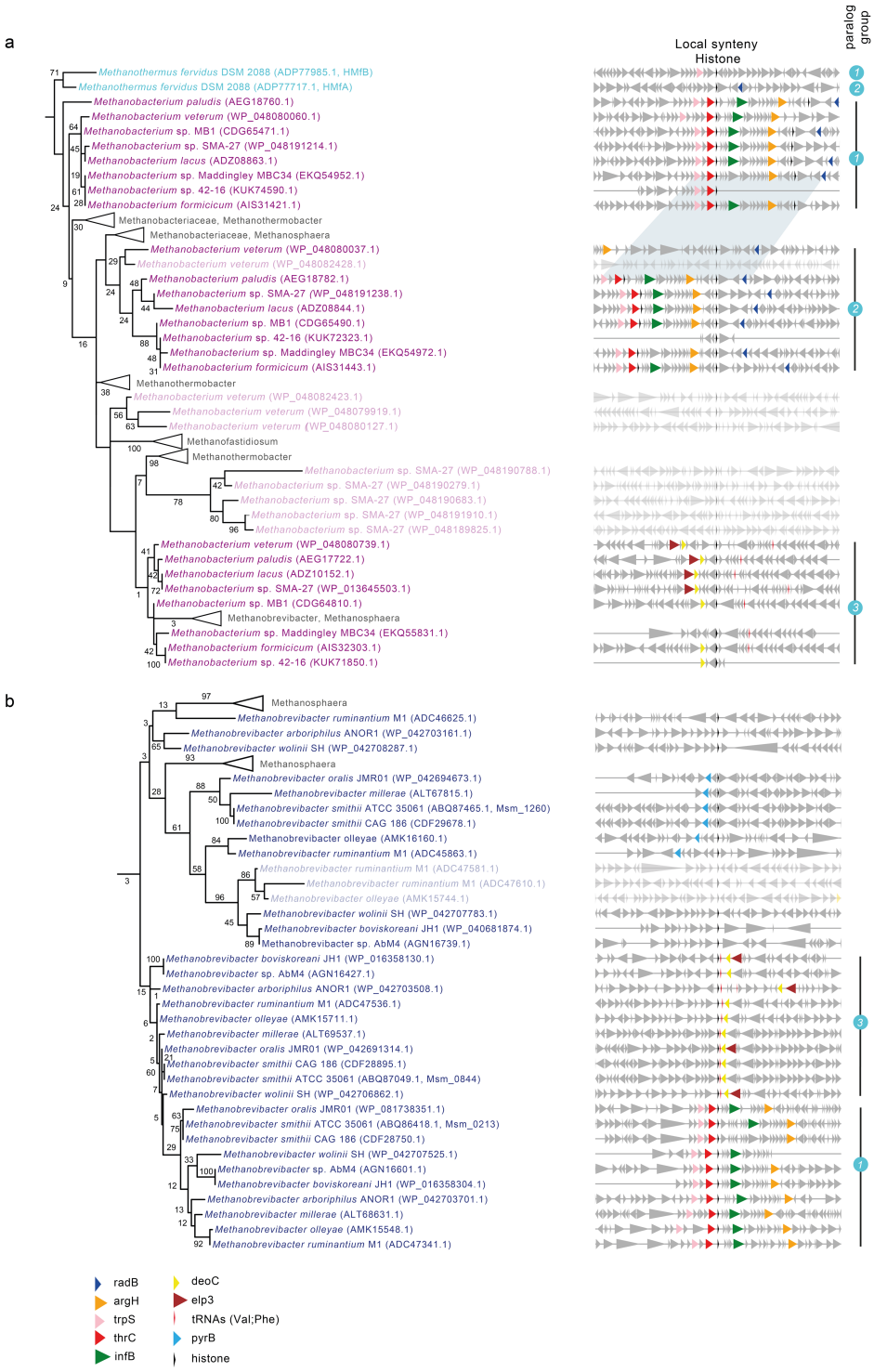


Figure 7.3 Long-term maintenance of histone paralogs in Methanobacteriales.
Protein-level trees showing recent duplications (light shaded taxa) and long-term maintenance of paralogs in the genera a) Methanobacterium, and b) Methanobrevibacter. Sequences from other Methanobacteriales have been collapsed: Methanosphaera in a), Methanofastidiosum, Methanothermobacter, Methanobrevibacter and Methanosphaera in b). On the right-hand side of the tree, the genes in the local neighbourhood of the histone are shown. For each tip of the tree, the syntenic context is centred on the histone for that tip. Genes are colour coded based on functional annotation. Three paralog groups are highlighted on the right-hand side of the synteny plot. Trees are rooted with reference to a larger tree of archaeal histones. Bootstrap values are shown as a percentage. In some cases, there are gaps in synteny due to incomplete annotations. This figure was produced by A. Hocher and has been published in (Stevens *et al.* 2020).

Shared synteny and phylogenetic clustering are suggestive of vertical inheritance of the three histone paralogs in the Methanobacteriales. If this is the case, the branching order of histones in paralog groups should follow the topology of the species tree. To compare the species tree to the protein-level tree, subtrees for each paralog group were taken from the larger tree (Figure 7.4). Branches with low support were removed (< 80%) as limited phylogenetic signal likely results in some discrepancies, and the resulting branching events were compared to the species tree.

In all three paralog groups, histones recapitulate the species history (Figure 7.4). There is one exception, a single bifurcation leading to *Methanothermobacter* sp. CaT2 and *Methanothermobacter thermautotrophicus* str. Delta H (Figure 7.4, shown in red). Aside from this case, the agreement between species and protein trees provides further support for vertical descent of each paralog in Methanobacteriales.

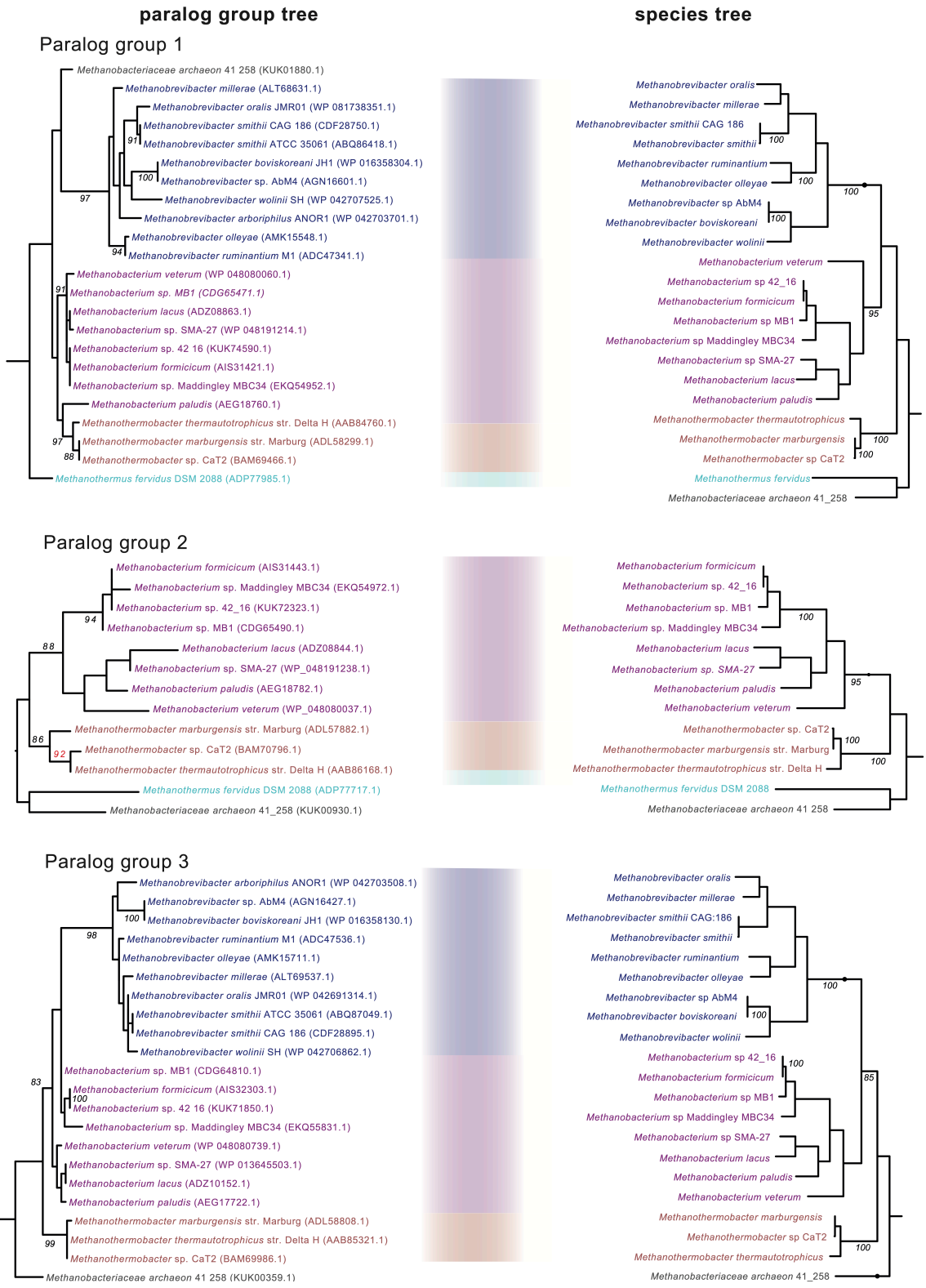


Figure 7.4 Comparison of protein-level tree and species tree for Methanobacteriales paralog groups. The protein tree for high support nodes for each paralog is shown on the left-hand side. The species tree for the Methanobacteriales species is shown on the right. Shading corresponds to branches containing species groups. The one branch in paralog group 2 which does not follow the species tree is highlighted in red. This figure was produced by A. Hocher and has been published in (Stevens *et al.* 2020).

7.4. Some Methanobacteriales species have an increased number of histones

Aside from the three paralogs, which appear to be shared, an expansion in the number of histones can be seen in some Methanobacteriales. For example, Methanosphaera genomes tend to have many more paralogs (> seven in most cases) than other genera (Figure 7.1). There are also more recent expansions occurring between closely related species. In the genus Methanobacterium, most genomes only have three histone paralogs, but in two genomes (*Methanobacterium veterum* and *Methanobacterium* sp. SMA-27), there are many more (seven and eight paralogs each for *M. veterum* and *M.* sp. SMA-27, respectively).

How can this gain of histone number be explained? There are several ways this could arise in theory; by horizontal gene transfer (HGT), recent duplications, or *de novo*. The birth of new histone genes in all these species is unlikely. There are only a few examples of *de novo* gene birth (Van Oss and Carvunis 2019) and, despite some differences, there are shared sequence features between histones within archaea. A more likely explanation could be either HGT or gene duplication. HGT plays a major role in increasing the repertoire of proteins in prokaryotes (Treangen and Rocha 2011). Duplication of genes (through mechanisms such as unequal crossing over or whole genome duplication) is also an important driving force, generating novel gene functions, and resulting in subfunctionalisation of duplicated genes (Magadum *et al.* 2013). This is thought to play a smaller role in generating novel proteins in prokaryotes than HGT (Treangen and Rocha 2011) which appears to be responsible for the gain of some NAP families in archaea (Hocher *et al.* 2021). In the case of recent duplication events followed by diversification, the emergence of one paralog from the duplicated ancestral paralogs would be more closely related to one another than paralogs across species, and would be expected to branch in a species-specific manner (Figure 7.2b).

In the Methanobacteriales, there are some cases where histones from the same species branch together, exemplified for the species where we find an increased number of histone paralogs (Figure 7.3a, b, shaded taxon labels). This is reminiscent of the toy model for recent duplication events (Figure 7.2b). For example, an increase in copy number of histones in *Methanobacterium veterum* results in two groups of paralogs branching together (Figure 7.3a, shaded taxon labels). Similarly, *Methanobrevibacter ruminantium* M1 and *Methanobrevibacter olleyae* have a higher-than-average number of histones for

Methanobrevibacter spp. (Figure 7.1). Histones in these species branch close to other paralogs from the same species indicative of recent duplication events (Figure 7.3b). In these cases, the histones appear to have differences in their genomic contexts (Figure 7.3). One explanation of histone gene duplication into divergent genomic contexts is a loss of synteny surrounding each gene after the duplication event. If these histones are also found in more closely related strains or species, comparing the genomic context may help aid the understanding of the timing and sequence of events leading to this apparent duplication; particularly in the *Methanobacterium* if the histones in a species have genomic contexts more similar to one another. Another possibility is that this expansion arose from HGT followed by recombination between paralogs within species, which could also account for clustering by species.

7.5. There is little evidence for recombination in Methanobacteriales histones

Eukaryotic histones have been suggested to be frequently subject to gene conversion and rearrangement events (DeBry 1998, Galtier 2003, Scienski *et al.* 2015). Are the histones in the Methanobacteriales also subject to gene conversion? Eukaryotic histones are generally arranged in tandem repeat cluster, making them more prone to gene conversion. In contrast, archaeal histones are generally found as single gene operons and multiple paralogs are not organised in tandem arrangements along the genome making gene conversion less likely (e.g., in *M. stadtmanae*, see section 6.2). To confirm gene conversion is unlikely to play a major role in the evolution of Methanobacteriales histones, several approaches to detect recombination events were used. As sequences of the histones are short, and therefore may be sensitive to the choice of parameters and algorithm, I used multiple tools (Figure 7.5). This included: GARD from DataMonkey (Kosakovsky Pond *et al.* 2006), the pairwise homoplasmy index (PHI) using the PhiPack package (Bruen *et al.* 2006) , which has previously been used to detect gene conversion in archaea and bacteria (Lassalle *et al.* 2015), and multiple metrics from RDP4 (Martin *et al.* 2015), which has been used to show recombination events in single genes (Esin *et al.* 2018).

There is no evidence for recombination or gene conversion when either GARD or PHI is used (Figure 7.5). There was some evidence for recombination from a few of the nine methods (primarily SciScan) computed in RDP4, but no consensus across methods (Figure 7.5– significant events shown in black, Table 7.1). All seven of these events were labelled as

tentative in RDP4, suggesting the signal may not come from true recombination events. Overall, there is no strong evidence for gene conversion at histones genes in the Methanobacteriales.

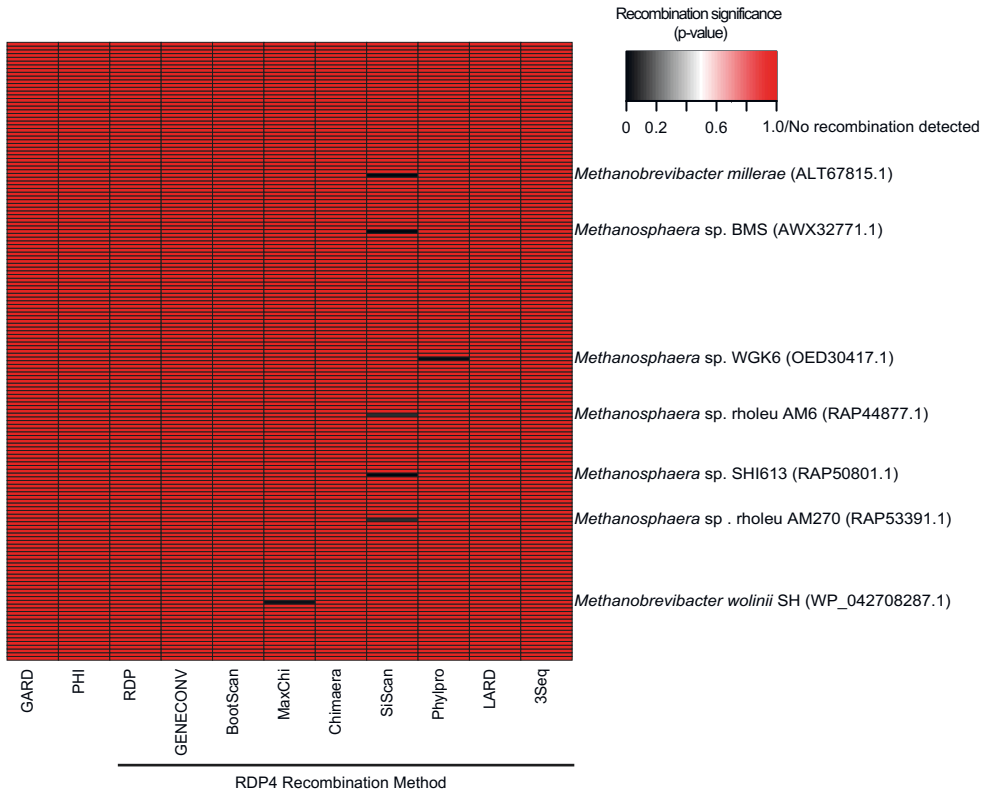


Figure 7.5 Detection of recombination events in the Methanobacteriales.
Algorithms from RDP4, GARD and PHI were used to detect potential gene conversion events. Potential recombination events meeting the significant threshold (P< 0.05) are shown in black (see Table 7.1). This figure was adapted from (Stevens *et al.* 2020).

Despite the weak signal, I reviewed these potential recombination events in the context of the phylogenetic trees. Taking these events into account does not change the interpretation of the paralog trees. Most proposed events are recent (Table 7.1), and tend to only affect a small part of the sequence.

Table 7.1 Significant recombination events detected by RDP4 in the Methanobacteriales. This table was adapted from (Stevens *et al.* 2020).

RDP4-detected recombinant events

Recombination event	Sequence(s) with recombination signal	Recombinant source sequence	Significance of detection								
			RDP	GENECONV	BootScan	MaxChi	Chimaera	SiScan	Phylpro	LARD	3Seq
1	Methanosphaera sp. BMS AWX32771.1 Methanosphaera sp. SHI613 RAP50801.1	Methanobacterium veterum WP_048082423.1	NS	NS	NS	NS	NS	6.73E-03	NS	NS	NS
2	Methanobrevibacter millerae ALT67815.1	Methanothermobacter marburgensis str. Marburg ADL57882.1	NS	NS	NS	NS	NS	2.90E-03	NS	NS	NS
3	Methanosphaera sp. rholeuAM6 RAP44877.1 Methanosphaera sp. rholeuAM270 RAP53391.1	Methanosphaera sp. WGK6 OED30333.1	NS	NS	NS	NS	NS	8.30E-02	NS	NS	NS
4	Methanosphaera sp. WGK6 OED30417.1	Unknown	NS	NS	NS	NS	NS	NS	1.36E-02	NS	NS
5	Methanobrevibacter wolinii SH WP_042708287.1	Methanosphaera sp. rholeuAM74 RAP50195.1	NS	NS	NS	2.33E-02	NS	NS	NS	NS	NS

7.6. Phylogenetic analysis of capstone histones in Methanobacteriales

Histones in the Methanobacteriales includes the structurally diverse capstone variant. Notably, the modelling results suggest at least two species have the capstone; both *M. stadtmanae* and *M. smithii* (see Chapter 6). What is the evolutionary history of this variant in the Methanobacteriales? Do we find evidence for shared vertical descent of the capstone histone variant in the Methanobacteriales?

The identity of each amino acid at position 49 (the capstone site, see Chapter 6) was mapped onto a protein-level tree containing all histones from the Methanobacteriales. Those with negative or hydrophobic amino acids at position 49 site cluster together to the exclusion of other sequences (Figure 7.6). These putative capstone histones are found in Methanosphaera and Methanobrevibacter *spp*. but not in Methanobacterium species. In one Methanosphaera species (*Methanosphaera* sp. rholeuAM6) there is an apparent reversion back to a positively charged residue (lysine, K) at position 49.

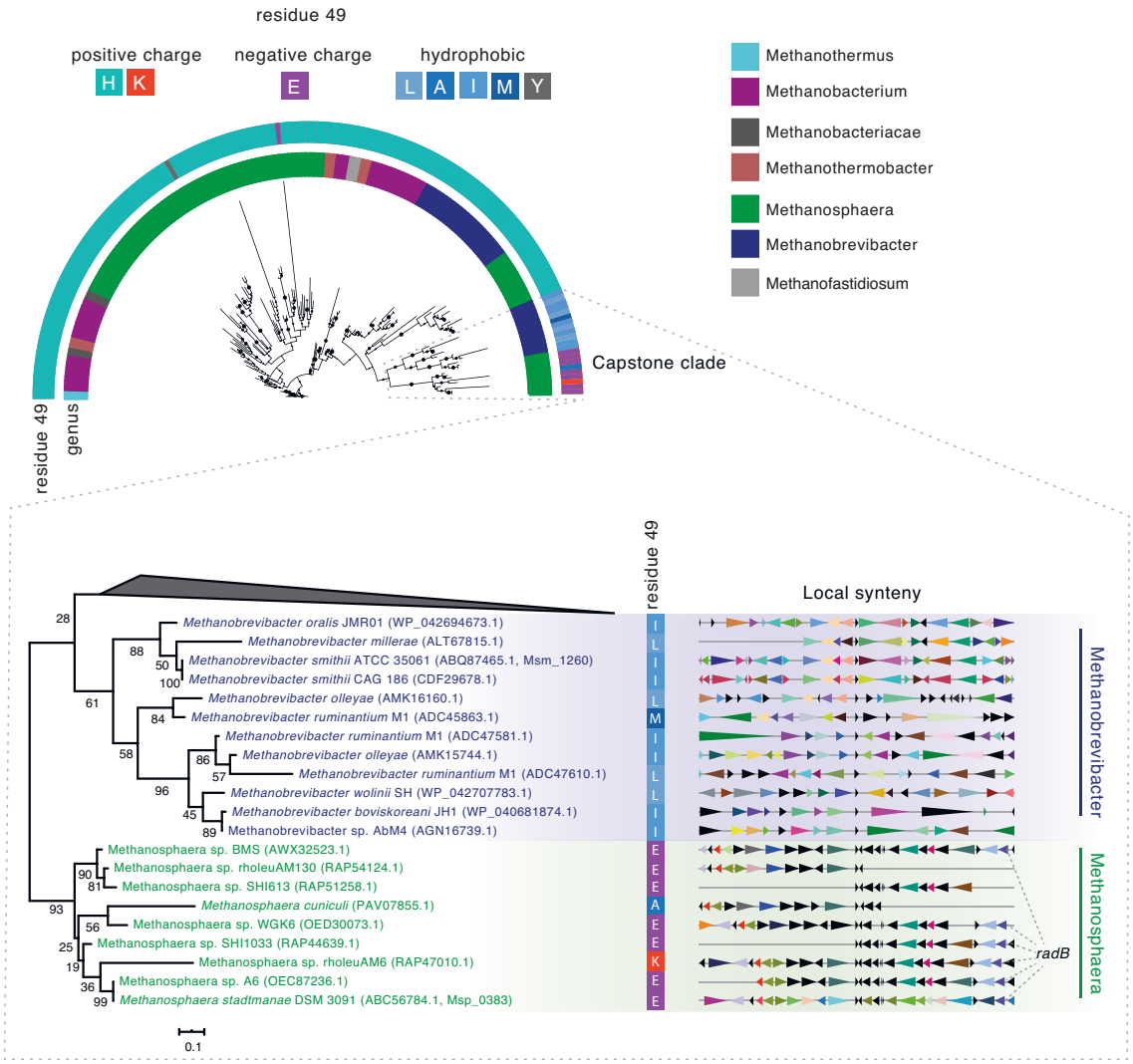


Figure 7.6 Phylogenetic analysis of potential capstone histones in Methanobacteriales.
On the top, a phylogenetic tree of histones from all Methanobacteriales genomes is shown. The amino acid found at position 49 (position number relative to HMfB, site of capstone identity) is shown for each histone. At the bottom, the local synteny for putative capstone histones in the Methanobacteriales is shown. Genes are coloured by gene annotation available in downloaded gff files, and radB is highlighted. This figure was produced by A. Hocher and is adapted from (Stevens *et al.* 2020).

Common genes can be found in the region neighbouring the capstone for both the Methanobrevibacter and Methanosphaera lineages. In particular, *radB* is found downstream of almost all paralogs with the H49 substitution in Methanosphaera species (Figure 7.6). The presence of the capstone histone in multiple Methanobrevibacter and Methanosphaera species, alongside shared synteny, is suggestive of vertical descent of the capstone in these lineages.

7.7. Summary

In the Methanobacteriales, histones cluster into paralogous groups which can be defined by phylogenetic positioning and shared local synteny. Three paralog groups could be defined in this manner, with results suggesting shared paralogy across species. Phylogenetic and syntenic analysis provide support for vertical inheritance of ancient paralogs present in the last common ancestor of the Methanobacteriales, and that gene conversion has played little role, if any, in the evolution of each histone. Conservation of paralogs across this clade includes the capstone histone, which is found in both Methanosphaera and Methanobrevibacter *spp*., the phyletic pattern of which supports vertical inheritance of this variant in each genus.

The evolutionary persistence of specific histone variants in the Methanobacteriales is on par with that of eukaryotic histone variants (Stevens *et al.* 2020). The last common ancestor of eukaryotes, which had some universal histone variants such as H2A.Z, existed around 1.2-2 Gya (Eme *et al.* 2014, Betts *et al.* 2018). In comparison, the Methanobacteriales clade arose around 1.6 Gya, and the two lineages where we find capstones (represented by *M. stadtmanae* and *M. smithii*) diverged ~1.3 Gya (Wolfe and Fournier 2018). Although the exact *in vivo* roles of these paralogs remain unknown, this conservation further supports the presence of important functional differences between histone paralogs in Methanobacteriales species, something which requires further investigation.

8. Conservation of histone paralogs in Thermococcales

8.1. Introduction

Evidence in Chapter 7 suggests that some histone paralogs in the Methanobacteriales are conserved across species separated by over a billion years (Stevens *et al.* 2022). Is the case of ancient paralogy in the Methanobacteriales unique, or are there other examples in distantly related archaeal lineages?

To identify interesting groups, a large CDS (coding sequence)-level tree of histone sequences from species present in the structural analysis was built (Chapter 5, see Methods). Automated phylogenetic analysis across this large set of archaeal species is difficult. Most histones are short (< 80 amino acids for single HFD histones), predisposing the phylogeny to low bootstrap values. The timescales since species diverged from one another are also large, reducing confidence when looking across clades (Wolfe and Fournier 2018). Looking within groups, rather than across all species, can provide better insight. One other interesting case could be identified by combining branching patterns and conserved synteny as done in Chapter 7.

Based on this larger phylogeny, the Thermococcales, which includes Pyrococcus, Paleococcus and Thermococcus species, were identified as a potentially interesting second cases study. Notably, this group contains *T. kodakarensis*, a hyperthermophilic euryarchaeon which grows at an optimum temperature of 85ºC, and can survive between temperature ranges of 60 ºC to 100ºC (Spaans *et al.* 2015). Unlike many other species of archaea, we know a little about chromatin architecture in this species. This includes the properties of the two histones and other NAPs, having been extensively studied in particular by Thomas Santangelo and colleagues.

T. kodakarensis is highly polyploid, exhibiting changes to chromosome copy number in a growth-phase dependent manner, ranging from 7 to ~19 copies (Spaans *et al.* 2015). Copy number is high in early exponential phase (~19), and decreases to around 7 in stationary phase. The genome of *T. kodakarensis* encodes many different NAPs which can bind to and shape the DNA. This includes histone proteins, which can coat the entire genome even at high copy numbers due to their high levels of expression (Spaans *et al.* 2015, Sanders, Lammers, *et al.* 2019).

T. kodakarensis has two histone proteins, HTkA (TK1413) and HTkB (TK22989) which are 84% identical at the protein sequence level (differing at 11 out of 67 amino acid positions). These histones can form homo- and heterodimers, and bind to DNA as tetrameric complexes, which can then assemble into longer oligomeric complex structures both *in vivo* and *in vitro* in a manner akin to HMf histones from *M. fervidus* (Figure 2.5) (Maruyama *et al.* 2013, Nalabothula *et al.* 2013, Mattiroli *et al.* 2017, Bowerman *et al.* 2021, Sanders *et al.* 2021).

The function of HTkA and HTkB appear to overlap with one another. Both histones cannot be deleted simultaneously, but strains with either HTkA or HTkB deleted can be constructed, suggesting their roles intersect (Čuboňová *et al.* 2012, Mattiroli *et al.* 2017). The deletion of either *htkA* or *htkB* results in gene expression changes, with overlaps in transcriptome alterations (Čuboňová *et al.* 2012, Sanders *et al.* 2021).

Given the similarities between the two, is there any evidence these two histones could be considered histone variants, and what, if any, are the differences between them? From prior work, we know that HTkA and HTkB do display differences from one another. Cells with *htkB* deleted show changes in expression of genes encoding ribosomal proteins and translation factors, potentially contributing to the reduced growth rate in these cells (Čuboňová *et al.* 2012). Strains deficient in HTkA can no longer be transformed, and exhibit decreased expression of genes encoding membrane proteins (Čuboňová *et al.* 2012). Therefore HTkA, and not HTkB, may play an important role in DNA integration and/or DNA uptake (Čuboňová *et al.* 2012). Each histone also shows differences in relative expression levels: in exponential phase, HTkA contributes 1.1% of the proteome, compared to 0.66% for HTkB (Hocher *et al.* 2021). Comparing the DNA binding properties between the two histones revealed further differences between the two histones. HTkB binds DNA more tightly than HTkA, and forms more compact protein-DNA complexes, as shown by faster migration in agarose gel electrophoresis (Higashibata *et al.* 1999).

Based on the above, it appears the two histones have overlapping, but distinct, functions. But are the differences between HTkA and HTkB important and physiologically relevant for *T. kodakarensis* cells, and enough to consider these two histones as variants? One way to approach this question is from an evolutionary stand point: to ask whether the two paralogs appear to be important and conserved as distinct entities. To see whether this was the case, I considered the structural and sequence properties of histones in the order Thermococcales.

8.2. Identifying histone paralogs with HTkA-like and HTkB-like properties in Thermococcales species

Histone proteins in Thermococcales were identified by searching 61 proteomes using HMM models and jackhammer searches (see Methods). This revealed that histones consisting of a singlet histone fold domain, similar to canonical archaeal HMf histones (e.g. those similar to HMfB in *M. fervidus* (Sandman *et al.* 1990)), are found in all Thermococcales species (Figure 8.1). Alongside the HMf-like histones I discovered shorter bacteria-type histones, and longer bacteria-type doublet histones in many Thermococcales (doublet refers to an end to end duplication of the histone fold domain, see section 2.8.1.1), consistent with previous findings (Alva and Lupas 2019) (Figure 8.1). The distribution of these bacteria-type histones is notably much patchier than that of the HMf-like histones (discussed further below).

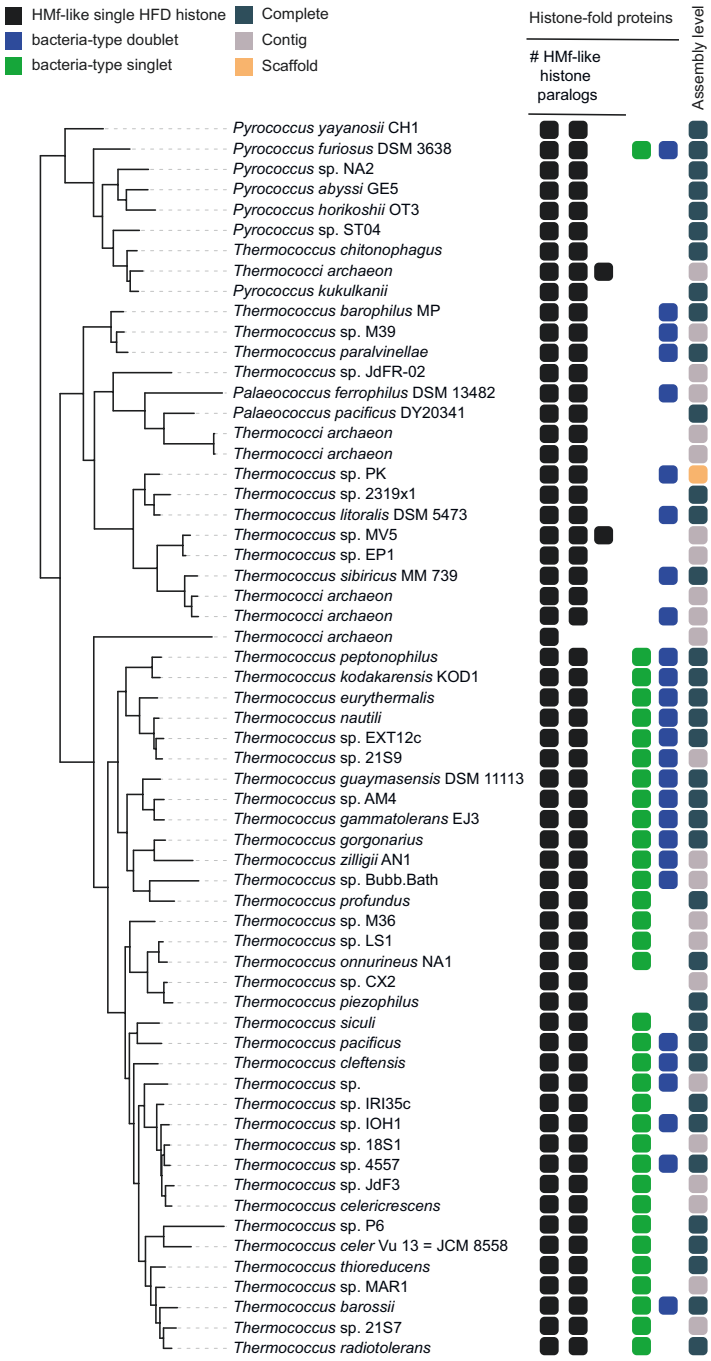


Figure 8.1 Number of histones in Thermococcales genomes.
GTDB species tree of Thermococcales archaea showing the number of HMf-like single HF proteins for each genome is shown on the right-hand side as black squares. Bacteria-type singlet and bacteria-type doublet histones are indicated as green and blue squares, respectively. The assembly level for each species is shown. Figure adapted from (Stevens *et al.* 2022).

All but one species, for which the genome is not complete, have two (and in two instances three) HMf-like histones. This is parsimonious with the common ancestor having two histones, similar to the pattern of three found in Methanobacteriales. Is there evidence for shared paralogy in extant Thermococcales genomes?

As we know the sequence differences between HTkA and HTkB, I first looked at the sequence properties of all HMf-like histones. Clustering these sequences by amino acid properties revealed two groups separating on the first principal component: one containing HTkA (designated HTkA-like) and one containing HTkB (designated HTkB-like) (Figure 8.2a). The separation of histones in Thermococcales into two groups is comparable with previous classification of histones in Thermococcus spp. based on protein sequences (Henneman 2019). Consistent with (Henneman 2019), I also find that some sites are conserved and diagnostic of histones in these two groups. For example, HTkA-like histones generally have E (glutamic acid) at position 18 (59/61 sequences), and Y (tyrosine) at positions 31 and 35, whereas HTkB-like histones have swapped the negatively charged E at position 18 for A (alanine), P (proline) or Q (glutamine), and Y at positions 31 and 35 have been substituted for positively charged amino acids H (histidine) and K (lysine) (60/62, and 2/62 sequences), respectively (Figure 8.2b). Note that position numberings here are relative to the HMfB sequence. The end of the alpha two helix is also more variable in HTkA-like histones.

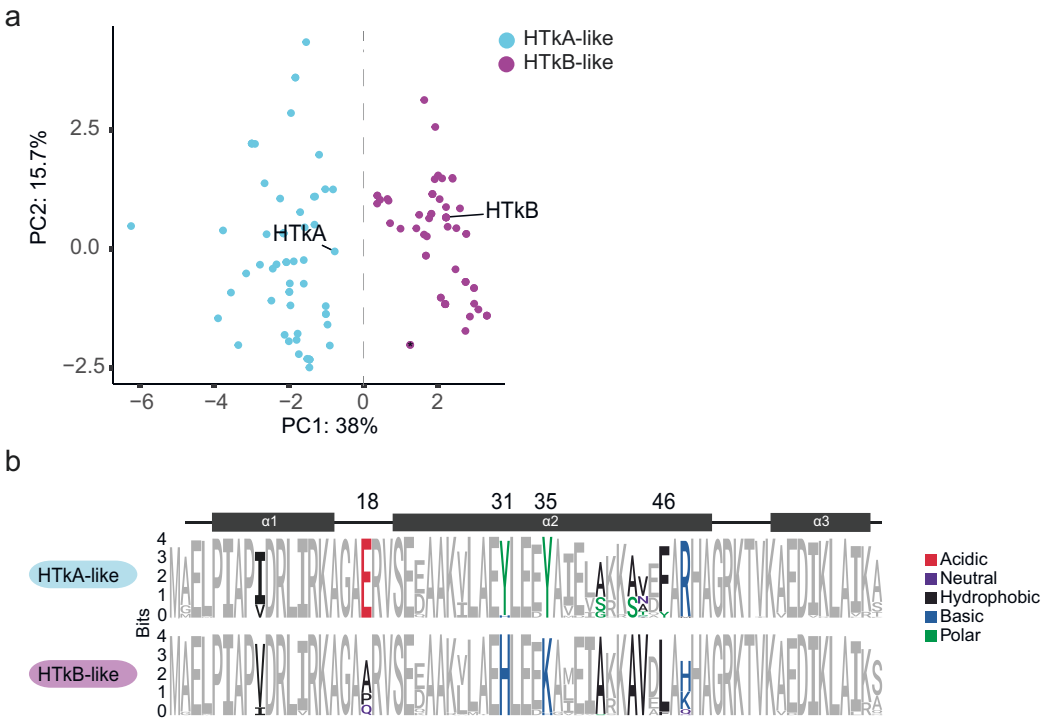


Figure 8.2 Identification of HTkA-like and HTkB-like histones in Thermococcales.
a) PCA clustering of HMf-like histones from Thermococcales spp. based on amino acid properties. HTkA and HTkB from *Thermococcus kodakarensis* are labelled. Histones are coloured by group – those which cluster with HTkA are HTkA-like, and those with HTkB are HTkB-like. b) Weblogo showing sequences of HTkA and HTkB-like histones as bit score. Weblogos were plotted using ggseqlogo (Wagih 2017). Residues which discriminate each group are coloured by amino acid property. The secondary structure of the histone is shown above (a1, a2 and a3 are alpha helices 1, 2 and 3, respectively). Figure adapted from (Stevens *et al.* 2022).

Some of these signature substitutions, in particular positions 18 and 35, have been characterised from structural modelling and previous *in vitro* experiments. Much of this mutation work has been performed with the histone HMfB from *M. fervidus*. Although HMfB does differ substantially from HTkA/B histones (see below), this can be used to infer some of the potential structural ramifications which may arise from individual substitutions at conserved sites. In HMfB, substitution of E at position 18, similar to HTkA-like histones, for a P (found at position 18 in some HTkB-like histones) has no effect on DNA binding affinity (Soares *et al.* 2000). This may suggest that differences between HTkA-like and HTkB-like histones at this position do not directly alter DNA binding. However, E18 and K53 form an inter-molecular salt bridge in HMfB which stabilises the interactions between monomers in a dimer (Decanniere *et al.* 2000, Sandman *et al.* 2001). Loss of this stabilising interaction may result in lower stability or reduced rigidity for HTkB-like histone dimers compared to HTkA-like histones. Another discriminatory site, position 46, is known to be involved in tetramerization (Soares *et al.* 2000, Marc *et al.* 2002). However, there is no evidence for differences in DNA binding in HMfB

histones with L46F (phenylalanine) substitution (Soares *et al.* 2000, Marc *et al.* 2002). A substitution of K at position 35 for Y (similar to HTkB-like v HTkA-like histones) in *M. formicicum* histone HFoB increased the overall stability of this histone *in vitro* (Li, Shriver, *et al.* 2000), suggesting that this may result in a lower overall stability of HTkA-like histones.

Can we determine if any of these substitutions will result in conserved structural differences between the two groups? From previous *in vitro* experiments, *T. kodakarensis* histones HTkA and HTkB have been shown to exhibit differences in their DNA binding affinities, with HTkB binding more strongly and forming more compact complexes (Higashibata *et al.* 1999) (see section 8.1). To determine whether conserved sequence features in HTkA-like and HTkB-like proteins confer similar structural properties, structural modelling was used to assess predicted DNA binding and tetramerization strengths for each histone. Using this approach, on average, HTkB-like histones are predicted to have stronger affinity for DNA than HTkA-like histones, similar to the *in vitro* results for *T. kodakarensis*, suggesting that histones in each group have conserved structural properties (Figure 8.3). On average, HTkA-like histones form stronger tetramers than HTkB-like histones, but this difference is weaker (Figure 8.3).

In some species, HTkA-like histones do not follow the trend of DNA binding, and are predicted to have stronger binding than the HTkB-like histone from the same species (Figure 8.3). There was no clear link to sequence differences for these cases; single sites in the protein sequence are not determinative of each group.

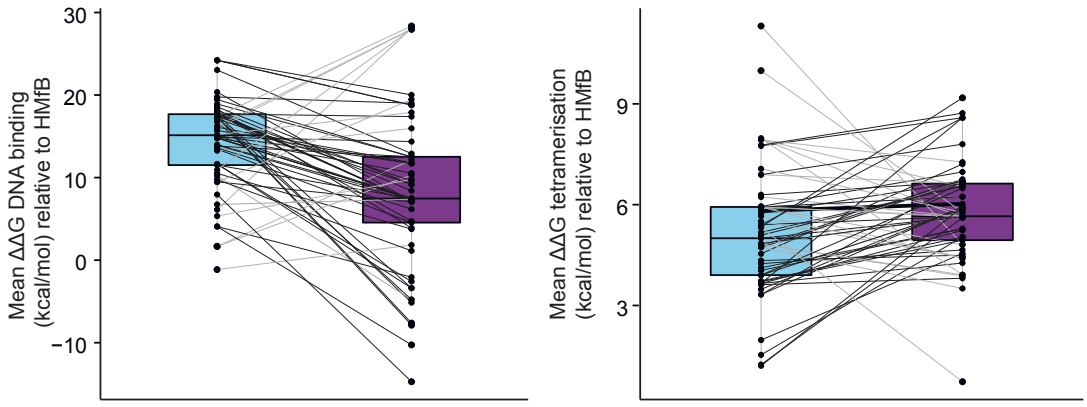


Figure 8.3 Predicted DNA binding and tetramerization strengths for Thermococcales HTkA- and HTkB-like histones.
Left: mean predicted DNA binding from five replicates for each HTkA and HTkB-like histone. Right: Mean predicted interaction strength between dimers from five replicates for each Thermococcales histone. Lines connect histones from the same species. Figure adapted from (Stevens *et al.* 2022).

Another way in which sequence differences between the two groups could confer distinct properties is through stacking interactions. Almost none of the residues identified as being important for HTkA or HTkB stacking interactions (K14, G17, K26, E30, E34, Q48, E58, K61, K65) for longer oligomer formation are discriminatory between HTkA-like and HTkB-like, with the exception of Q48 (Figure 8.2b) (Mattiroli *et al.* 2017, Henneman *et al.* 2018, 2021). Q48 has been suggested to mediate oligomer formation in HTkB, based on the structure (Henneman *et al.* 2018). However, in HTkB-like histones, Q48 is relatively rare. Therefore, HTkB may be an outlier compared to other HTkB-like histones, in that it may form relatively stronger oligomeric complexes with DNA.

There may be other functionally important changes which are not captured by structural modelling or DNA binding assays. For example, mass spectrometry results suggest the difference at position 35 may be particularly relevant. A lysine acetylation at this site was identified in *T. kodakarensis* HTkB. However, as HTkA has a tyrosine at this position, HTkA cannot be subject to lysine acetylation at this site (Figure 8.4) (Alpha-Bazin *et al.* 2021). Similarly, sequence differences in some eukaryotic histone variants can result in variant specific post-translational modification, as discussed in the introduction in Chapter 2. For example, H2A variant H2A.X has a unique motif at the C-terminus, of which the serine is a site of phosphorylation (Redon *et al.* 2002, Mah *et al.* 2010). Phosphorylation at this site occurs

at the location of double strand breaks on the DNA, and initiates a cascade of proteins which help to repair the damage at this site (Mah *et al.* 2010).

The positions of known lysine acetylation in the core of eukaryotic histones and *T. kodakarensis* or *T. gammatolerans* are not conserved. This suggests that the ramifications of this post-translational modification are likely substantially different between eukaryotes and archaea (Figure 8.4). There is currently no information about the stoichiometry or the functional role that K35ac plays in *T. kodakarensis*.



Figure 8.4 Comparison of lysine acetylation in eukaryotic histone H3, and Thermococcales histones. The secondary structure for histone H3 is shown at the bottom. Top: lysines subject to post-translational acetylation modification in the core histone fold of histone H3 in *S. cerevisiae* and *H. sapiens*, based on data from (Tessarz and Kouzarides 2014), are highlighted with a blue shaded box. *H. sapiens* histone H3 has lysines at positions 56 (Yuan *et al.* 2009), 64 (Di Cerbo *et al.* 2014) and 122 (Tropberger *et al.* 2013) which can be acetylated. Histone H3 in *S. cerevisiae* can be acetylated at lysines at positions 56 (Xu *et al.* 2005) and 79 (Bheda *et al.* 2012). Bottom: Lysines which are sites for post-translational acetylation in *T. kodakarensis* histones HTkA and HTkB are highlighted (Alpha-Bazin *et al.* 2021) in a blue box. Lysines which may be important for regulating hypernucleosome formation in HTkB through stacking interactions are indicated with an arrow (Henneman *et al.* 2018). This figure has been published in (Stevens and Warnecke 2022).

8.3. HTkA-like and HTkB-like histones are almost ubiquitous in Thermococcales

Based on sequence and structural properties, all Thermococcales species appear to have at least one HTkA-like and one HTkB-like histone (Figure 8.5). There are a few species with two histones belonging to the same group, for example *Thermococcus* sp. MV5 which has two HTkA-like histones (Figure 8.5).

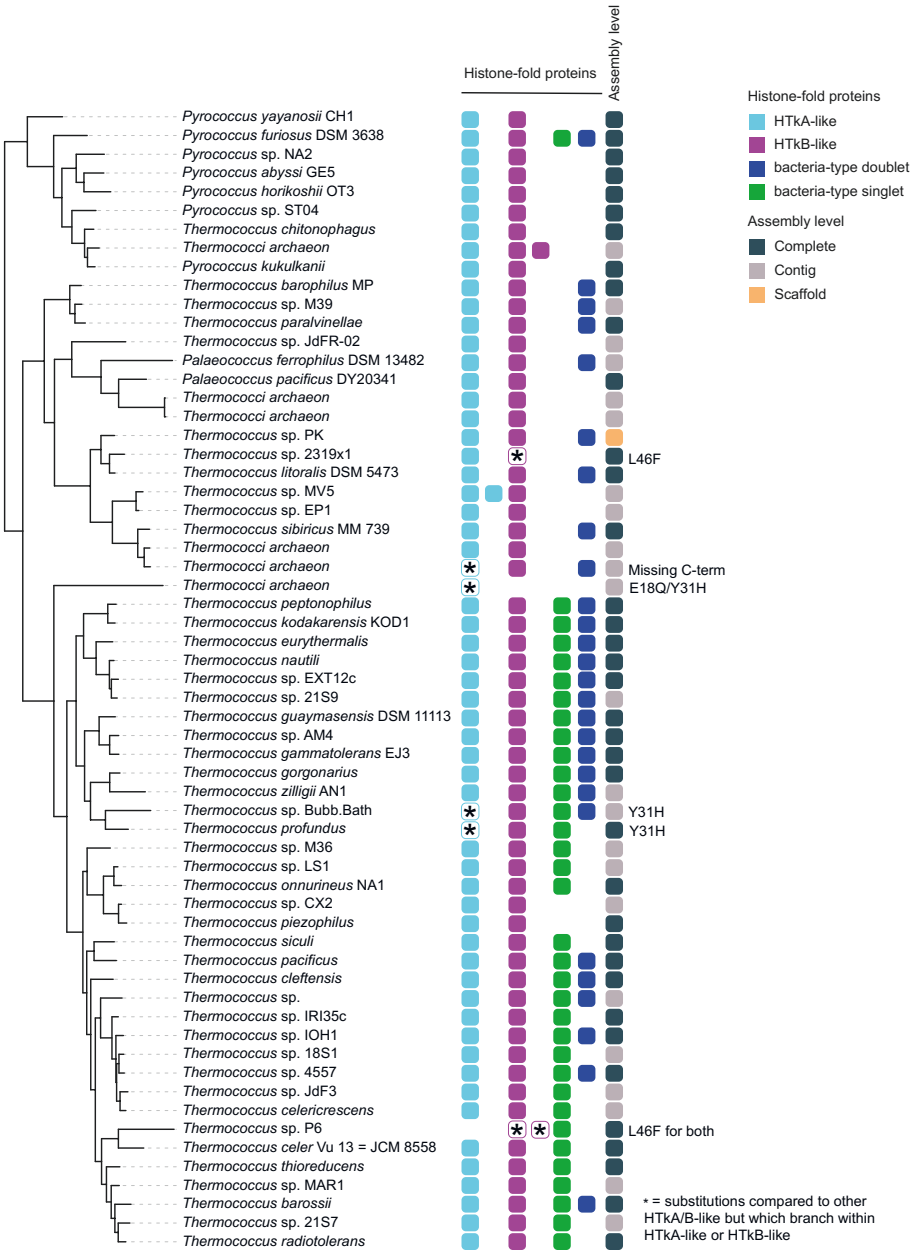


Figure 8.5 HTkA-like and HTkB-like histone paralogs are found in almost all Thermococcales species. GTDB species tree of Thermococcales archaea showing the presence of different histone fold proteins: HTkA-like, HTkB-like, bacteria-type singlet and bacteria-type doublet histones. The assembly level for each species is shown. For species with a HTkA/B-like histone without defined key residues (highlighted with an asterisk) the substitution is annotated on the right-hand side.

Some species have HTkA- or HTkB-like paralogs with amino acids which discriminate the other group (indicated with an asterisk in Figure 8.5). Intriguingly, in the one species which has a complete genome and only HTkB-like histones (*Thermococcus* sp. P6), the two HTkB-like both have a HTkA-like amino acid substitution (L46F, Figure 8.5). This raises the possibility that the two HTkB-like histones in *Thermococcus* sp. P6 may have gained the properties of the lost HTkA-like histone.

The global presence of HTkA- and HTkB-like histones suggests that these two groups may have been present in the last common ancestor of the Thermococcales. To test whether this observation was indicative of ancient paralogy, protein-level phylogenetic trees of the canonical HMf-like Thermococcales histones were constructed (Figure 8.6). Using this approach, the 123 HMf-like histones separate into two groups identified in the previous PCA analysis, the two HTkA-like and HTkB-like histones defined by their position on the tree.

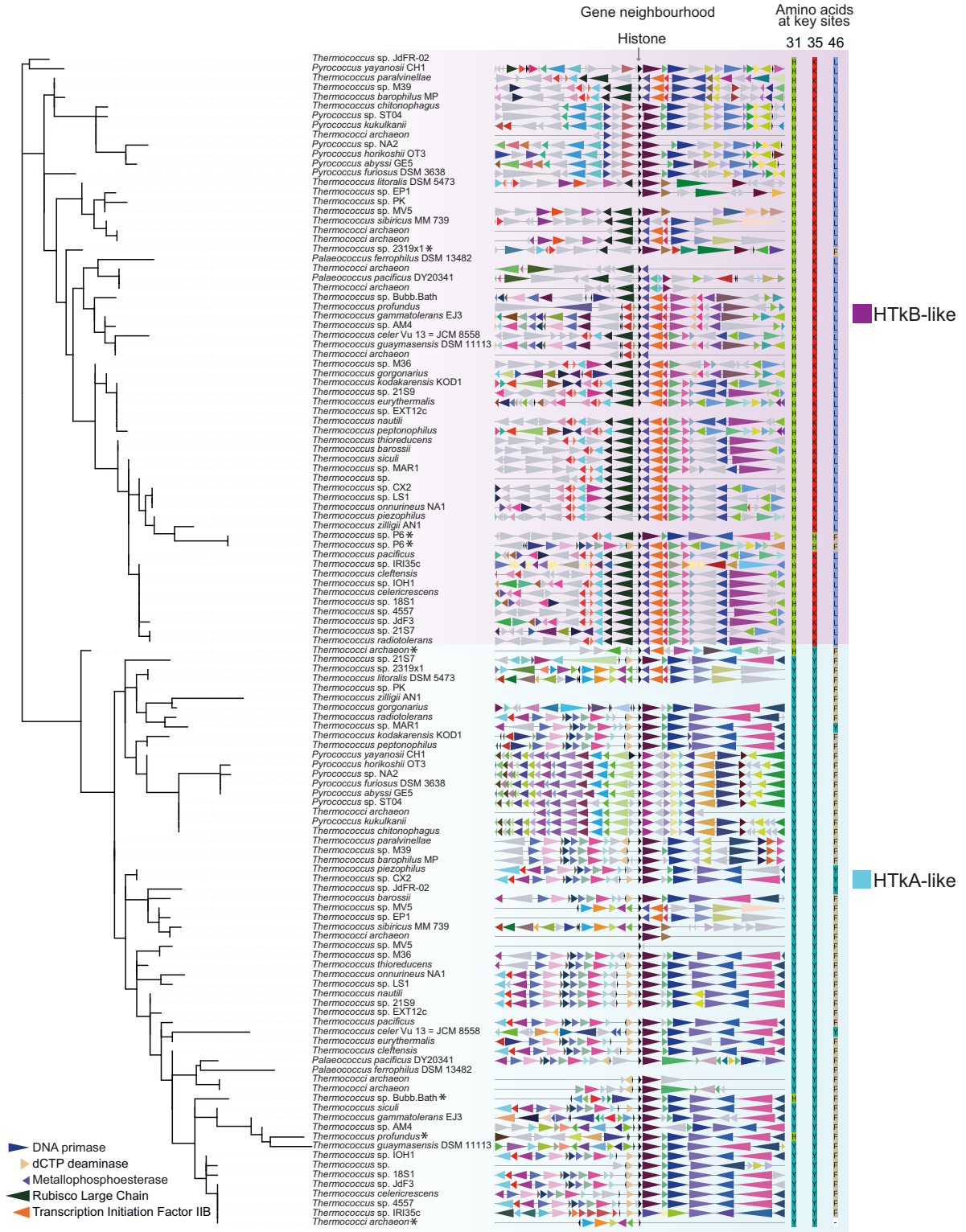


Figure 8.6 Protein-level tree of Thermococcales HTkA-like and HTkB-like histones.
Protein-level phylogenetic tree of HTkA/B-like histones from Thermococcales species. Genes in the genomic neighbourhood for each histone were plotted using GeneSpy (Garcia *et al.* 2019). Orthologs are coloured and identified as best reciprocal hits, with *T. kodakarensis* as a reference (see Methods The identity of amino acids at key residues is shown on the right with asterisks denoting those with deviations from the most common amino acid. Figure adapted from (Stevens *et al.* 2022).

In the Methanobacteriales, synteny was conserved for paralogous groups. In the Thermococcales HTkA-like and HTkB-like there are some paralogs within groups which display high syntenic conservation (Figure 8.6). Genes encoding proteins involved in transcription and DNA replication are found closely located to each paralog. For example, DNA primase is often found directly downstream of HTkA-like and Transcription Initiation Factor IIB is downstream of HTkB-like genes, although neither are predicted to form part of an operon (Figure 8.6).

The conserved sequence changes and *in silico* structural properties support the separation into two groups. However, compared to the Methanobacteriales, analysing the tree in conjunction with synteny is more complicated. There are some examples of HTkA-like histones in Pyrococcus spp. which appear to have proximal genes which are found neighbouring HTkB-like genes in Thermococcus spp. (and vice versa, Figure 8.6). The protein-level tree also does not directly follow the species tree.

The hypothesis that the common ancestor of the Thermococcales having HTkA- and B-like histones does not fall apart based on synteny alone. Within the Thermococcales there has been a global loss of synteny (Figure 8.7) which is particularly apparent for the HTkA/B-like histones.

By identifying orthologous genes throughout Thermococcales species and comparing the genes found upstream and downstream of each between genomes, it is possible to determine the amount of reshuffling which has occurred between genomes and calculate a gene order conservation score (GOC) (Rocha 2006). In the context of the HMf-like histones, this may hint at whether differences in the genomic context are more likely given high levels of gene rearrangements. Comparing gene order conservation between the two lineages (e.g., between Thermococcus spp. and Pyrococcus spp.), conservation is lower between than within the lineages (Figure 8.7a, b). This is indicative of even larger rearrangements *between* the two groups than *within* groups throughout evolution, as expected based on the time since each diverged.

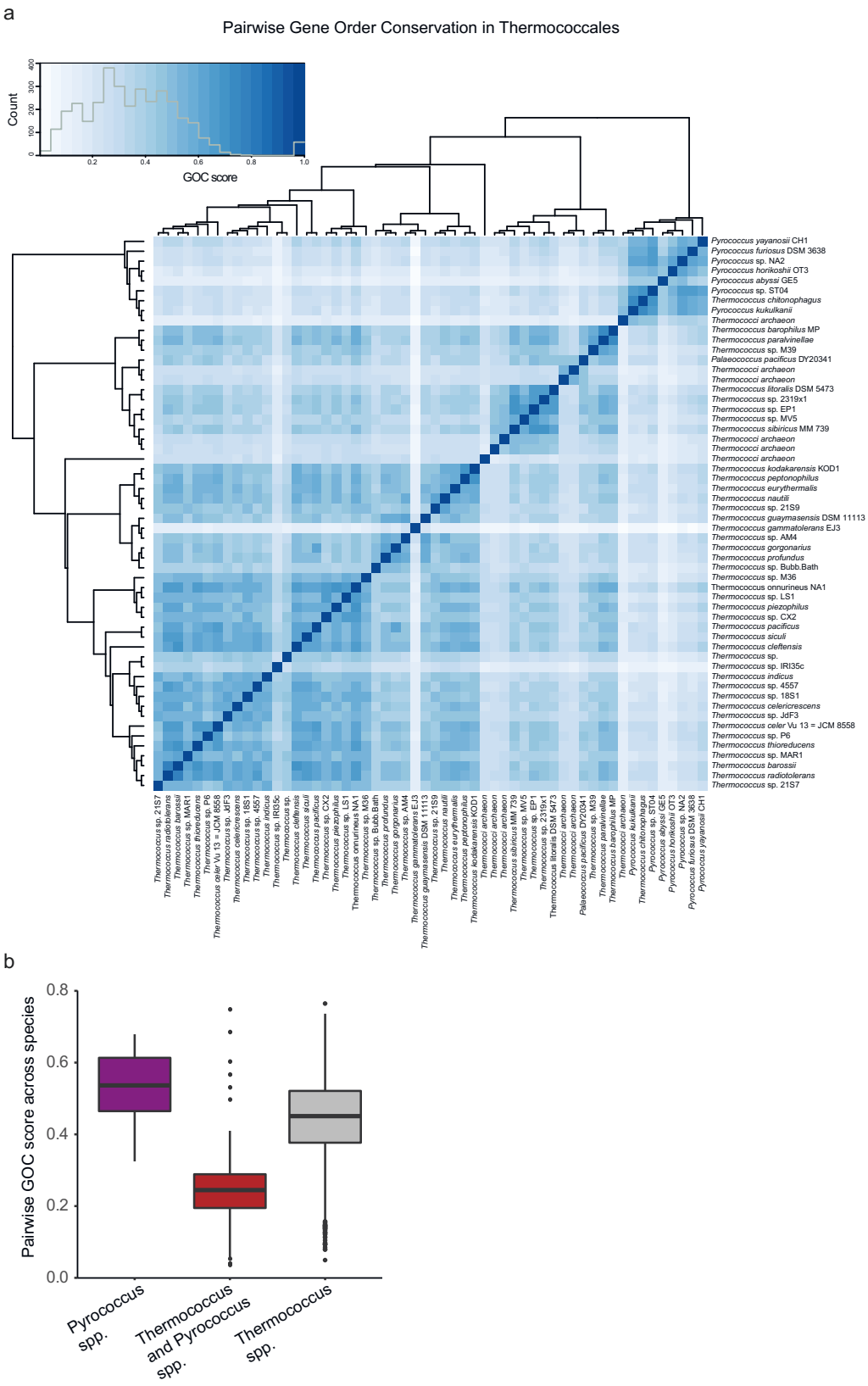


Figure 8.7 Gene order conservation in Thermococcales.
a) Pairwise gene order conservation (GOC) scores across Thermococcales species (see Methods) shown on the GTDB species tree. GOC score of 1 = complete conservation. b) Pairwise GOC scores for species belonging to Thermococcus, Pyrococcus, or pairwise between Thermococcus and Pyrococcus species.

This loss of global synteny can be seen particularly in the region around the histone, and help explain divergent genomic contexts within paralogous groups. Performing whole genome alignments with species which have HTkA-like or HTkB-like histones with divergent genomic neighbourhoods confirms the high level of rearrangements between species (Figure 8.8a, b). HTkA-like and HTkB-like histones align separately between species, with breakpoints either side (Figure 8.8 shows an example for *P. kukulkanii* and *T. barophilus*) suggesting that the differences in genomic context between Pyrococcus and Thermococcus species could reflect loss of synteny.

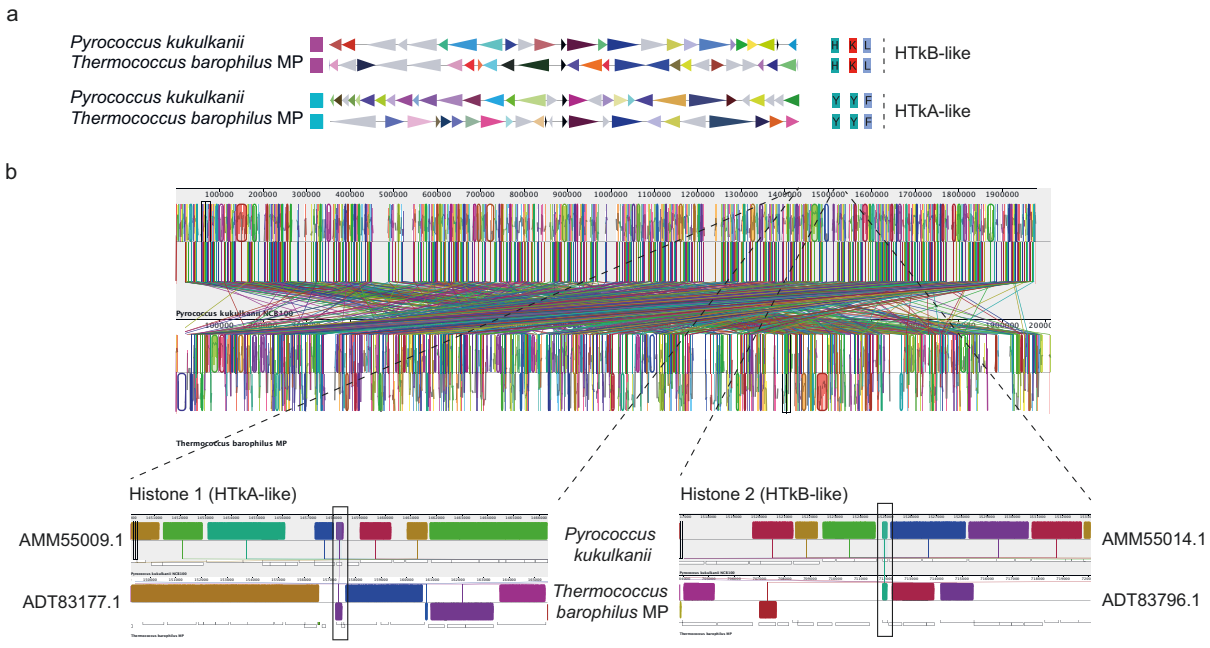


Figure 8.8 Whole genome alignment centred on HTkA/B-like histones.
a) An example of HTkA-like (bottom) and HTkB-like (top) histones with divergent genomic neighbourhoods from *P. kukulkanii* and *T. barophilus* MP. *P. kukulkanii* HTkB-like histone has a downstream neighbourhood consistent with HTkA-like histone in *T. barophilus* MP and other Thermococcus species. b) Whole genome alignment from progressiveMauve (Darling *et al.* 2010) centred on HTkA-like (left) or HTkB-like (right) histones from species in a).

8.4. HTkA/B-like histone paralogs do not extend beyond the Thermococcales

The presence of both HTkA- and B-like histones in almost all Thermococcales species suggests both were likely present in the last common ancestor. It is also possible that these two paralogs arose at an earlier point still, and may be found in outgroups to the Thermococcales. To

investigate whether this could be the case, histones in the three closest sister clades (NM3, Theionarchaeota, and Methanofastidiosa) were subject to the same analysis (Borrel *et al.* 2020).

All three clades have histones with a single histone fold domain. Unlike in the Thermococcales, there are no bacteria-type singlet or doublets in any species in these lineages (see section below). To determine whether these singlet HMf-like histones could be classified as HTkA/B-like, PCA clustering was performed as above, and amino acid identity analysed at key sites. Some histones in Methanofastidiosales spp. have short N/C-terminal extensions, and cluster outside the HTkA/B-like histones when clustered based on protein sequence properties (Figure 8.9a, c). Removing these sequences, reveals some Theionarchaeota histones which cluster with either HTkA- or HTkB-like histones (Figure 8.9b). All histones from NM3 species cluster with the HTkA-like histones.

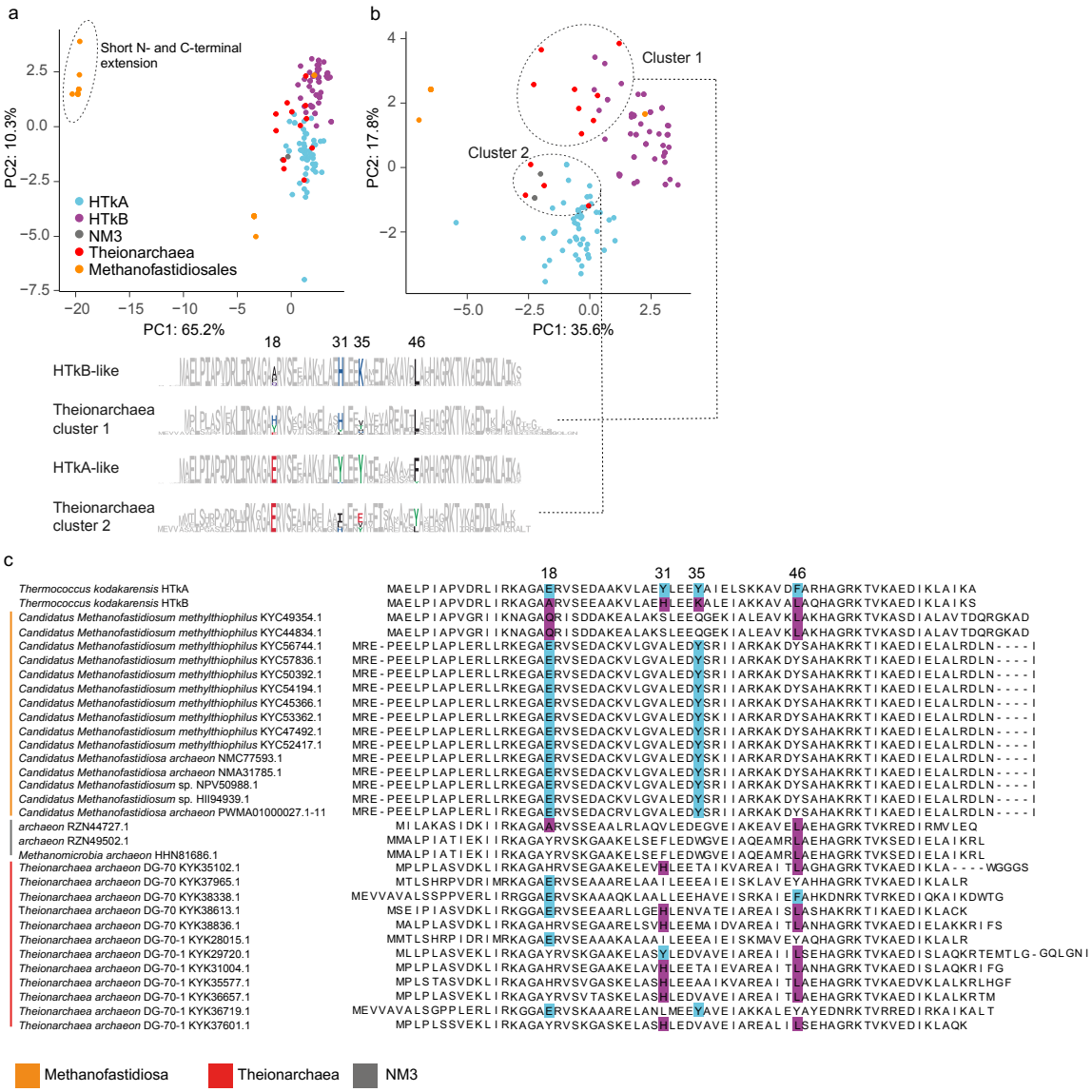


Figure 8.9 Clustering of histones from outgroup Methanofastidiosales, Theionarchaeota and NM3 species.
a) PCA analysis of outgroup species histones and HTkA-/HTkB-like histones using amino acid properties from Seqinr AAStats (Charif and Lobry 2007). b) Left: PCA analysis of outgroup spp. histones and HTkA-/B-like histones with longer histones with N/C-terminal extensions in Methanobacteriales removed. Two groups of Theionarchaeota histones cluster with either HTkA-like or HTkB-like histones with positions which discriminate the Thermococcales histone groups highlighted. c) Alignment of *Thermococcus kodakarensis* histones with histones in outgroup species in b). Sites which determine HTkA/B group identity are highlighted in blue and purple, respectively.

Looking at the sequence of the Theionarchaeota histones which separated into HTkA/B-like groups, there are different amino acids at residues which distinguish between HTkA/B-like histones (Figure 8.9b, c). Most Theionarchaeota histones have either HTkA- or HTkB-like amino acids at sites which are deterministic of either paralog, and only a few have mixed HTkA/B residues at these sites (Figure 8.9c). However, the overall identity between these

histones and those in Thermococcales is low. These histones also do not clearly separate with HTkA/B-like histones on a protein-level tree (not shown). Taken together, these results do not support HTkA/B-like histones extending beyond outside the Thermococcales and it is conservative to assume that these variants are only as old as the origin of the Thermococcales.

8.5. Some Thermococcales species encode for histones similar to those in some bacteria

As well as the canonical histones, some Thermococcales species also encode bacteria-type singlet and doublet histones (Figure 8.1). These consist of a singlet histone fold domain (bacteria-type singlets) and are relatively shorter on average in the Thermococcales (59 aa compared to ~66 for HTkA/B), or are organised as an end-to-end duplication of the HFD (in the case of bacteria-type doublets). A comprehensive overview of the histone superfamily in archaea, bacteria and eukaryotes identified these two groups, which contain sequences from bacteria and archaea, as separate clusters which are distinct in sequence from HMf-like histones (Alva and Lupas 2019).

These histones are restricted to the Thermococcales and are not present in the closest outgroup archaeal species (Figure 8.10c). The bacteria-type histones are also, on average, less well-conserved than the canonical HMf histones when considering pairwise sequence identity across species (Figure 8.10a).

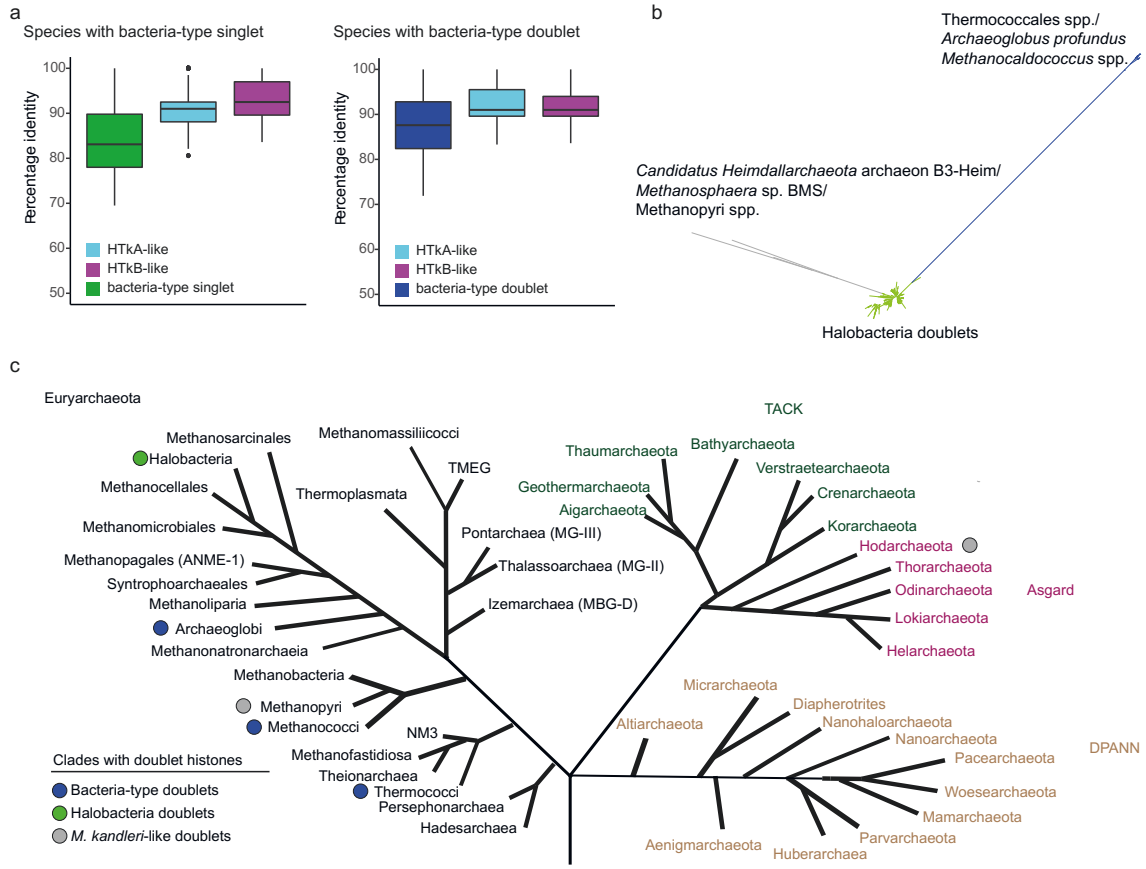


Figure 8.10 Bacteria-type histones in Thermococcales species.
a) Pairwise percentage sequence identity for HTkA/B-like histones compared to bacteria-type singlets (left) and bacteria-type doublets (right). Calculations were done for all species with a bacteria-type singlet, HTkA-like and HTkB-like histone (left), and a bacteria-type doublet, HTkA-like and HTkB-like histone (right). b) Protein-level phylogenetic tree of doublet histones from across archaeal species. Histone doublets are coloured by their identity/origin. c) Archaeal phylogeny (adapted from (Borrel *et al.* 2020, Stevens *et al.* 2020)) showing the distribution of bacteria-type doublets and other archaeal doublet histones. Histone doublets are coloured by their classification (as shown in Figure 43e). Figure adapted from (Stevens *et al.* 2022).

As discussed in the section 2.8.1.1, there are other known doublet histones, for example HMk in *M. kandleri* (Slesarev *et al.* 1998, Fahrner *et al.* 2001) and those in the Halobacteria (Figure 8.10c). The doublet histones in Thermococcales differ in sequence from HMk and those in Halobacteria. On a protein-level tree built from doublet histones throughout archaea, they branch as a distinct, separate monophyletic group with other doublet histones from *Methanocaldococcus* and *Archaeoglobus* (Figure 8.10b). Species encoding the bacteria-type doublet are all hyperthermophilic or thermophilic. *Archaeoglobus profundus* and Thermococcales species all grow in high temperatures and, in *Methanocaldococcus*, I only find these doublets present in species which are thermophilic or hyperthermophilic. This might indicate a specific role for the doublet within the high temperature environment (see below).

Unlike HMf-like histones, the distribution of the bacteria-type doublet and singlets is relatively patchy and restricted to a few species (Figure 8.5). The bacteria-type singlet histone is found in a single lineage of Thermococcus species, in the branch containing *T. kodakarensis*, and in *Pyrococcus furiosus* (Figure 8.5). Compared to HTkA/B paralogs, the syntenic context is highly conserved in all Thermococcus species containing the bacteria-type singlet (Figure 8.11a). The conserved genomic neighbourhood and presence in almost all species in the Thermococcus lineage is consistent with the bacteria-type singlets having a single origin. Centring the syntenic plot on a neighbouring gene close to and conserved in almost all the Thermococcales bacteria-type singlets (TK1039), there appears to be a single loss event in the Thermococcus branch leading to *T. piezophilus* and *Thermococcus* sp. CX2 (Figure 8.12a). The bacteria-type singlet in *P. furiosus* has a different genomic context than those in Thermococcus species. This could be again due to high levels of genome reshuffling between the two lineages (Figure 8.7, see above). Based on the syntenic context, presence in Thermococcus spp. and absence from other Pyrococcus spp., the histone in *P. furiosus* was likely acquired from an ancient or extant *Thermococcus* species. However, this case of HGT appears to be rare, and evidence does not clearly support this as a driver of histone diversity more globally.

The bacteria-type doublets also have a conserved genomic neighbourhood and are found distributed throughout the Thermococcales (Figure 8.6, Figure 8.11b). This distribution and conservation of syntenic neighbourhood is consistent again with a single origin for the bacteria-type doublets and subsequent loss in several species. Centring the genomic context again on a better conserved gene close to the bacteria-type doublet histones (TK0752), loss of the doublet from the conserved region can be clearly seen (Figure 8.12b).

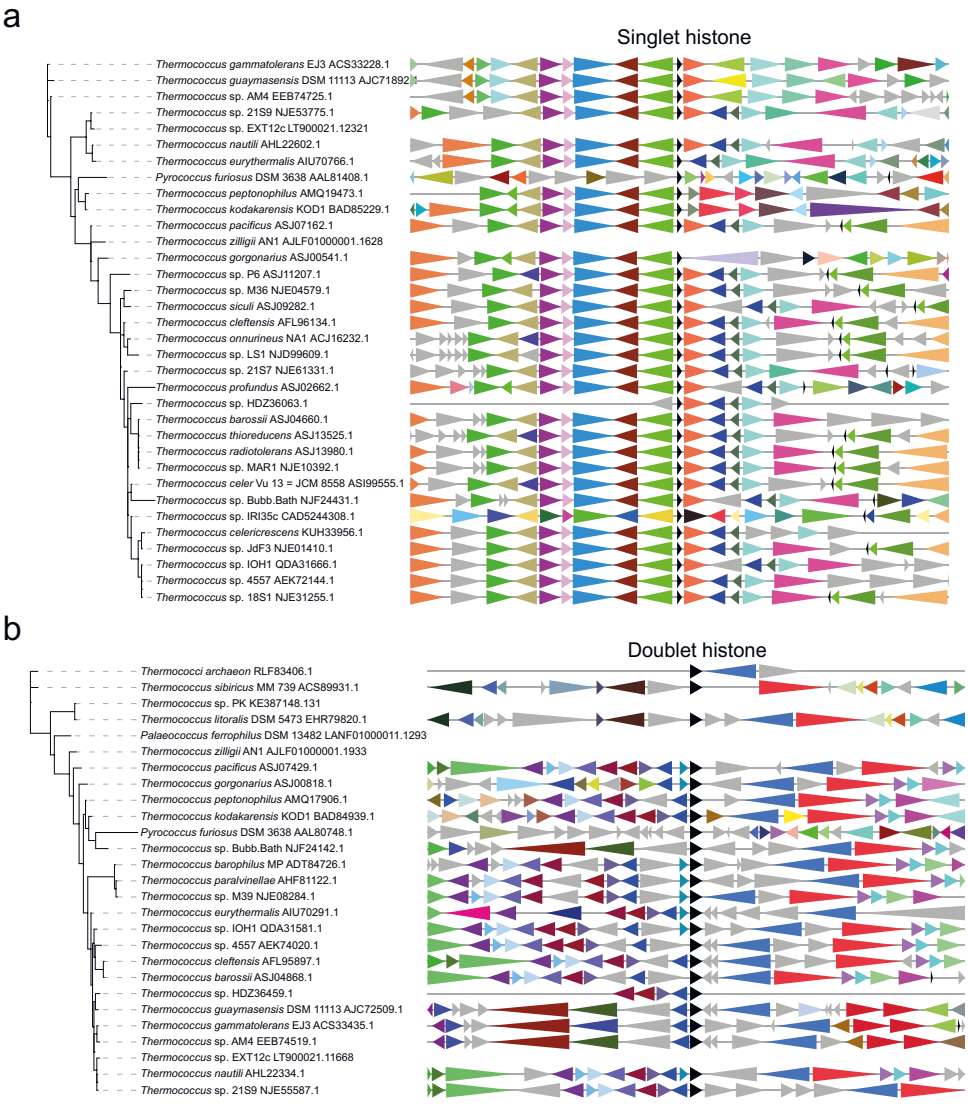


Figure 8.11 Protein-level phylogenetic trees of bacteria-type histones showing genomic neighbourhood. Protein-level trees of a) bacteria-type singlet, and b) bacteria-type doublet histones, showing genes in the syntenic region coloured by ortholog identity (see Methods) and centred on the histone. Figure adapted from (Stevens *et al.* 2022).

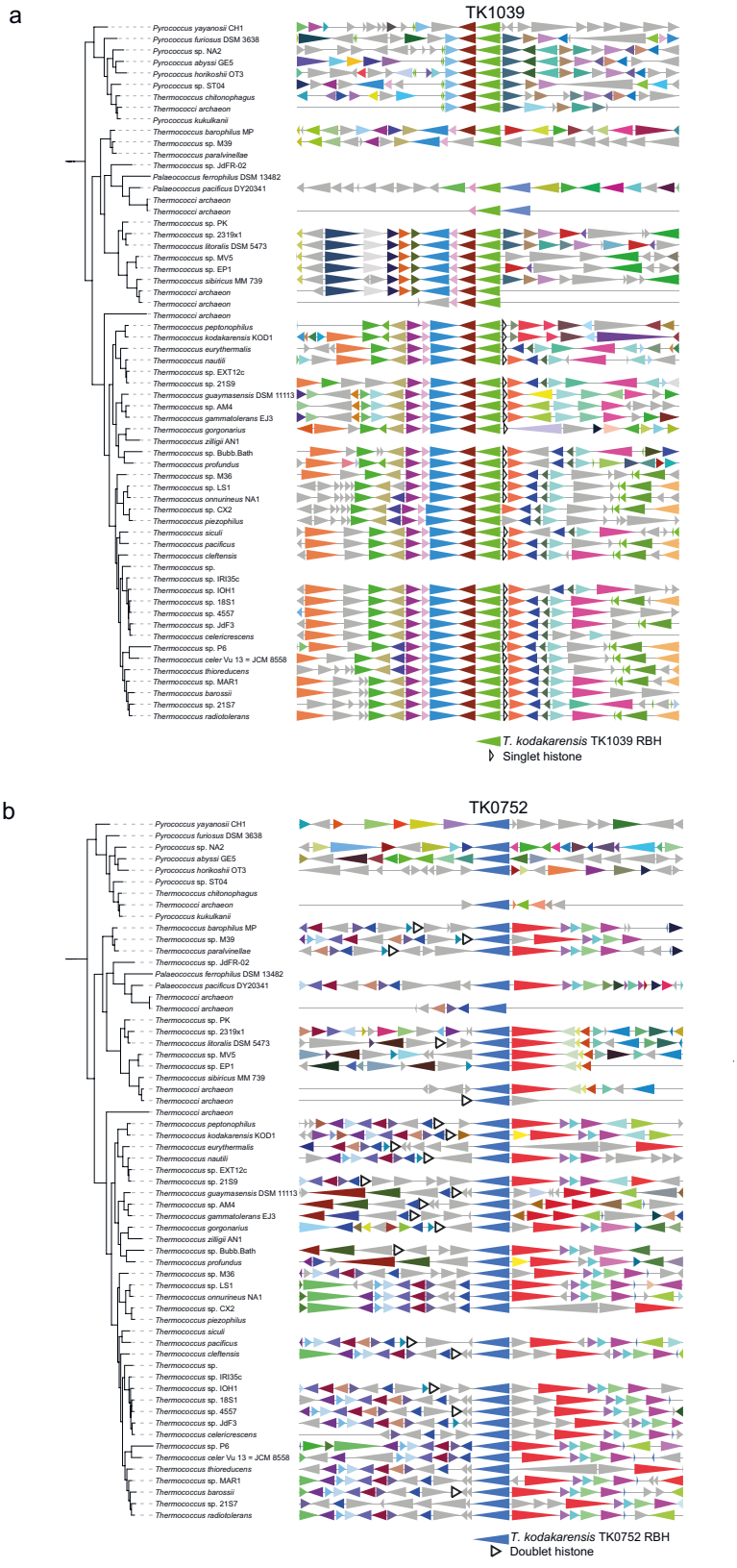


Figure 8.12 Species-level phylogenetic trees of bacteria-type histones showing genomic neighbourhood centred on conserved syntenic region.
Species level phylogenetic trees of Thermococcales species showing genes in the syntenic region as in Figure 8.11, but centred on conserved close neighbouring genes of c) bacteria-type singlets and, d) bacteria-type doublets. Histone proteins, where present, are highlighted in each. Figure adapted from (Stevens *et al.* 2022).

We do not currently know what know the function of the bacteria-type histones could be. From previous data, we know in *T. kodakarensis* they have a relatively lower expression level compared to HTkA and HTkB, so are less likely to be involved in global genome compaction to the same extent (Figure 8.13a). This is true both at the transcript level, as well as the protein level (Jäger *et al.* 2014, Sas-Chen *et al.* 2020). In exponential phase, both bacteria-type singlet and doublet histones together contribute just 0.37% of the proteome, compared to 1.1% for HTkA and 0.66% for HTkB (Figure 8.13a).

Previous studies have linked the bacteria-type doublets to heat shock response. In *T. kodakarensis* and *P. furiosus*, both doublets are controlled by the heat shock regulator Phr and are upregulated when Phr is deleted (Kanai *et al.* 2010, Keese *et al.* 2010). At lower temperatures in *T. kodakarensis*, the bacteria-type doublet is downregulated and shows decreased expression level (Hocher *et al.* 2021). This suggests that the bacteria-type doublet may have a specific role in adapting to temperature change, consistent with the presence in thermophilic and hyperthermophilic species of archaea (Figure 8.10c).

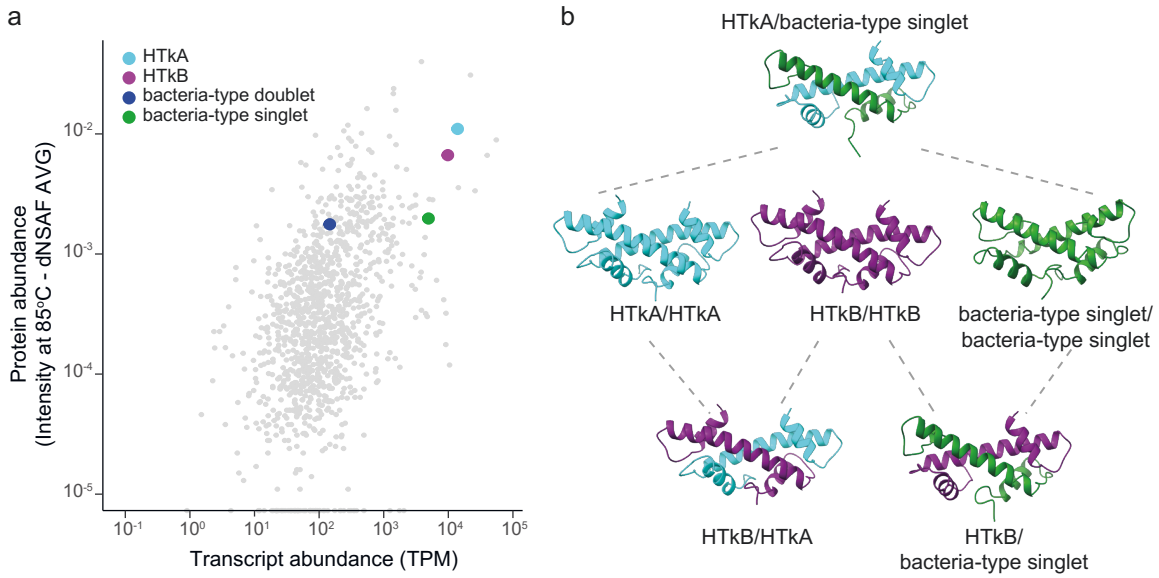


Figure 8.13 Expression level of histones in *T. kodakarensis* and alpha fold prediction of HTkA/B with bacteria-type singlets.
a) Expression level of histones in T. kodakarensis at the protein level and transcript level (Jäger *et al.* 2014, Sas-Chen *et al.* 2020). Histone proteins are coloured. b) AlphaFold predicted homodimeric and heterodimeric structures of histones in T. kodakarensis (HTkA-like, HTkB-like and bacteria-type singlets/doublets) showing all possible combinations. Monomers are coloured by histone identity. Figure adapted from (Stevens *et al.* 2022).

The bacteria-type singlet in *T. kodakarensis* (TK1040) was previously found in low amounts in the chromatin (Maruyama *et al.* 2011), suggesting that it may be able to associate with, and bind to, DNA, either through direct interaction or indirectly. Although unlikely to be as key as HTkA/B-like histones, the presence of TK1040 in the chromatin suggests that it might be interacting with HTkA and HTkB in *T. kodakarensis*, furthering the potential diversity of histone-based chromatin in this species. To predict whether TK1040 can form stable dimers with the HMf-like histones, AlphaFold (Jumper *et al.* 2021) was used to model all homodimers (e.g. HTkA+HTkA, or bacteria-type singlet with bacteria-type singlet), as well as all combinations of histone proteins to form heterodimers (e.g. bacteria-type singlet with HTkB). Using this approach, as expected, HTkA and HTkB are predicted to form both stable homodimers and heterodimers (Figure 8.13b). Interestingly, the bacteria-type singlets are also predicted to form stable homodimers, and are predicted to heterodimerise with HTkA and HTkB (Figure 8.13b).

8.6. HTkA- and HTkB-like histones have a separate origin to A/B histones in the Methanobacteriales

Methanothermus fervidus also encodes histones designated as A and B which have shared synteny with paralogous groups in Methanobacteriales (see Chapter 7). The labels "A" and "B" do not imply ancient paralogy with Thermococcales histones; they were named in ascending alphabetical order independently in different species as a simple matter of convention. However, this does not rule out the possibility that ancient paralogy does, in fact, exist. To determine whether the HMfA/B histones are orthologous to HTkA/B-like histones, HMfA and B from *M. fervidus* were compared to the HTkA/B-like histones in Thermococcales. Clustering the Thermococcales HMf-like histones with HMfA and HMfB, I find that both *M. fervidus* histones cluster more closely with HTkA-like paralogous histones, and do not separate into the two clusters (Figure 8.14a). Similarly, building protein-level trees of these histones, both HMfA and HMfB branch together within the HTkA-like histones, and not into the two paralogous histone groups in Thermococcales (Figure 8.14b). Taken together, this suggests that the AB duplication which lead to HMfA and HMfB was a separate event to the AB duplication in Thermococcales HMf-like histones, and that it is unlikely that they have a common origin.

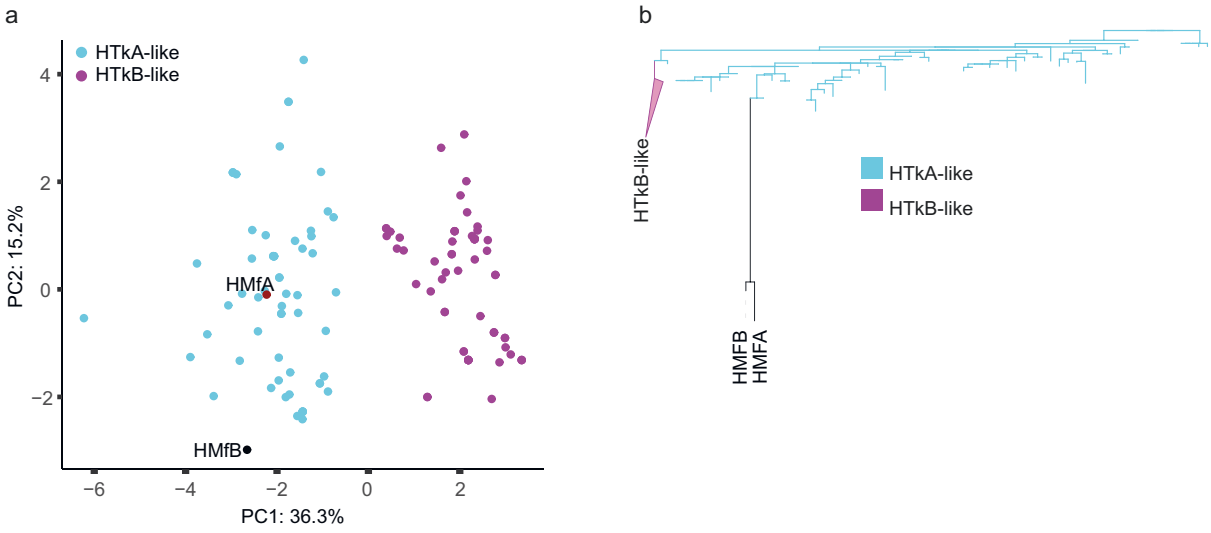


Figure 8.14 A/B type histones in Methanobacteriales do not have the same origin as HTkA/B histones in Thermococcales spp.
a) PCA clustering of HTkA-/B- like histones from Thermococcales spp. and HMfA/B based on amino acid properties from Seqinr AAStats (Charif and Lobry 2007). b) Phylogenetic tree of HTkA/B-like histones with *M. fervidus* histones HMfA (red) and HMfB (black). The branch containing HTkB-like histones has been collapsed.

8.7. Summary

Histones in the Thermococcales can be separated into HTkA-like and HTkB-like paralogous groups based on sequence properties and structural differences. Compared to the Methanobacteriales, synteny across the Thermococcales is less well conserved and points either side of the histones have served as breakpoints in genome shuffling events. Although lack of shared synteny is convoluted in some cases, there is no strong support for a separate origin within either paralog. It is therefore parsimonious that both an HTkA-like and an HTkB-like histone were present in the last common ancestor of the Thermococcales and each variant has been maintained in almost all extant species. The Thermococcales arose as a clade roughly 750 million years ago, putting the presence of HTkA/B-like histones in a similar, although shorter, time frame to paralogs in the Methanobacteriales (Wolfe and Fournier 2018, Stevens *et al.* 2022). The conservation of these paralogous histones supports the hypothesis that sequence differences between the histone paralogs are functionally important and under selection (Stevens *et al.* 2022). This adds further support to the hypothesis that histone variants, as conserved paralogs, exist in archaea, evolving and diversifying in parallel to histone variants in eukaryotes.

The presence of other histone fold proteins consisting of either a singlet (bacteria-type singlet) or doublet (bacteria-type doublet) histone fold domain throughout Thermococcales adds to the potential diversity of histone-based chromatin in these species (Stevens *et al.* 2022). However, further research is needed to characterise these proteins. Do they interact with HTkA/B-like histones *in vitro* and *in vivo*? And do they play a role in adaptation in these species – i.e., can they alter DNA accessibility in response to changes in environmental conditions or at different stages of cell growth?

9. Histone diversification through specific amino acid changes in archaea and eukaryotes

9.1. Introduction

Chapters 7 and 8 were motivated by trying to understand whether related multi-histone archaea have shared paralogs. This information provides glimpses at likely evolutionary histories for different paralogs, but does not tell us how, on a more global level, histones have diversified in sequence. From the Msp_0383 capstone, we know that single substitutions can have potentially important ramifications for the function of histones, and ultimately chromatin structure. This is similar to functional histone variants in eukaryotes, where sequence differences can alter the dynamics of the nucleosome (see section 2.4). To understand the sequence diversification of archaeal histone paralogs in archaea across species more generally, we combined structural approaches with a phylogeny-free approach.

The consequence of substitutions other than at the capstone H49 site have been documented through prior *in silico*, *in vitro* and *in vivo* work. This has mostly focused on *M. fervidus* histone HMfB and detailed residues which, when mutated, can alter properties of the histone-DNA complex, such as DNA binding, stability, and tetramerization (Figure 9.1). The effects of different mutations on DNA binding were extensively probed by Soares et al. (2000), who generated a large set of mutations across the HMfB sequence, and analysed the DNA binding of the mutants through gel shift assays. Other residues have known functional importance in intra- and inter-monomer interactions. For example, Marc *et al*. (2002) and Higashibata *et al*. (1999) identified residues key for tetramer stability and, in turn, DNA compaction in HMfB and *T. kodakarensis* histone HTkB through generation of mutants. Other residues have been shown to play a key role in oligomer formation. Mutating position 17 (16 relative to HMfB sequence) from a glycine to larger amino acids in HTkB prevents the formation of stable oligomeric complexes (Mattiroli *et al.* 2017). Using sequence-based approaches, Henneman *et al*. (2018) identified further positions at the stacking interface between non-adjacent dimers. Other substitutions have been identified which prevent the accumulation of histones in *E. coli*, a proxy for compromised folding or stability (Soares *et al.* 2000, Higashibata *et al.* 2003).

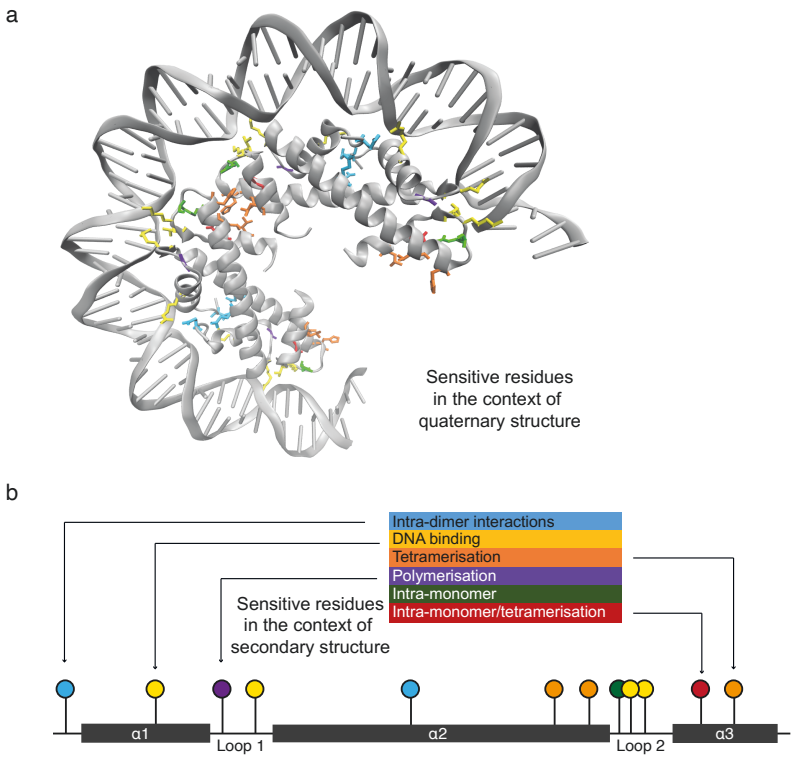


Figure 9.1 Sites sensitive to structural change in archaeal histones.
Residues which are known to affect structural interactions upon mutation (based on previous work, (Higashibata *et al.* 1999, Soares *et al.* 2000, Marc *et al.* 2002, Mattiroli *et al.* 2017)) are shown on a) the quaternary structure of HMfB (PDB 5t5k), and b) the secondary structure of the histone fold. This figure was adapted from (Stevens *et al.* 2020).

In this chapter, the global sequence diversity of archaeal histones in the context of residues known to be structurally important is explored, and compared to the sequence diversity of eukaryotic histones H3 and H4. These core histones, and not H2A/H2B were chosen for two main reasons. Firstly, archaeal histones are more similar to H4 (and H3 but less so) (Bailey *et al.* 1999, Malik and Henikoff 2003, Henneman *et al.* 2018, Grau-Bové *et al.* 2022). Secondly, the $(H3\text{-}H4)_2$ tetramer has specific residues at the tetramer interface which direct assembly and are therefore constrained in a manner similar to archaeal histones (Figure 2.4d) (Luger *et al.* 1997, Postberg *et al.* 2010).

9.2. *In silico* mutational analysis of the HMfB hexamer complex

To identify sites in *M. fervidus* HMfB which are structurally sensitive to change, the HMfB crystal structure (pdb: 5T5K, (Mattiroli *et al.* 2017)) was subjected to *in silico* mutational analysis. FoldX (Schymkowitz *et al.* 2005) was used to relax and then mutate each residue and

the changes in Gibbs free energy of DNA binding, interaction between protein subunits, as well as stability was determined. Each single mutation was performed for all six HMfB monomers simultaneously.

These results were compared to sites known from prior *in vitro* experiments (largely from HMfA/B) to be sensitive to DNA binding, tetramerization, stability and oligomerisation (Soares *et al.* 2000, 2003, Marc *et al.* 2002, Higashibata *et al.* 2003, Mattiroli *et al.* 2017). The effects from the modelling approach are broadly concordant with those observed *in vitro*. In general, there are changes to predicted DNA binding when residues in and around the loop regions are mutated, in agreement with these regions being proximal to the DNA and containing residues known to be key for DNA binding (Figure 9.2a, (Soares *et al.* 2000)). Residue 19, which plays an important role in DNA binding (Soares *et al.* 2000), similarly shows large deviation from the wild-type structure in DNA binding (Figure 9.2a). The sites essential for tetramer stability *in vitro* (46/49/59/62) (Marc *et al.* 2002) also show large changes upon mutation ($\Delta\Delta G$) at the interface between adjacent dimers compared to the wild type relaxed HMfB structure (Figure 9.2b). Adjacent dimer pairs are AB-CD, and CD-EF (Figure 9.2b).

The change in energy at the interface between stacked dimers (AB-EF) is smaller than that between adjacent dimers (Figure 9.2b). The distance between the terminal stacked dimers in the crystal structure is too great to be probed using FoldX.

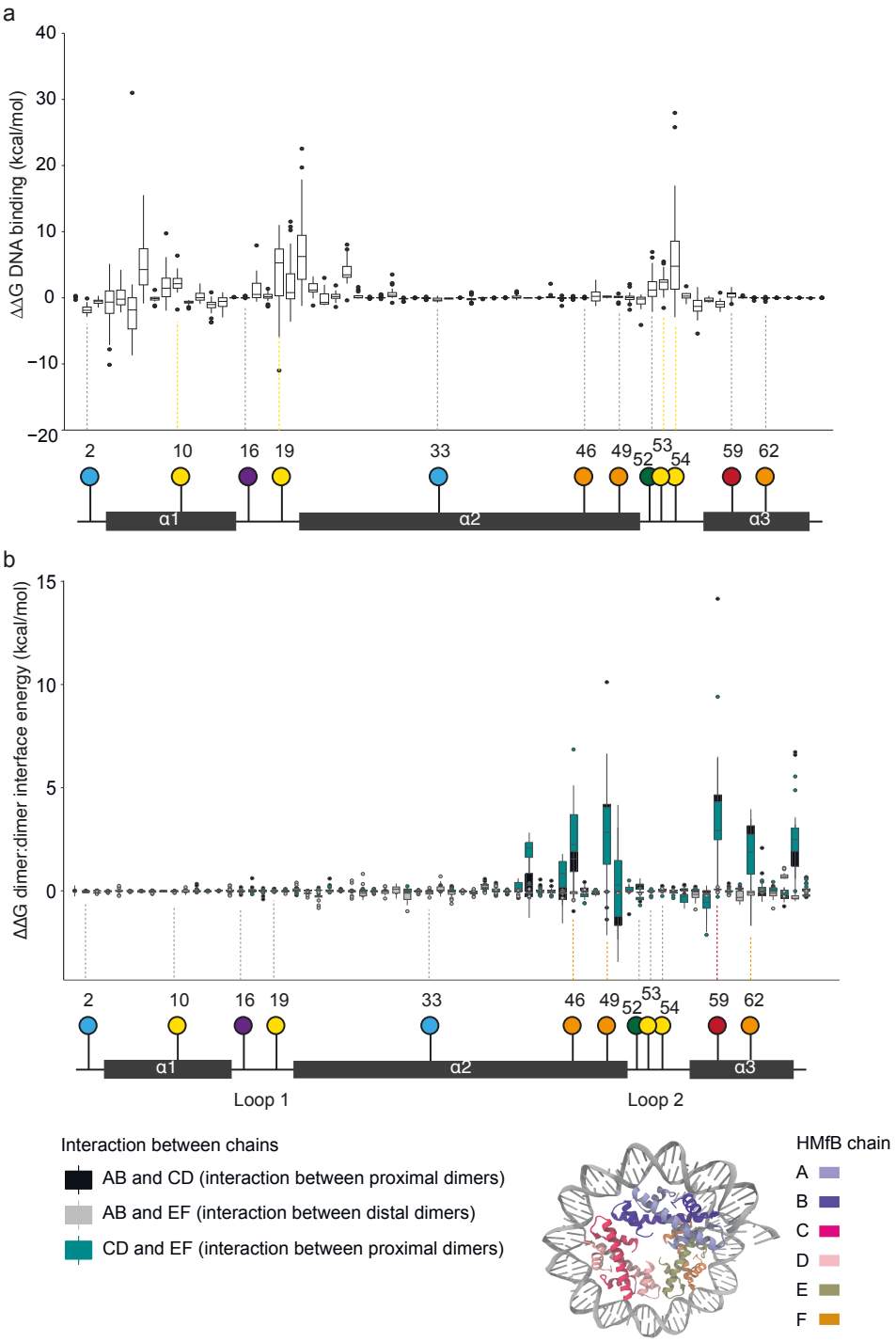


Figure 9.2 Interrogation of the HMfB structure using in silico single mutations.
Changes to a) DNA binding and b) Interface energy between dimers (a proxy for tetramerization strength), for all possible single mutants of HMfB. ΔΔG was calculated using FoldX relative to the repaired wild type HMfB hexamer. For tetramer strength, this is calculated at the interface between pairs for all three dimers. Adjacent and stacked dimers are shown on the HMfB structure. The position of amino acids with known structural importance for experimental work are shown on the secondary structure (see Figure 9.1). This figure was adapted from (Stevens *et al.* 2020).

These results were also compared to the gel shift assay results for the HMfB mutants produced *in vitro* by Soares *et al*. (2000). When the mutated HMfB histone is incubated with DNA and run on a gel, the migration is compared to the wild type. Complexes with faster migration compared to wild-type HMfB indicate stronger DNA binding and compaction, whereas those with slower migration indicate weaker DNA binding and overall compaction levels. More often than not, our *in silico* mutation results predict which mutations increased or decreased DNA binding *in vitro* (Figure 9.3, Appendix table 14.2). Single mutants with no gel shift or reduced gel shift from *in vitro* data are generally predicted to have weak DNA binding, whereas mutants with increased gel shift are predicted to have stronger DNA binding using our modelling approach.

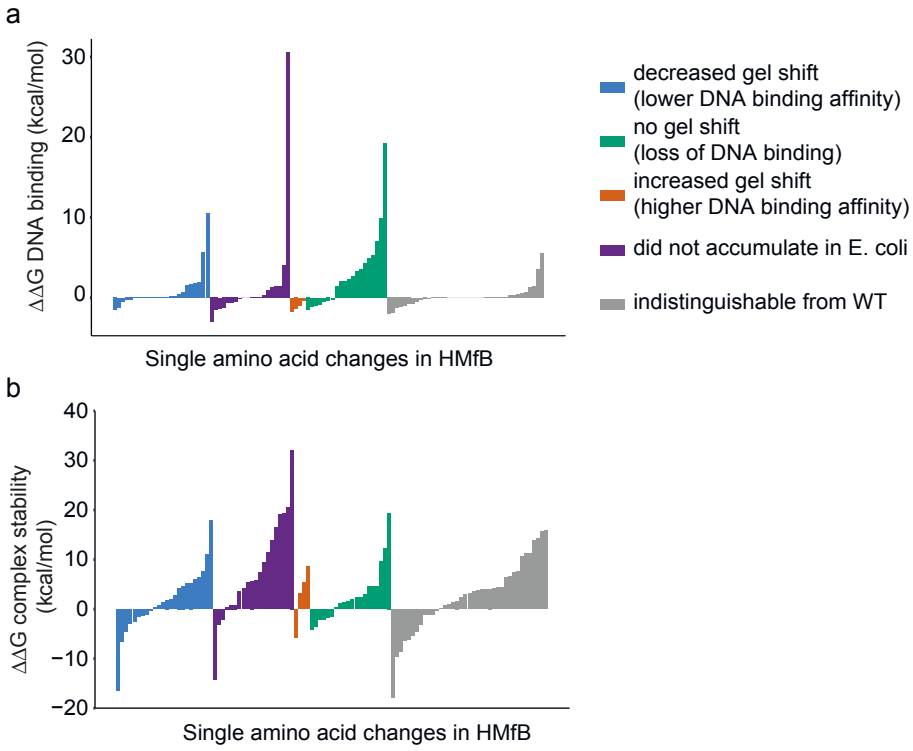


Figure 9.3 In vitro gel shift assay for HMfB single mutants compared to in silico results.
FoldX changes to a) DNA binding and b) stability for single mutants of HMfB which have been characterised in vitro by transformation into E. coli and in vitro gel shift assays (Soares *et al.* 2000). See appendix table 14.2 for results for individual mutations. This figure was adapted from (Stevens *et al.* 2020).

The agreement between the *in silico* mutations and *in vitro* data provides additional validation for the modelling approach taken in Chapter 5. Aside from validation, the modelling also provides additional insights into other residues which may be integral for DNA binding or

structural stability, but have not yet been investigated experimentally. For example, residue 50 in HMfB shows large deviation in the strength of tetramerization upon mutation (Figure 9.2b).

9.3. Substitutions drive the diversification of archaeal histones

How does the plethora of potential structural effects compare with what has occurred during evolution? Are structurally sensitive sites well conserved, or are there many changes at key positions, such as that seen in the *M. stadtmanae* capstone? A non-phylogenetic approach was applied to the alignment of archaeal monomeric histones (506 sequences) used to generate tetrameric histone complex models, alongside our knowledge of structurally sensitive sites to answer these questions.

To see how histones have diversified in archaea, sequences were split into two groups: those from species where the genome only encodes a single histone gene, and those where the genome encodes multiple (two or more) histone genes and where, as a consequence, positions along the protein can in principle diversify to assume different properties and encode different functionality. Such diversification could be driven by positive selection or relaxed constraint after duplication in systems with multiple histone paralogs.

The Shannon entropy formula (*H*) is a measure of diversity based on both richness and evenness (Shannon 1948, Shannon and Weaver 1964). Shannon's entropy was used as this is a sensitive measure, and takes into account both the types of amino acid present, and the number of each amino acid at each site. This was used to calculate diversity at each position in the protein alignment, keeping sites where we have no gaps in >95% of species to focus on the core conserved histone fold domain (see Methods). Shannon entropy is often used to measure species diversity (Spellerberg and Fedor 2003), but can also be applied to protein sequences to determine the probability of the presence of a specific amino acid at a specific site (Khan *et al.* 2008). Phylogenetic sampling, the number of histones in each of the two groups and other factors will affect the average ratio of diversity. However, information from sites which deviate from the average can tell us about sites which have diversified more or less often in the two groups.

Taking this approach, I find that diversification of archaea with multiple histone paralogs is strongly associated with sites which are structurally important (Figure 9.1, Figure 9.4). Positions with known sensitivity upon mutation, such as those involved in DNA binding or tetramerization, are relatively more diverse than other positions in the protein sequences for archaea with multiple copies of histones (H_M), compared to archaea with only one histone (H_S) (Figure 9.4). The loop regions, where the histone contacts the DNA, are particularly divergent compared to the rest of the histone fold. Perhaps the most obvious example is residue 19, which is completely conserved as an arginine in archaea with one histone ($H_S = 0$) but diversifies to eight different amino acids in systems with multiple histone paralogs ($H_M/H_S = \infty$, Figure 9.4). Other DNA binding residues (10, 53, 54) are also highly diverse.

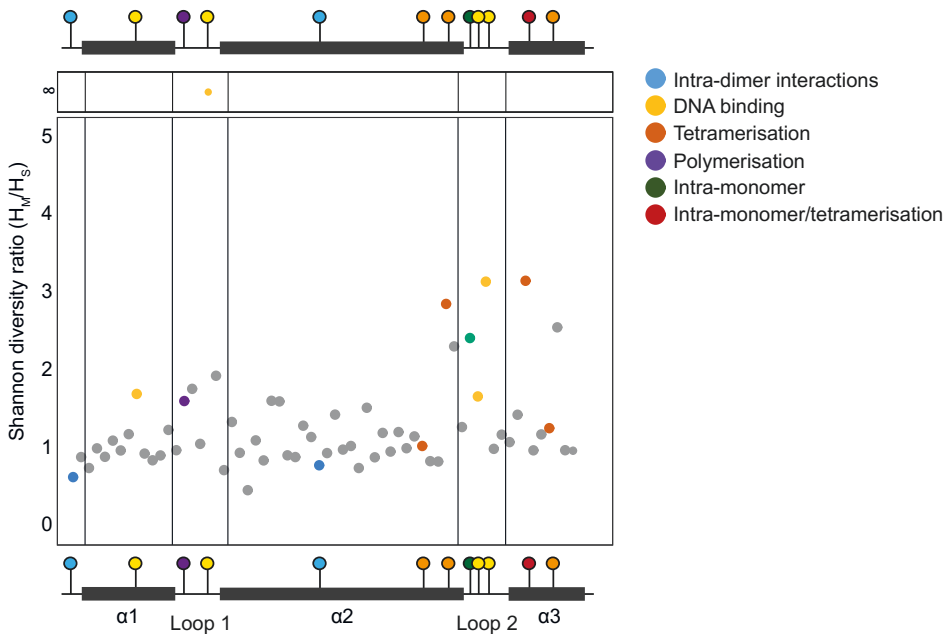


Figure 9.4 Sequence diversification of archaeal histones.
Shannon diversity ratio (H_M/H_S, archaea with multiple histones v those with only one) for each position in the core histone fold. The location of residues with structural properties identified by in vitro work is shown above and below. This figure was adapted from (Stevens *et al.* 2020).

As well as diversity in residues involved in DNA binding, I also find significant diversification at sites key for tetramerization. This includes residue 49 (position relative to the HMfB sequence) which defines the capstone (Figure 9.4).

The sequence diversification of archaeal histones shows some phylogenetic structure. By plotting the distribution of substitutions away from the most commonly found amino acid

across species, we can see phyletic patterns of particular sequence diversification (See Appendix Figure 14.2 for substitutions at structurally sensitive sites per species). In some lineages, structurally sensitive sites (Figure 9.1, 9.2) are all highly conserved. This is particularly noticeable in the Thermococcales (Figure 14.2). There are also substitutions at other structurally sensitive sites found across diverse archaea. This suggests there may be other structural histone variants evolving in parallel across species similarly to the capstone.

The pattern of other substitutions is narrower, and some appear to be restricted to certain lineages. In particular, it is apparent that diversification at the DNA binding residue 54 almost exclusively occurs in the Asgard archaea. Concurrently, when this clade is excluded from the diversity analysis, the relative diversity at this position dramatically reduces (Figure 9.5).

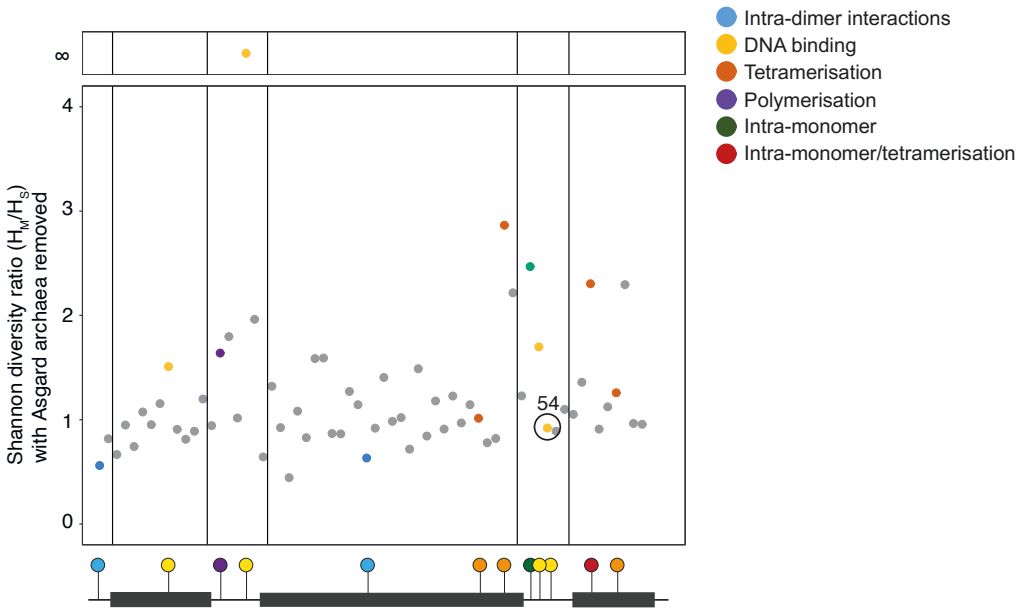


Figure 9.5 Sequence diversification of archaeal histones without Asgard archaea dramatically reduces diversity at position 54.
Shannon diversity ratio (H_M/H_S, archaea with multiple histones v those with only one) without the Asgard archaea at each position in the core histone fold. This figure was adapted from (Stevens *et al.* 2020).

9.4. Diversification of eukaryotic histone sequences in comparison to archaeal histones

How does this sequence diversification compare to histones from eukaryotes? The structure of archaeal and eukaryotic histones is highly similar (Mattiroli *et al.* 2017), and some residues

have conserved interactions. For example, the residues involved in tetramerization of archaeal histones are also important for inter-dimer interactions between individual H3 histones (Luger *et al.* 1997, Postberg *et al.* 2010). In eukaryotic nucleosomes, the additional constraint of the corset that is the nucleosome likely restrict certain positions more than in flexible archaeal histones. Given the structural similarities, do eukaryotic histones diversify at the same sites as archaeal histones?

To compare the sequence diversification of eukaryotic histones to archaeal histones, I added sequences of eukaryotic histones H3 and H4 to the archaeal histone alignment using MAFFT (See Appendix Table 14.3 for relative positioning for each group in the alignment). Histone sequences matching the eukaryotic core histone folds were obtained from Uniprot (Bateman *et al.* 2021) and filtered to remove those originating from non-eukaryotic species (see Methods). This left some larger proteins which were clearly not histones (Figure 9.6a, b). Size thresholds were used to restrict the dataset to retain histone proteins (Figure 9.6a, b).

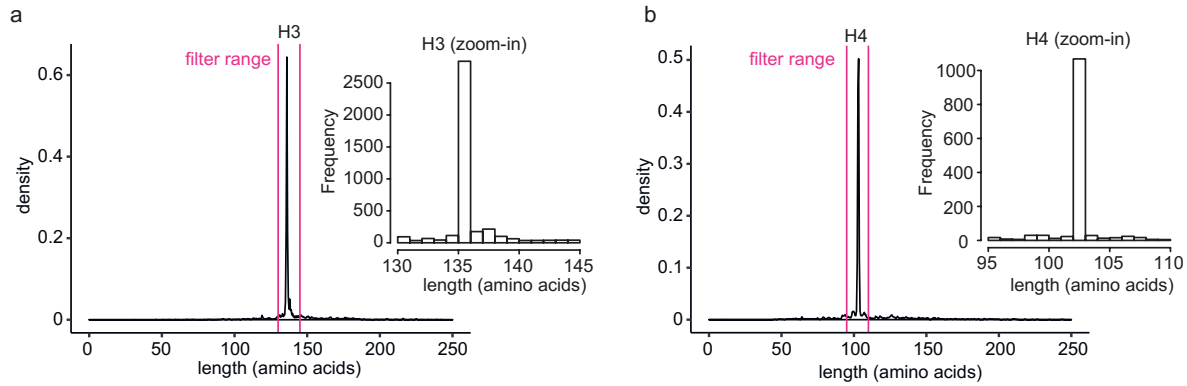


Figure 9.6 Length filtering of eukaryotic H3 and H4 histones.
The distribution of protein length (in amino acids) for a) H3 and b) H4 is shown before and after filtering. The size thresholds used to filter each group is shown as pink lines. This figure was adapted from (Stevens *et al.* 2020).

The alignment of archaeal and eukaryotic histones was divided into the following groups; archaea with one paralog, archaea with two or more histone paralogs, eukaryotic histone H3, and eukaryotic histone H4. Shannon diversity was calculated for H3 (H_3) and H4 (H_4). Diversity at conserved sites in the histone fold for H_3/H_S and $H_4/H_{S\,\mathrm{was}}$ then compared to that of archaeal histones (H_M/H_S).

In archaea with multiple histones, H3 and H4, the patterns of diversification are similar (Figure 9.7a). This trend is more apparent for H3 (rho=0.40, P=0.0081) than for H4 (rho=0.24, P=0.063).

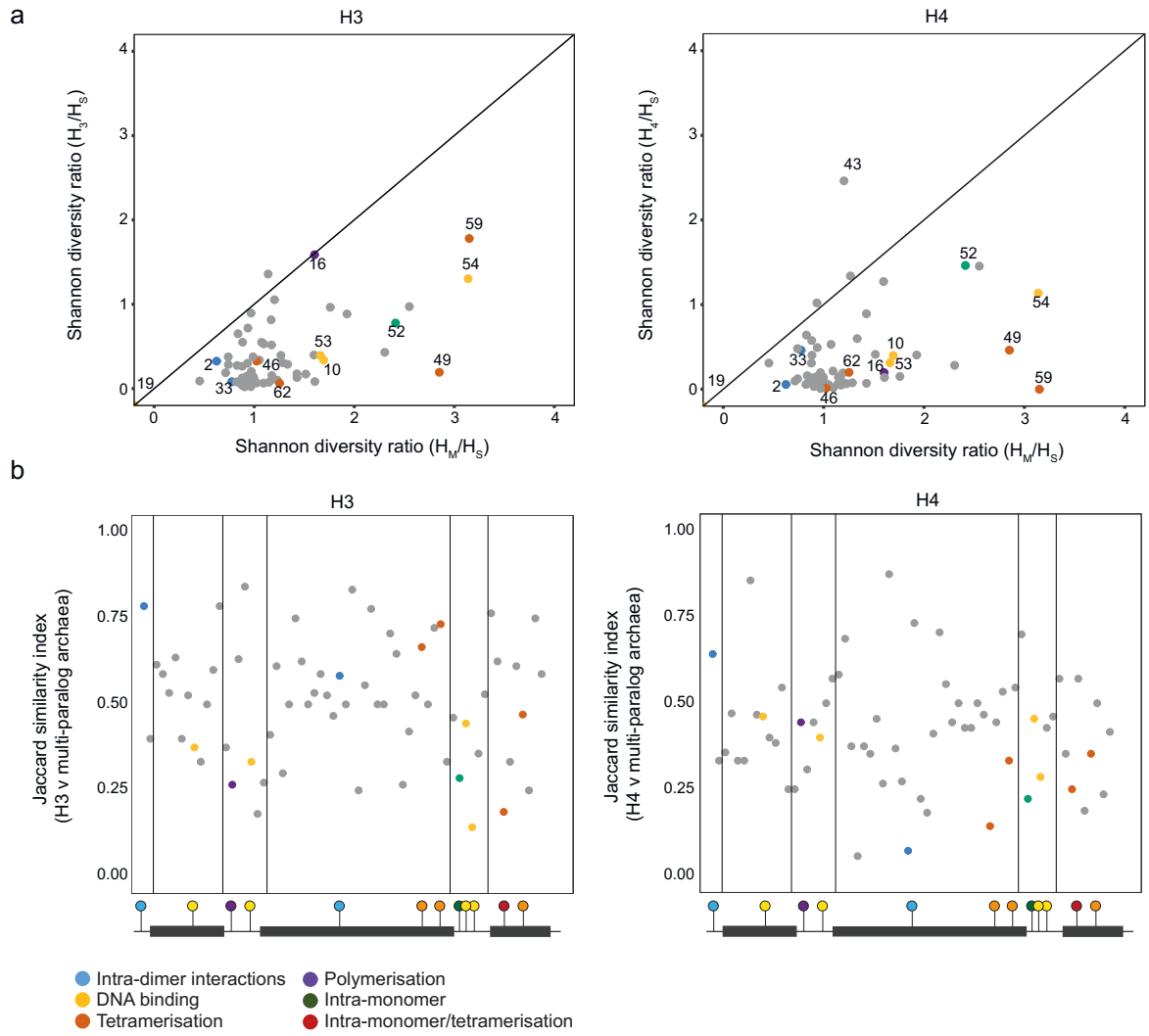


Figure 9.7 Diversification of eukaryotic histones H3 and H4 in comparison to archaea with multiple histone copies.
a) Shannon diversity ratio for H3 (H_3/H_S) and H4 (H_4/H_S) compared to archaea (H_M/H_S). b) Jaccard similarity index for H3 and H4 compared to archaea, showing the extent of overlap in which amino acids are found at each position between each group. Structurally sensitive residues are highlighted in colour, and are numbered in a) and shown on the secondary structure for b). This figure was adapted from (Stevens *et al.* 2020).

Some sites show significant diversification in eukaryotic histones, but not archaeal histones (and vice versa). This includes residue 43, which is highly diverse in H_4/H_S but not H_M/H_S. This residue is not implicated in DNA binding or tetramerization based on our *in silico* mutational analysis (Figure 9.2a, b). Residues involved in tetramerisation appear to be more diverse in

archaea than in eukaryotes (Figure 9.7, 9.2b). In particular, residue 49 is diverse in archaeal histones but not eukaryotic histones (high H_M/H_s, low $H_{3/4}/H_s$), and residue 59, which is diverse in archaea but not histone H4 (high H_M/H_s, low H_4/H_s). There are also differences in the amino acid landscape explored by eukaryotic histones compared to archaeal histones. When comparing which amino acids are found at each site using the Jaccard index (Jaccard 1908), there is generally a different set of amino acids at sites which are diverse in both groups (Figure 9.7b). For example, at position 52, the most common amino acid found for H4, H3 and archaea is arginine (R) (Postberg *et al.* 2010), but we also find diversification to cysteine (C) and lysine (K) in H4, and K in H3, whereas archaeal histones have diversified to alanine (A), K, isoleucine (I), leucine (L) and valine (V).

9.5. Summary

Probing the crystal structure of HMfB *in silico* revealed sites which display deviations in DNA binding and tetramerization upon mutation, in accordance with prior experimental work. In archaea with multiple histone paralogs, diversification has primarily occurred at these structurally sensitive sites. This includes position 49, which is key for the capstone histone functional properties. The relatively high diversification at these sites in archaea with multiple histones suggests a significant change in the evolutionary regime compared to archaea with one histone (Stevens *et al.* 2020). The strong change from the average baseline of diversity, particularly at sites which are functionally constrained, implicates positive selection as opposed to just relaxation of functional constraint (Stevens *et al.* 2020).

Histones from both eukaryotes and archaea have diversified in the loop 2 region in particular, but with a higher constraint in eukaryotes. Why do some eukaryotic residues not diversify, and why have histone variants in each domain diversified to a different sequence set?

Within eukaryotic lineages, core histones are highly conserved to preserve nucleosome structure and maintain interactions with machinery, therefore many mutations are likely disruptive and/or lethal (Amatori *et al.* 2021). This can be seen for some of the relatively more conserved sites. For example, a comprehensive mutational analysis of histones H3 and H4 in *S. cerevisiae* documented the effect of substitutions to alanine (and alanine to serine) (Dai *et al.* 2008). Although many mutations are viable – particularly in the tail regions, suggesting retention of the general function of the nucleosome – those at sites interacting with DNA, and

some in the hydrophobic core, tended to be lethal including positions 49 and 59 in H3 and H4, which are relatively less diverse than in archaea (high H_M/H_s, low $H_{3/4}/H_s$) (Dai *et al.* 2008). Disruptive effects can also be seen in disease. Mutations at E97 (position 33 in HMfB) are one of the most frequent H3 mutations in a broad range of main cancer types and are highly conserved in H3 (Figure 9.7) (Nacev *et al.* 2019). Although not explored extensively here, many of the species responsible for the high diversity at structurally sensitive sites (e.g., position 52) in H3 and H4 are parasitic e.g., *Leishmania* species.

Even at sites which appear to have similarly diversified, the sequence landscape explored is significantly different in archaeal and eukaryotic histones. Although there are similarities in structure, the overall constraints of eukaryotic and archaeal histones therefore appear to be different. The consequence of assembly with DNA into octamers (Kornberg 1974, Luger *et al.* 1997) and requirements for dedicated chaperones, reader, writers and modifiers (Grau-Bové *et al.* 2022), may impose stricter requirements for amino acid identity at certain sites in eukaryotes. In this regard, it is important to note here that the eukaryotic H3 and H4 histones are all bona fide packaging agents. Whether the same is true for all of the archaeal histones is unknown. The comparatively flexible higher-order structure of archaeal histones, and lack of dedicated machinery, may remove some of these restraints but impose others. This suggests that eukaryotic and archaeal histone variants have independently explored the potential landscape of structure/function at structurally sensitive sites, evolving distinct properties in parallel (Stevens *et al.* 2020).

10. Exploration of recombinant *M. stadtmanae* histone Msp_0769 and capstone Msp_0383

10.1. Introduction

The diversification of archaeal histones at structurally sensitive sites suggests transition to histone variants with more extreme structural properties, exemplified by capstone from *M. stadtmanae* (Msp_0383) (Stevens *et al.* 2020). *In silico* tools hint at this functionality, but do not definitively show this capping property. I therefore sought to characterise the function of the capstone experimentally. This work was done in collaboration with, and under the supervision of, Antoine Hocher.

Like many archaea, *M. stadtmanae* is not currently genetically accessible, presenting a significant hurdle to full *in vivo* characterisation of each histone. The use of *in vitro* tools, and characterisation of recombinantly expressed histones, can alleviate this issue. In this regard, prior work in the lab (primarily by Maria Rojec) has established *E. coli* as a model naïve prokaryote for chromatinization by the histones HMfA and HMfB from *Methanothermus fervidus* (Rojec *et al.* 2019). Recombinantly expressed HMfA+B histones bound to, and protected, genomic DNA in *E. coli* from digestion in complexes ranging from 60 to >150 bp, consistent with those formed *in vivo* in *M. fervidus* (Rojec *et al.* 2019). This provides a good starting system in which to compare the properties of *M. stadtmanae* histones and the capstone. I therefore used recombinant expression and *in vitro* experiments to try to unpick the differences between two histones from *M. stadtmanae*; the capstone Msp_0383, and the histone with the highest pI and highest predicted DNA binding, Msp_0769 (See Chapter 6).

The results in this chapter do not fully characterise the histones in *M. stadtmanae,* but rather provide a platform for further research. These experiments were initially undertaken during Covid restrictions, and also at the end of my PhD. This chapter therefore represents a work in progress. The experiments presented here were done with Antoine Hocher and are referenced where appropriate in the figure legends.

10.2. Transformation of *M. stadtmanae* histones Msp_0383 and Msp_0769 into *E. coli* and novobiocin phenotype

In the common laboratory *E. coli* strain K-12 (MG1655), HMfA and HMfB are highly expressed, and can chromatinise the genome with only mild growth defects (Rojec *et al.* 2019). I also used this strain to express histones Msp_0769 and Msp_0383 from *M. stadtmanae*. Although *M. fervidus* and *M. stadtmanae* are both methanogens, the optimum growth temperatures for each are vastly different (See section 6.3). Other *E. coli* strains have also been used: the *E. coli* strain B834(DE3) was used to purify histones from mesophilic archaea, including *M. jannaschii* (Li, Sandman, *et al.* 2000) and *Methanobacterium formicicum* (Li, Shriver, *et al.* 2000). The B843(DE3) strain, and daughter BL21(DE3), are commonly used to express toxic proteins as they are deficient in proteases Lon and OmpT which degrade misfolded proteins (Wood 1966). BL21(DE3) was therefore also used as a background to express *M. stadtmanae* histones Msp_0383 and Msp_0769.

pD861 plasmids encoding codon-optimised HMfA and HMfB sequences were used for recombinant histone expression in *E. coli* (Rojec *et al.* 2019). In this set-up, the histones are under the control of a rhamnose-inducible promoter and carry a kanamycin resistance gene for selection. Maria Rojec, a former PhD student, generated plasmids with an identical backbone to that used for HMfA+B, with *M. stadtmanae* histones Msp_0383 and Msp_0769 codon optimised for expression in *E. coli* by ATUM (Newark, CA). Maria Rojec transformed these plasmids into *E. coli* K-12 MG1655, and I transformed each plasmid into *E. coli* BL21(DE3) using heat shock. All strains were verified using PCR. *E. coli* K12 expressing either Msp_0383 or Msp_0769 are further referred to as K12-msp0383 and K12-msp0769, respectively. *E. coli* BL12(DE3) expressing either Msp_0383 or Msp_0769 are further referred to as BL12-msp0383 and BL21-msp0769, respectively.

E. coli expressing HMfA have a novobiocin sensitivity phenotype, and show reduced growth upon exposure (Rojec *et al.* 2019). This sensitivity is not seen in strains expressing non-binding HMfA (with mutations at DNA binding residues; K13T, R19S, T54K shown to stop DNA binding in HMfB (Soares *et al.* 2000)) or the empty pD861 vector (Rojec *et al.* 2019). Novobiocin inhibits DNA gyrases – enzymes which constrains DNA in negative supercoils in *E. coli* – resulting in supercoiling stress. Rojec *et al*. (2019) found that *E. coli* expressing HMf histones exhibited downregulation of gyrases, and transcriptome changes similar to that seen

when cells are exposed to gyrase poison. The lower expression level of gyrases could compensate for an increase in negative supercoiling imposed by the introduction of HMf histones (Rojec *et al.* 2019). Although this may help cells to adapt to chromatinisation with histones, this may make cells less tolerant of novobiocin (Rojec *et al.* 2019). This stress test gives an initial insight into whether the histones from M. stadtmanae may impose the same supercoiling stress as HMfA.

Using the same protocol and strains from Rojec *et al.* (2019) as controls, I find that K12-msp0769 displays similar sensitivity to novobiocin as HMfA, and shows reduced growth when exposed to the antibiotic (Figure 10.1). In contrast, the growth of K12-msp0383 does not reduce upon novobiocin addition, and the growth phenotype is similar to the non-binding HMfA, or empty vector (Figure 10.1). This suggests that, at least on a phenotypic level, histone Msp_0769 may play a more similar role to HMfA than Msp_0383 in *E. coli*.

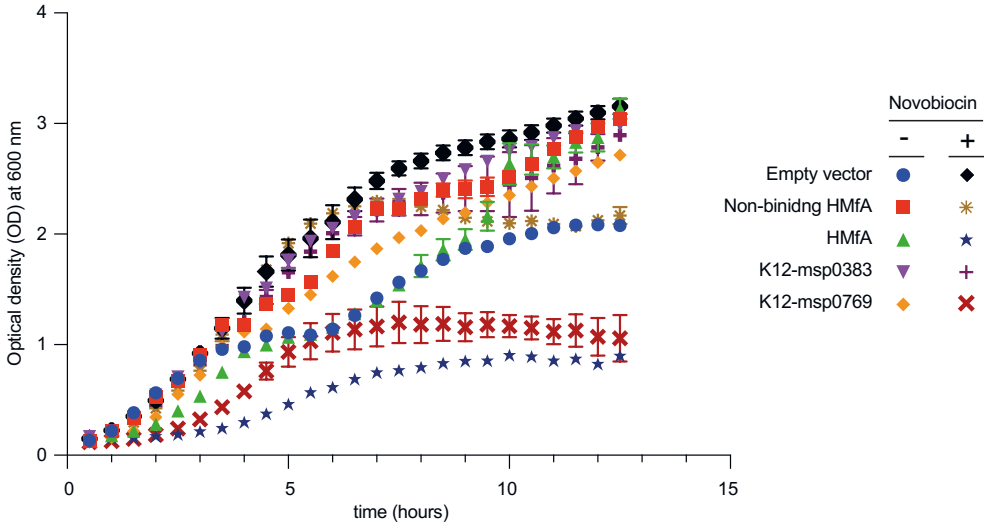


Figure 10.1 Novobiocin sensitivity assay for *E. coli* strains expressing *M. stadtmanae* histones Msp_0769 and Msp_0383.
E. coli strains expressing HMfA (HMfA), non-binding HMfA (Non-binding HMfA) and the empty vector containing no histones (Empty vector) from Rojec et al. (2019) were used as controls, alongside *M. stadtmanae* histones Msp_0383 and Msp_0769 from *M. stadtmanae* in the same *E. coli* background (K12) (K12-msp0383 and K12-msp0769, respectively). *E. coli* strains expressing histones or the empty plasmid were pre-induced and added to LB medium either with or without novobiocin. Growth for each strain was measured every 30 minutes over 15 hours, and is shown as optical density (OD) at 600 nm. These experiments were done in collaboration with Antoine Hocher.

10.3. MNase digestion of *E. coli* chromatin expressing Msp_0769 and Msp_0383

To determine whether the phenotypic differences between Msp_0769 and Msp_0383 can be traced back to altered DNA binding, I performed micrococcal nuclease (MNase) digestion of the histone-expressing *E. coli* strains. MNase is an endo-exonuclease which degrades double- or single-stranded DNA, with a slight preference for AT-rich sequences (Hörz and Altenburger 1981). Regions of the DNA tightly bound by proteins resist digestion, resulting in fragments of the size wrapped by the protein after digestion and removal of proteins.

First, expression tests were performed in a range of conditions to determine the optimum conditions for expression of each histone. For reference, the theoretical molecular weight (Mw) of each protein, as calculated by the Compute pI/Mw tool on the ExPASy server (http://web.expasy.org/protparam/, accessed 31-07-20) (Bjellqvist *et al.* 1994), is as follows: HMfB = 7.67 kDa, Msp_0383 = 7.42 kDa, Msp_0769 = 7.43 kDa. Note that the archaeal histones tend to run at a lower Mw than predicted. In K12-msp0383, the histone is relatively highly expressed, at levels similar to those seen for HMfA and HMfB in almost all conditions (Rojec *et al.* 2019) (Figure 10.2a, b, c). In contrast, Msp_0769 is lowly expressed in the *E. coli* K12 background (K12-msp0769, Figure 10.2a, b). Neither reducing the temperature nor the addition of sorbitol increased expression levels of Msp_0769 (Figure 10.2b). In BL21-msp0769, expression levels of Msp_0769 were slightly higher than in K12-msp0769, but still much lower than Msp_0383 (Figure 10.2c).

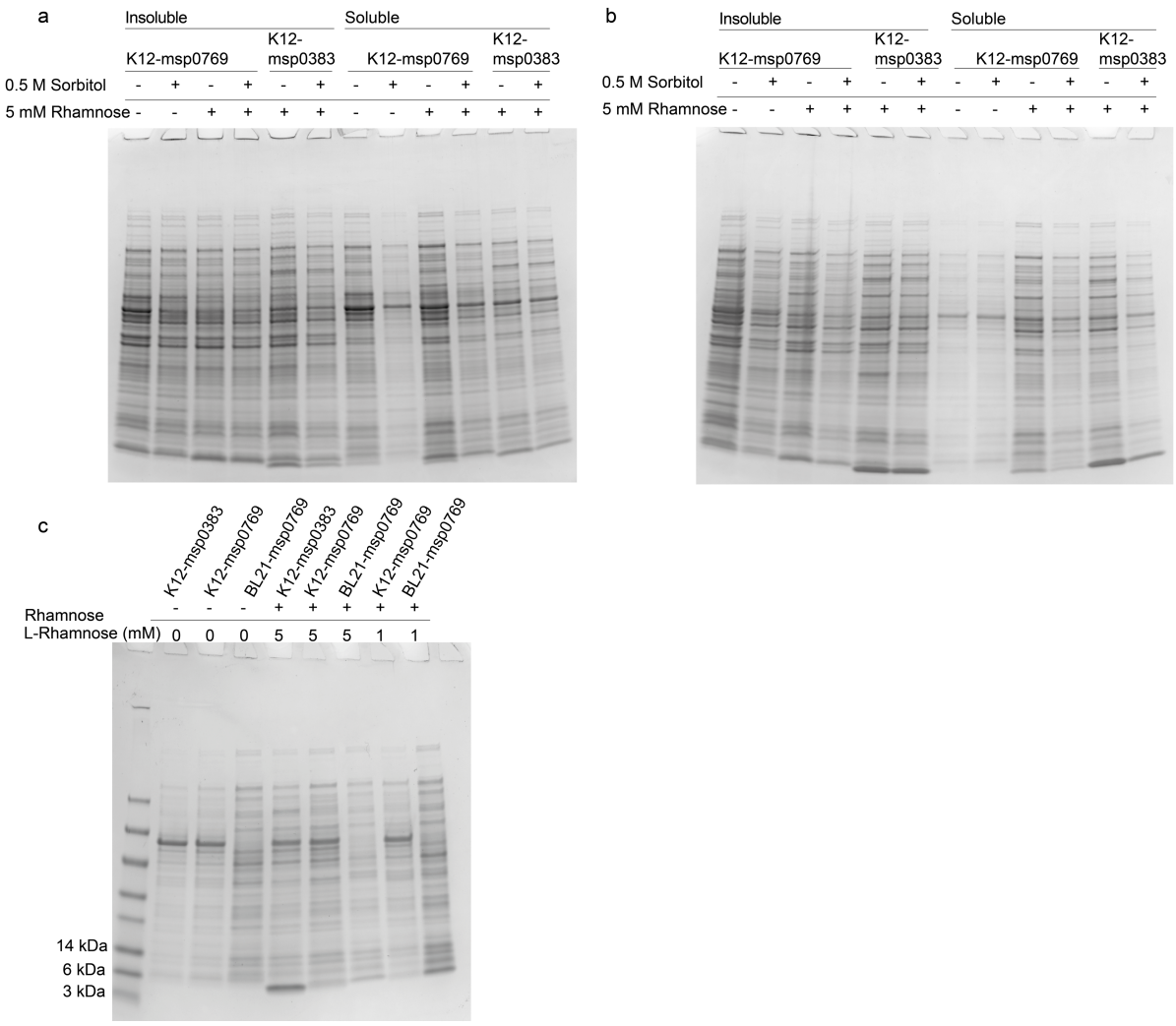


Figure 10.2 Expression tests for *E. coli* strains expressing histones from *M. stadtmanae*.
Proteins found in the insoluble and soluble fraction of lysed exponential phase cells is shown for cells with (+) and without (-) expression of the histones. q) Proteins from K12-msp0383 and K12-msp0769 strains in the insoluble and soluble fractions. Expression was induced with 5 mM rhamnose at room temperature for 16-17 hours with or without 0.5 M sucrose. b) As in a), with the exception of induction conditions. Expression of histones was induced by addition of 5 mM rhamnose for 4 h at 37°C c) Proteins found in the soluble fraction of K12-msp0769, BL21-msp0383 and BL21-msp0769. Expression was induced by the addition of either 0 mM, 1 mM or 5 mM rhamnose for 4 h. These experiments were done in collaboration with Antoine Hocher.

MNase digestion was then used to investigate whether the histones Msp_0383 and Msp_0769 appeared to bind to and protect genomic DNA in the *E. coli* K12 background from digestion, as is seen for HMfA and HMfB (Rojec *et al.* 2019). I could recapitulate the 30 bp banding pattern for *E. coli* expressing HMfA, both using fixed cells (Figure 10.3b, method from Rojec *et al*. (2019)) and unfixed cells subject to mechanical lysis (Figure 10.3a, protocol from Hocher *et al*. (2019)). Higher order fragments (over 300 bp) corresponding to large oligomers were observed frequently in unfixed chromatin (Figure 10.3a). Despite the novobiocin sensitivity phenotype for Msp_0769, unlike HMfA and HMfB, there was no clear protection from MNase

for either of the two histones from *M. stadtmanae* using both methods. When fixed chromatin was used, I found a smear for both Msp_0769 and Msp_0383 similar to *E. coli* strains transformed with an empty vector and not indicative of non-specific binding (Figure 10.3b, this result was previously found and documented in Rojec *et al*. (2019)). The lack of ladder pattern for Msp_0769 could be due to relatively low expression levels in *E. coli* K12 (strain k12-msp0769, Figure 10.2c).

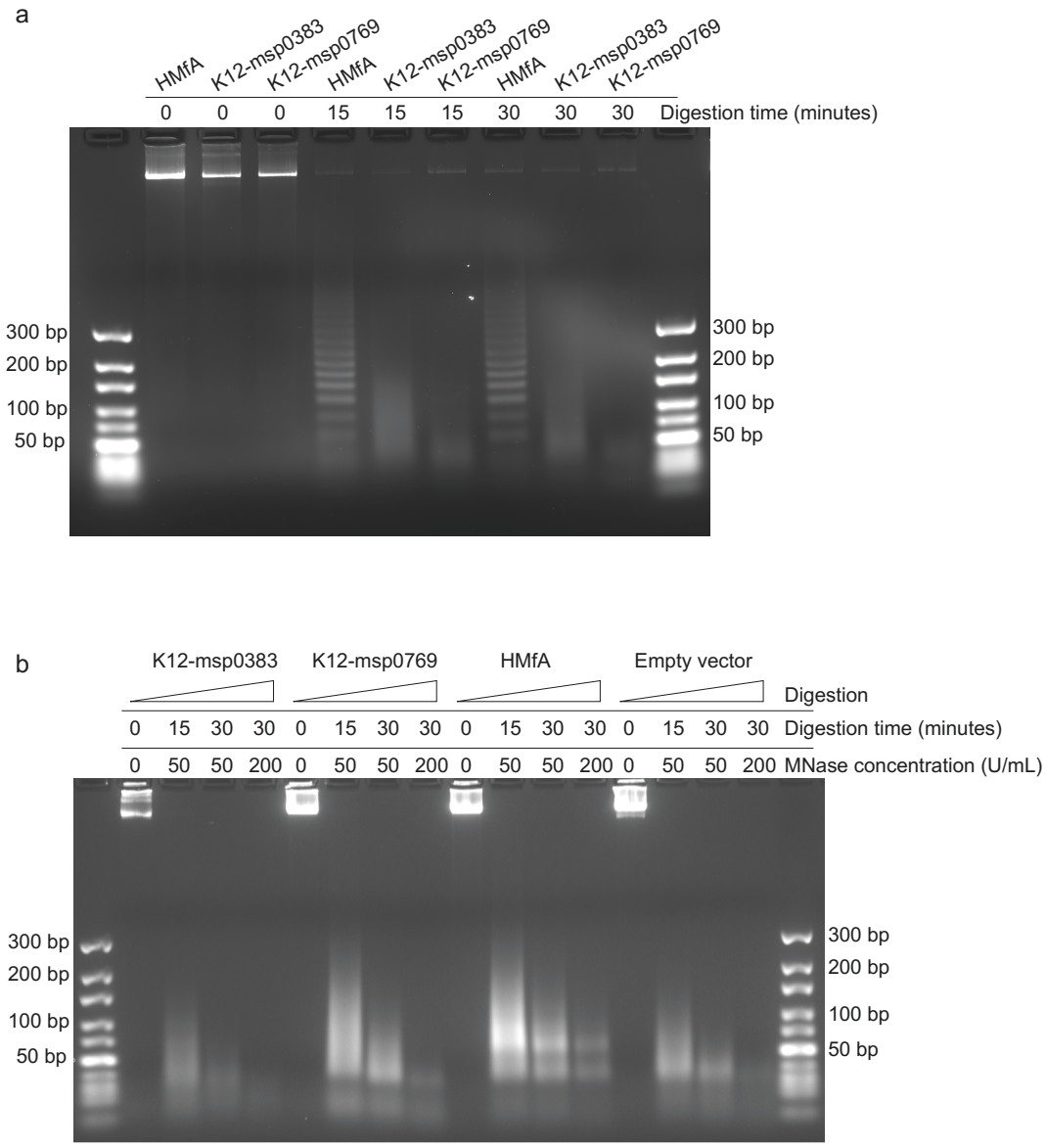


Figure 10.3 in vivo MNase digestion of *E. coli* K12 strains expressing histones HMfA from *M. fervidus* and Msp_0383, Msp_0769 from *M. stadtmanae*.
a) Unfixed chromatin was isolated from *E. coli* expressing archaeal histones using mechanical lysis and was digested for 0, 15 or 30 minutes with MNase. The resulting banding pattern of DNA protection is shown for each. This experiment was done under the supervision of Antoine Hocher. b) Fixed chromatin isolated from *E. coli* expressing archaeal histones subjected to increasing digestion time and MNase concentrations (Rojec *et al.* 2019).

The inability to use identical *E. coli* expression systems to characterise histones from *M. stadtmanae* and *M. fervidus* is perhaps unsurprising as we know there are likely stability differences between the two. Histones in hyperthermophilic archaea, including *M. fervidus*, have been shown to be very heat stable compared to those in mesophiles. After incubation at 95^{o}C for 4 hours, histones from hyperthermophiles *M. fervidus* (HMfA and HMfB) and *Pyrococcus* (rHPyA1) still retain full DNA binding activity in high salt conditions (Soares *et al.* 1998). Circular dichromism assays show that the stability of archaeal histones increases with optimum growth temperature (Li *et al.* 1998, Topping and Gloss 2004).

I therefore repeated the MNase digestion for BL21-msp0769 strain to see whether the slightly higher expression level could elucidate the binding mode (as described above). Unlike HMfA, there is no strong ladder pattern after digestion for BL21-msp0769 (Figure 10.4a). There is some undigested DNA at the top of each sample, likely due to remaining unbroken cells after lysis (Figure 10.4a), making visualising smaller fragments more difficult. After removing the top of the gel, faint bands can be seen more clearly (Figure 10.4b). The presence of these bands was confirmed by running the sample on a TapeStation (Aligent, CA), revealing a ladder pattern in the BL21-msp0769 reminiscent of HMf binding footprints (Figure 10.4c/d). It is possible that the banding pattern could be contamination from HMf, but the distribution bands suggests this is unlikely. The most prominent peak of digested DNA product for HMfA (over ~120 bp) is much higher than Msp_0769 (120 bp and lower). If the DNA in Msp_0769 was HMfA contamination, a similar peak at larger size digestion products would be expected. Therefore, the digestion pattern for Msp_0769 appears to be real, albeit much weaker than that of HMfA.

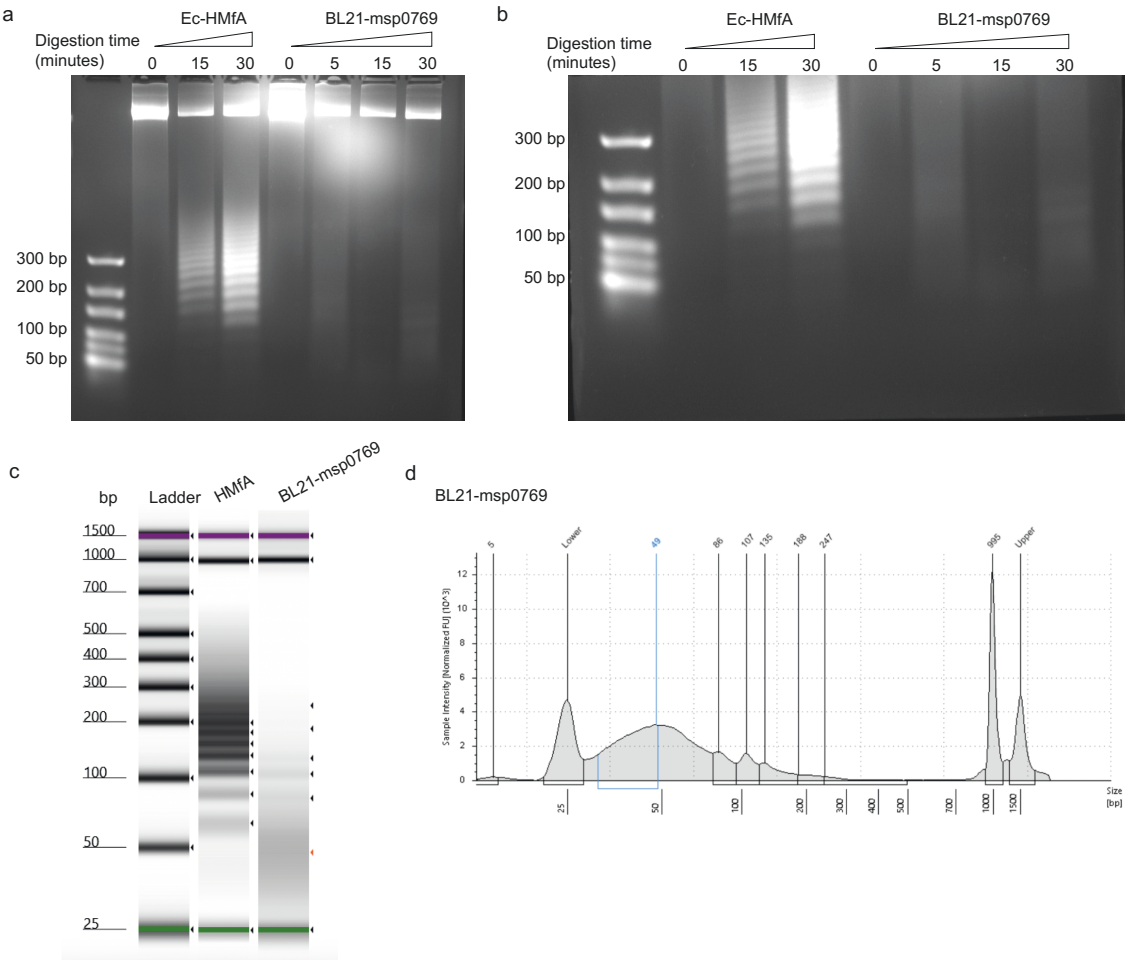


Figure 10.4 in vivo MNase digestion of *E. coli* K12 strain expressing histone HMfA from *M. fervidus* and *E. coli* strain BL21(DE3) expressing histone Msp_0769 from *M. stadtmanae*.
a) full gel, and b) cut gel showing bottom section, of MNase digestion products for each strain with increasing digestion time. c), d) TapeStation results for MNase digestion products showing the size of protected fragments in HMfA and BL21-msp0769.

The digestion pattern for BL21-msp0769 is much less clear than that of HMfA, and there appear to be few cases of stable oligomers larger in size than an octamer (~140 bp) (Figure 10.4c/d). This does not preclude the formation of longer histone-DNA oligomers by Msp_0769 *in vivo*. Msp_0769 may not be expressed at high enough level to form larger oligomeric complexes, as the expression level of this histone is still relatively low compared to HMfA (Figure 10.2c).

The relatively low abundance of Msp_0769 makes further characterisation of the function of this histone compared to Msp_0383 in *E. coli* difficult. To show the capping function of Msp_0383 in *E. coli*, a comparison between *E. coli* expressing Msp_0769 with and without the capstone would be needed. A weak ladder pattern means it would be hard to see changes to

oligomerisation when the capstone was added (Figure 10.4). I therefore decided to explore *in vitro* approaches to probe capstone function.

10.4. Purification of recombinant *M. stadtmanae* histones Msp_0769 and Msp_0383 from *E. coli*

To try to characterise histones *in vitro*, Msp_0383 and Msp_0769 were purified from *E. coli*. Given the slight increase in expression levels, the strain BL21(DE3) was used to purify Msp_0769.

Much of the prior *in vitro* work has focused on archaeal histones from extremophiles, more specifically hyperthermophiles – such as *T. kodakarensis* and *M. fervidus*. Often, protocols to purify these recombinant histones use a heat pre-clearing step. As histones from hyperthermophilic archaea tend to be heat stable, this denatures most other proteins in *E. coli*, leaving the histones (Sandman *et al.* 1994, Weng *et al.* 2004). However, histones from mesophilic archaea, such as *M. formicicum*, are not as heat stable as histones from hyperthermophiles (Darcy *et al.* 1995) suggesting this technique will not work for *M. stadtmanae*. Three HFoB histones from the mesophile *M. formicicum* have been successfully purified from cells *in vivo* (Darcy *et al.* 1995) and recombinant rHFoB has been purified from *E. coli* using a similar method (Li *et al.* 1998, Topping and Gloss 2004). Key to these purification efforts is the use of a pre-clearing step with ammonium sulphate. This precipitation step, known as salting out, can be used to separate proteins from *E. coli* lysate based on their solubility (Duong-Ly and Gabelli 2014). This approach was used to purify recombinant *M. stadtmanae* histones.

Pre-clearing with ammonium sulphate removed many non-histone proteins from *E. coli* (Figure 10.5a, 10.6a). To further remove contaminants, the pre-cleared lysate was loaded onto heparin columns. Msp_0769 appeared to bind to the heparin column, whereas Msp_0383 was found in the flow-through (Figure 10.5a, 10.6b). Fractions from the heparin column containing Msp_0769 were pooled. Msp_0383 was obtained from the flow-through of the heparin column. The purity was then further improved by loading onto anion/cation columns.

Msp_0383 has a low theoretical pI (4.64) (Stevens *et al.* 2020). Therefore, the anion and cation columns were set up so that the anion was at the top and cation at the bottom. (Klinker *et al.*

2014). In this set-up, the anion column acts as a 'waste', removing proteins with a low pI, and the cation column should bind negatively charged proteins including the histone (Klinker *et al.* 2014). The histone is expected to elute from the cation column during the chromatography run. In practice, Msp_0383 was found in the flow through from both columns, as well as in fractions 1-5 with high purity (Figure 10.5b). This could be due to an uneven charge distribution on the surface of the histone. Proteins from fractions 1-5 had a similar mass to HMfB (Figure 10.5c).

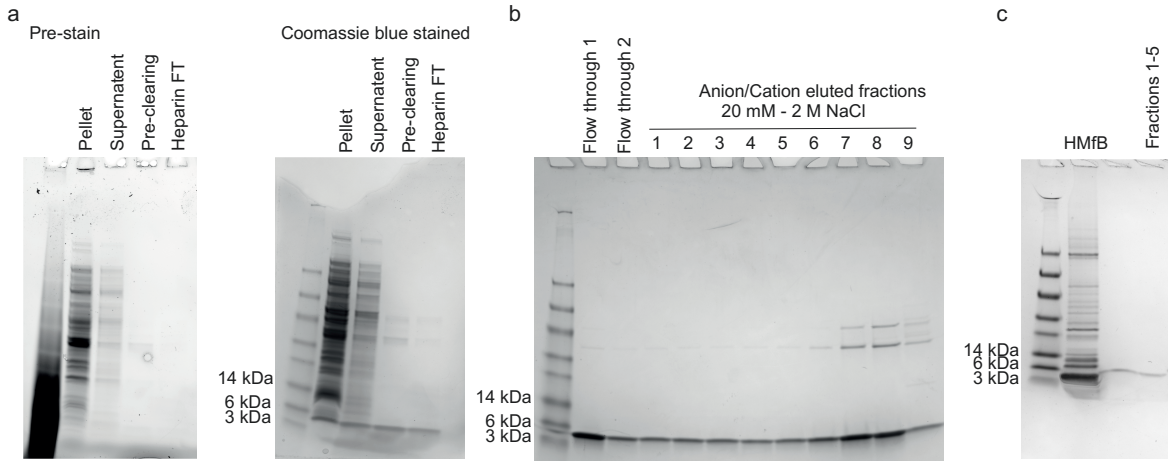


Figure 10.5 Purification of Msp_0383 from K12-msp0383.
a) Proteins from lysed cells in the soluble and insoluble fraction, after pre-clearing with 70% ammonium sulphate, and after flowing through a heparin column. Gels are shown pre-stain (L) and post-stain (R). b) Anion/cation flow through and eluted fractions. Fractions 1-5 were pooled and stored. c) The pooled fractions from b) have a similar molecular weight to HMfB purified by M. Rojec. The purification of Msp_0383 was done in collaboration with Antoine Hocher.

In comparison to Msp_0383, Msp_0769 has a higher pI. I therefore used a similar set up to purify this histone with the anion/cation in the opposite configuration (with cation as waste at the top). The histone was found in the flow through from both columns (Figure 10.6b).

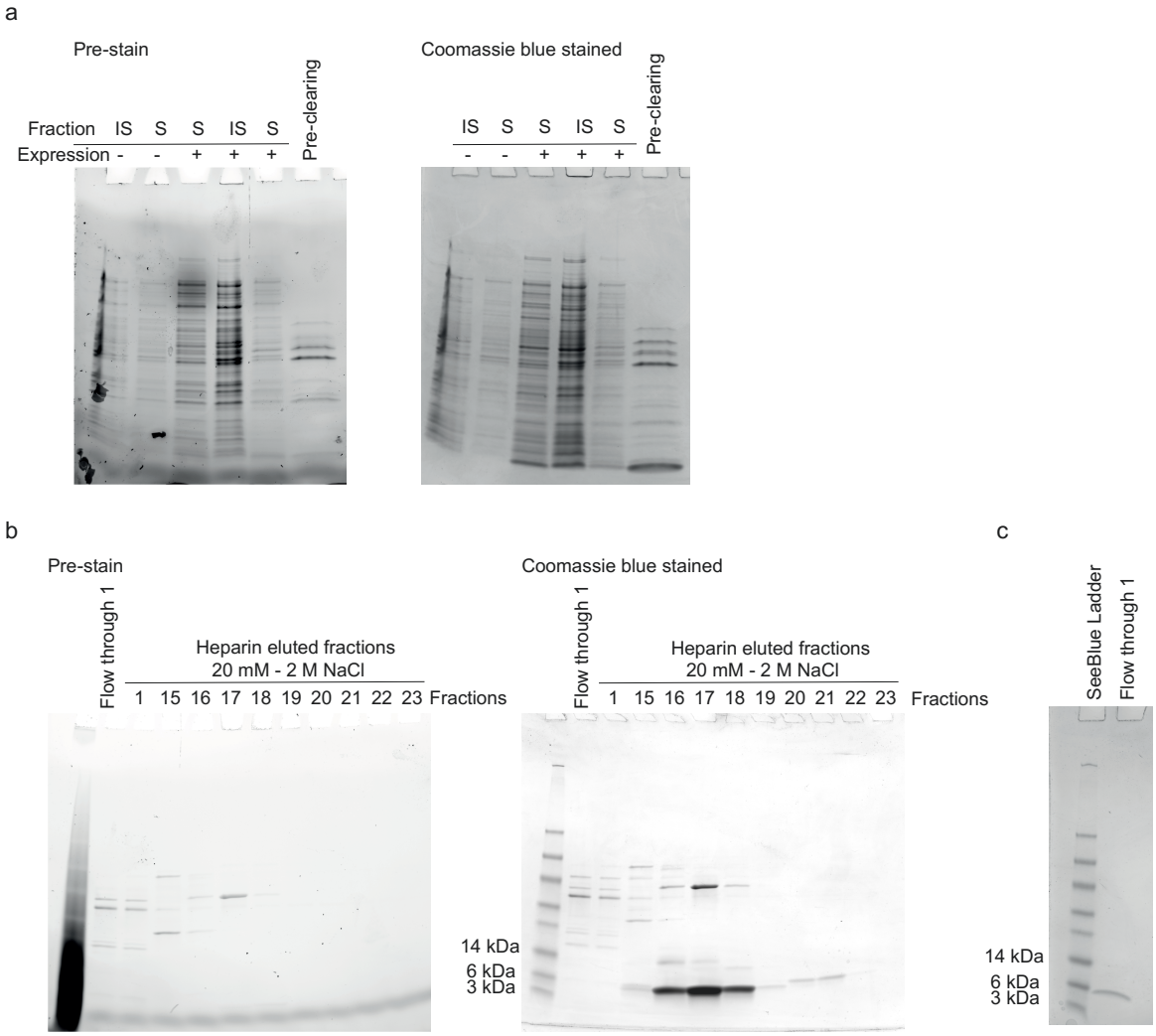


Figure 10.6 Purification of Msp_0769 from BL21-msp0769.
a) proteins from lysed cells in the soluble (S) or insoluble (IS) fractions for BL21-msp0760 with expression of the histone induced (+) or not (-). Gels are shown before (left) or after (right) Coomassie blue staining. b) Pre (left) and post (right) Coomassie blue stained 4-20% TGX protein gels showing flow through and fractions 1, and 15-23 from heparin column. Fractions 16-18 were pooled. c) Flow through from anion/cation column from pooled fractions 16-18 from the heparin column in b).

To provide a good control for the *in vitro* experiments, I also re-purified histone HMfB from stocks originally purified by Rojec *et al*. (2019). To increase the purity of the sample, an anion/cation step was added to the original sample (Figure 10.7). This process removed most contaminants.

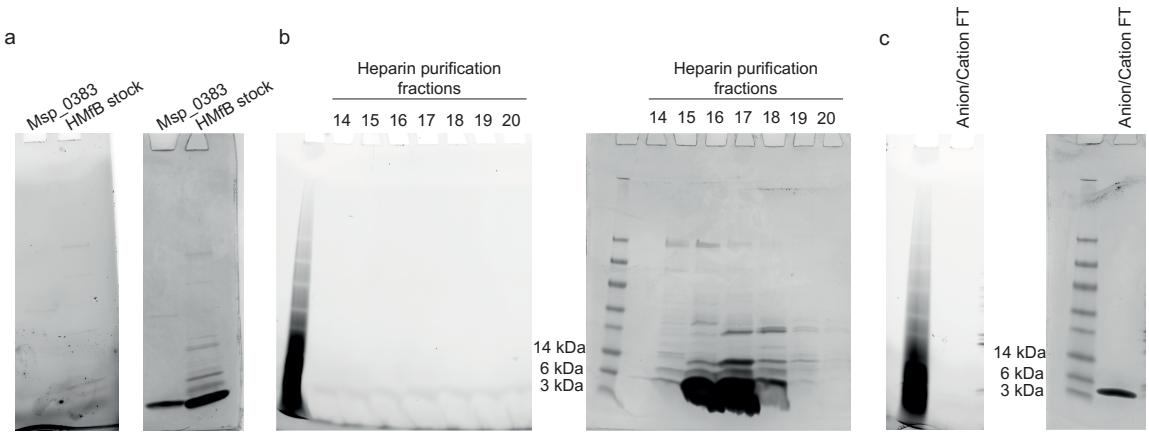


Figure 10.7 Purification of HMfB.
For a), b) and c), the left image shows gel pre-stain and right image shows post-Coomassie blue staining. a) Purified Msp_0383 from *E. coli* strain k12-msp0383 and previously purified HMfB. b) Fractions 14-20 from heparin purification of HMfB stock from a). c) Flow through from anion/cation columns for pooled fractions containing the highest concentration of protein from b). This was done under the supervision of Antoine Hocher.

The identity of each purified protein was verified by mass spectrometry, and the yield was estimated using absorbance measurements at 205 nm. The concentration for each protein was as follows; re-purified HMfB (0.499 mg/mL), Msp_0383 (0.140 mg/mL), Msp_0769 (0.0334 mg/mL). The yield for Msp_0769 was particularly low, likely due to the low expression in *E. coli*, despite attempts to increase expression level and concentrate the sample (Figure 10.2c).

10.5.Electrophoretic mobility shift assay

HMfB and other archaeal histones have been shown to bind and compact DNA in relatively low salt conditions in electrophoretic mobility shift assays (EMSA) (Sandman *et al.* 1994, Grayling *et al.* 1996, Higashibata *et al.* 1999, Marc *et al.* 2002). I therefore performed gel shifts with the purified proteins and linearised plasmid, using HMfB as a control. Neither of the histones from *M. stadtmanae* bound and compacted DNA in 100 mM KCl, unlike HMfB (Figure 10.8a).

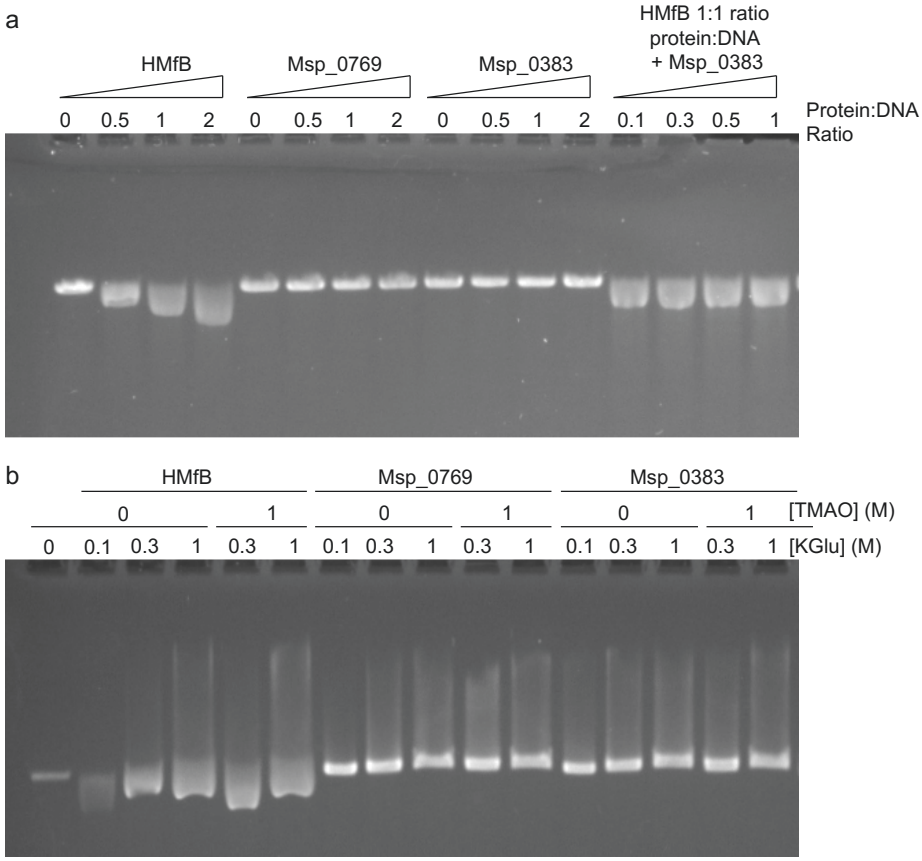


Figure 10.8 Electrophoretic mobility shift assay (EMSA) for purified histones HMfB, Msp_0383 and Msp_0769.
Histones incubated with linearised plasmid pBR322 DNA. a) Increasing ratio of protein: DNA from 0:1 to 2:1. Histones HMfB, Msp_0769 and Msp_0383 were incubated with DNA alone. HMfB and DNA at a 1:1 ratio was also incubated with increasing ratios of Msp_0383 to DNA (from 0.1 to 1) – lanes 13-16. b) 1:1 ratio of histone protein: DNA at different concentrations of TMAO and potassium glutamate (KGlu) at 100 mM KCl, 50 mM Tris pH8.

The lack of binding in low salt does not rule out a DNA binding role for *M. stadtmanae* histones Msp_0769 and Msp_0383. This low salt concentration is unlikely to be physiologically relevant to *M. stadtmanae*. From prior work, we know that different salt concentrations are required to stabilise different histones. Compared to histones from hyperthermophiles, HFoB is particularly unstable and requires higher salt concentrations to reach maximum stability (Li *et al.* 1998, Topping and Gloss 2004). At 1 M KCl, rHFoB was ~95% folded (Li *et al.* 1998) and only at 1 M TMAO HFoB had similar stability to HMfB and HPyA1 in low salt conditions (Topping and Gloss 2004). Purified histones from *M. formicicum* also produced a gel shift after incubation with DNA at 210 mM KCl with sensitivity to high temperatures (Darcy *et al.* 1995).

I therefore incubated the histones and DNA with additional salts – potassium glutamate (KGlu) and the osmolyte TMAO – which promotes protein folding (Baskakov and Bolen 1998) – at increasing concentrations, to see whether this would stabilise the histones and allow DNA binding. At high TMAO and KGlu concentrations, HMfB binds to and compacts DNA (Figure 10.8b). The level of compaction appears to vary between different salt conditions. At 1 M TMAO, 0.3 M KGlu, compaction appears greatest (biggest gel shift), and this compaction reduces without TMAO, or when the concentration of KGlu is increased to 1 M (Figure 10.8b). This could be a more general artefact of the salt concentration.

For *M. stadtmanae* histones Msp_0769 and Msp_0383, there is no evidence for DNA binding and compaction in any of the salt conditions tested (Figure 10.8b). There are slight shifts to the positions of bands between salt conditions, but this is no different from the shifts seen for DNA alone in each condition. The apparent lack of DNA binding at highly stabilising salt conditions may indicate that, unlike histones from hyperthermophilic archaea, Msp_0769 and Msp_0383 are relatively unstable and may not homodimerize spontaneously in these conditions tested. It is important to note that an exhaustive range of salt conditions was not tested. The two histones may also require another osmolyte or a co-factor to allow binding.

10.6.*In vitro* MNase digestion

The suggested functional model of the capstone requires the capstone Msp_0383 to be incorporated into a growing oligomeric complex. To determine whether addition of the capstone could alter the landscape of histone complexes, I used *in vitro* MNase assays. Based on this model, it is expected that histones with the capstone added will have an increased abundance of smaller protected DNA fragments after digestion.

In previous work, authors have successfully used *in vitro* MNase digestion to elucidate binding modes of archaeal histones, notably for *T. kodakarensis* histone HTkB (Mattiroli *et al.* 2017). A similar approach was used to probe the capstone function.

Given the difficulty finding conditions in which Msp_0769 binds in vitro and relatively low concentration of this protein (0.0334 mg/mL), this necessitated the use of HMfB and Msp_0383 for *in vitro* experiments. Other work has shown that histones from different species of archaea can universally form stable heterodimers (Li, Sandman, *et al.* 2000). Therefore, this strategy of

attempting to create heterodimers to show the capping function should be applicable to histones HMfB and Msp_0383, despite the fact they originate from different species. Concurrent with this, HMfB is predicted to form a stable heterodimer with Msp_0383 based on AlphaFold v2 structural modelling predictions (Figure 10.9a).

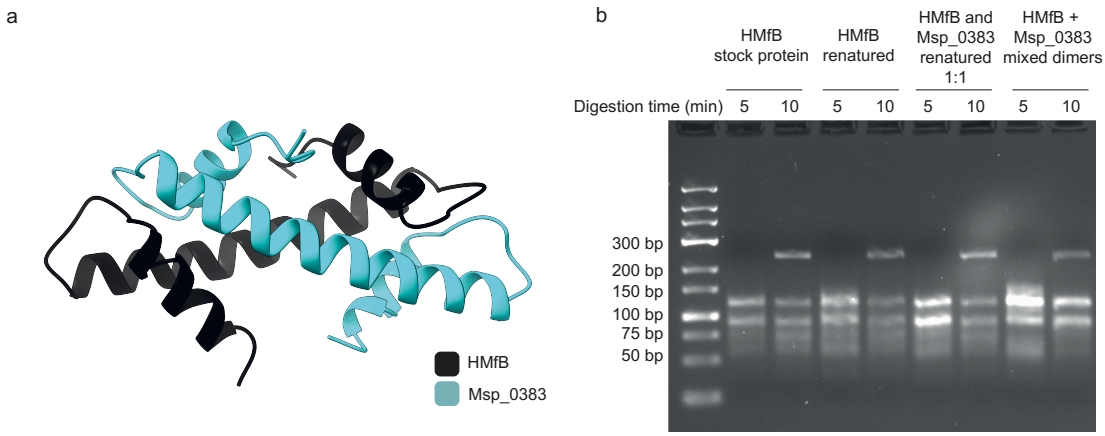


Figure 10.9 in vitro MNase digestion patterns for HMfB with and without Msp_0383.
a) Top scoring AlphaFold predicted heterodimeric structure of HMfB (black) and Msp_0383 (light blue/cyan). The structure shows a histone heterodimer in the canonical histone fold domain handshake configuration. b) Digestion was done for either 5 or 10 minutes with 1:500 diluted MNase. From left to right: purified HMfB protein, renatured HMfB, HMfB and purified Msp_0383 renatured in a 1:1 ratio of protein: protein, and HMfB and Msp_0383 mixed dimers (see methods). Except for HMfB stock protein, all samples were denatured then renatured. HMfB and Msp_0383 renatured were denatured and renatured together. HMfB + Msp_0383 were denatured and renatured separately, then mixed. The ratio of HMfB to DNA was 1:1. The in vitro MNase was done in collaboration with Antoine Hocher.

It is possible that the histones may not be properly folded in solution, and that heterodimers may not spontaneously form between HMfB and Msp_0383. Along with Antoine Hocher, I therefore first optimised the *in vitro* MNase assay for HMfB, before adding the capstone histone. For HMfA and HMfB, heterodimers spontaneously form quickly at high temperatures (95ºC) (Grayling *et al.* 1996). Histones from the mesophile *M. stadtmanae* are likely less thermostable, as discussed above, so a heating step may denature the protein. Each histone was therefore denatured using urea, before both were renatured separately, as well as together to increase the likelihood of spontaneous formation of heterodimers.

Using this approach, there was not a clear difference in digestion pattern when HMfB was incubated with DNA alone, compared to with Msp_0383 (Figure 10.9b). This is true when the histones are mixed both before and after renaturation (renatured v mixed dimers, Figure 10.9b).

10.7.Summary

The experiments described above do not provide concrete evidence that Msp_0383 has a capping property *in vitro* or in *E. coli*. However, there are hints at differences between the capstone and the most highly expressed histone in *M. stadtmanae*, Msp_0769. *E. coli* expressing Msp_0769, but not Msp_0383, display a weak ladder-like pattern upon MNase digestion, and show sensitivity to novobiocin similar to *M. fervidus* histones. Given these results, the tools used to probe the effects of chromatinization in *E. coli* using HMf do not appear to be directly applicable to histones from the mesophile *M. stadtmanae*.

The negative EMSA results, and weak MNase pattern here do not preclude DNA binding and compaction by histones in *M. stadtmanae*. The salt conditions tested here are likely substantially different from those in the cell, which are unknown. Requirement of specific salt, co-factors or ions for binding has implications for the modelling in Chapters 5 and 6, which was done in neutralising NaCl conditions. Whether histone binding changes in different salt conditions requires further work. It is also possible that histones could require modification for DNA binding, or that the histones were misfolded in *E. coli* and did not renature correctly in our *in vitro* experiments. The requirement for a chaperone for binding is unlikely, as all archaeal histones studied appear to spontaneously dimerise *in vitro*.

The timing of these experiments did not allow a full investigation into the *M. stadtmanae* histones, but I have included these preliminary results here to document my approach for future characterization of the full complement of seven histones *in vitro*.

11. Discussion

In archaea, histones are widespread (Figure 2.6) and in some species are one of the most abundant proteins (Rojec *et al.* 2019, Hocher *et al.* 2021). The variable DNA binding affinities, difference in expression level throughout growth and transcriptomic/phenotypic changes upon deletion from prior work suggested that histone paralogs have diverse functions in archaea (see section 2.8). The implications of these differences and comparison to the diversity of histones in eukaryotes previously remained largely unexplored. In this work, structural differences between paralogs and phylogenetic patterns reveal histones with unique structural properties, and the presence of distinct paralogs shared between species with a common ancestor hundreds of millions of years ago, similar to eukaryotic histone variants (Stevens *et al.* 2020, 2022). Across archaea, both the composition of NAPs and those identified in the chromatin by mass spectrometry (Maruyama *et al.* 2020, Hocher *et al.* 2021) can vary substantially, making the conservation of distinct paralogs across species separated by long time periods even more striking. The biological importance of each variant at the cellular level remains unknown, but it appears that individual histone paralogs are likely unique and important.

Although the classification of archaeal histones into variant classes in this work is based on structural and phylogenetic patterns which bear resemblance to that in eukaryotes, variants in the two groups likely operate by different mechanisms. The differences in higher-order complex formation between archaea and eukaryotes is particularly important to consider when thinking about the nature of histone variants in archaea (Stevens and Warnecke 2022). If promiscuous dimerization and tetramerization allows archaeal histones to occupy *any* position in the histone-DNA complex, variants are not defined by position (see section 11.2.2). That is to say, there are unlikely to be specific variants which occupy a single spatial niche in the oligomer, unlike eukaryotic variants which can only replace their canonical counterpart – such as H2A.X replacing H2A in the nucleosome. (Stevens and Warnecke 2022). Archaeal histones must therefore be considered separately to eukaryotic histone variants.

My work presented in this thesis provides only a starting point to understanding histone variants in archaea. Below, I will highlight key evolutionary and physiological questions that remain to be addressed.

11.1. Outstanding questions

11.2. The role of histones in archaea in general

Eukaryotic nucleosomes are an essential complex playing many roles outside of compaction, affecting DNA replication, repair, expression and organisation of the genome. Not only does the wrapping of DNA into nucleosomes generate a barrier for many processes, but this assembly also provides a platform to integrate external information. These roles are linked by an entrenched interplay between histones, PTMs, transcription factors and other associated proteins. Is this also true in archaea? A comprehensive overview of chromatin proteins showed that the diversification of eukaryotic histone variants, chaperones and readers likely occurred *within* eukaryotes (Grau-Bové *et al.* 2022). The same tight association is therefore less likely in archaea, and a global role for archaeal histones is not well supported, given the differences in expression level across species and variable essentiality (Rojec *et al.* 2019, Sanders, Marshall, *et al.* 2019).

Differences in the need for an integrated histone-centric system between eukaryotes and archaea is not unexpected. Genome organisation is dramatically different between the two kingdoms; archaea have smaller circular genomes, with a high density of genes (Brochier-Armanet *et al.* 2011, Sela *et al.* 2016). One of the main drivers of NAP abundance more generally in archaea appears to be growth temperature, suggesting a role in adaptation to high temperature environments (Hocher *et al.* 2021). However, there is still the capacity for chromatin complexity in archaea through exchanging histone variants, PTMs and combinations of other NAPs.

One eukaryotic-like mechanism for histone diversification not explored in this thesis are the tails found in a select number of archaea. Charged histone tails are a key feature in eukaryotes and represent an important thread holding together the association between different chromatin regulators (e.g., PTMs and other associated proteins). Tails are less common in archaea, predominantly found in the Asgard archaea. These are dissimilar to eukaryotic histones in sequence (Henneman *et al.* 2018, Grau-Bové *et al.* 2022) but could provide a platform similar to eukaryotic histones to integrate information and alter the dynamics of histone-DNA complexes, as well as simply stabilising DNA interactions. Whether this holds true will be an interesting future avenue of research. Further sequencing in this important group may also

elucidate patterns of inheritance for lysine-rich histones, and determine whether these can be considered as variants in their own right.

11.2.1. What drives change in the number of histones?

The number of histones in archaea is highly variable (Figure 2.6). This is particularly pronounced in the Methanobacteriales members, where closely some species have many more histones than their neighbours (Figure 7.1). Growth temperature may dictate overall levels of chromatinization, but the number of histones itself does not necessarily correlate with overall chromatinization, as not all histones are necessarily expressed to the same level (Rojec *et al.* 2019). Indeed, growth temperature does not appear to drive the increase in histones in the Methanobacteriales, which are mesophilic. Both *M. stadtmanae* and *M. smithii* have more histones (with seven and three histones, respectively) than the hyperthermophilic *M. fervidus* (two histones). There is no clear link with genome size, nor obvious change in environment more broadly for the species with an increased number.

What favours the expansion of histones in some clades over gaining other NAPs? In some cases, this diversity could be driven by specific adaptive requirements. The histone fold is clearly amenable to change, exemplified by the vast array of functionally diverse eukaryotic histone variants and diversification at key structural sites in archaea (Chapter 9). If this feature is more pronounced in histones compared to other domains, this may have aided the adoption of this fold over others for differential chromatin regulation.

The patchy gain of putative capstones across archaea is particularly puzzling. If this variant is needed to control oligomer size, how then do species without the capstone (such as *M. fervidus* or *T. kodakarensis*) limit higher-order complex formation? Without knowing their physiological or regulatory role, it is difficult to answer this fully. There are no unambiguous trends in relation to GC content or environment and the small number of species with capstones limits statistical support. Understanding the propensity of histones from different species to form oligomers, and the extent to which they exhibit breathing motions, may go some way to beginning to answer this question.

11.2.2. Archaeal histone variants: a global or a local role?

One of the key outstanding questions about archaeal histone variants is whether they are deposited in a global or local manner, and how this could be controlled. For the purposes of predicting the difference in chromatin between different growth phases in *M. stadtmanae*, the assembly of histone tetrameric complexes was predicted based on expression level alone, and changes thought of on a global scale. Altering the proportion of complexes through assembly of paralogs in an unsystematic, global manner to alter chromatin architecture seems likely given a few fundamental assumptions based on prior work.

The first is that archaeal histones dimerise randomly, with equal probability of each homodimer and heterodimer formation across pairs. Based on what we know from prior work, this seems a reasonable assumption. Many archaeal histones homo- and hetero-dimerise *in vitro*, but so far, we only have direct evidence for heterodimer formation *in vivo* for *M. fervidus* HMfA and HMfB ((Sandman *et al.* 1994), see Introduction in Chapter 2). *In silico* interrogation of homo- and hetero-dimerisation probability systematically across archaea is hampered by good controls; histone H4, for example, is predicted to form a stable homodimer both in docking experiments and some structure predictions.

Random dimerization and global deposition of histone paralogs is also likely given the lack of dedicated chaperones (Grau-Bové *et al.* 2022). Self-assembly and exchange of paralogs into hypernucleosomes is likely a dynamic process, enabled by the transient opening of oligomeric complexes (Bowerman *et al.* 2021). In this model, the composition of chromatin is directed by mass action, allowing rapid exchange (Stevens and Warnecke 2022). This marks a substantial deviation from what is known about eukaryotes, where many histone variants are integrated in a highly localised and targeted manner.

In the absence of targeted placement by chaperones, one remaining possibility is that paralogs bind preferentially to different DNA sequences. Based on prior work, this also seems unlikely. Archaeal histones appear to bind non-specifically, but preferentially to bendable, DNA (Grayling *et al.* 1997, Bailey *et al.* 2000, Nalabothula *et al.* 2013, Rojec *et al.* 2019). HMfA and HMfB have highly correlated binding footprints in *E. coli* with only slight differences between the two (Rojec *et al.* 2019). Lacking evidence for DNA sequence specificity, it is possible that the DNA sequence associated with the histone complex could alter the relative

hypernucleosome stability, thus exposing some sites to more frequent opening events and facilitating dynamic exchange of variants at specific sites. Some variants may be more localised through increased expression at certain time points and subsequent incorporation at more exposed sites.

Whether the incorporation of histone variants is truly random will require extensive testing *in vivo* using tools such as ChIP-seq. Determining the location of histones on the genomes of archaea is confounded by difficulties in generating specific antibodies to histones paralogs. As archaeal histones are short, generating tags is also difficult without disrupting DNA binding, but this may eventually work (Stevens and Warnecke 2022).

If histone variants in archaea affect chromatin on a global rather than local scale, changes to the relative expression of each in different conditions may help elucidate their function. Building systematically on the differential expression data for diverse archaea (see Table 2.1) will give a comprehensive overview and allow comparison between species.

11.2.3. The interplay between histone variants and other NAPs

The work presented here suggests that complex combinatorial histone-based chromatin may exist in archaea. But this is not the only possible protein-based mechanism able to modify chromatin state. In bacteria, changes to the ensemble of diverse NAPs can alter the architecture of chromatin in a cell cycle-dependent manner (Dorman 2013) (see section 2.2). Other chromatin-associated proteins in archaea likely play a similar role.

Variants of NAPs other than histones may regulate chromatin architecture alongside histone variants. Alba is one such example. In some species, such as *Sulfolobus solfataricus*, there are no histones, but two copies of Alba (Alba1 and Alba2) which have been recently duplicated (Jelinska *et al.* 2005, Laurens *et al.* 2012). Alba2 only functions as a heterodimer with distinct properties, such as DNA binding, from the Alba1 homodimer (Jelinska *et al.* 2005, Laurens *et al.* 2012). Slight changes to the ratio of each may provide an alternative mechanism to alter genome accessibility and genome organisation (Jelinska *et al.* 2005). Aside from the diversity of NAPs themselves, the interplay between histones and NAPs is also important to consider (Stevens and Warnecke 2022). Histones and the NAP TrmBL2 in *T. kodakarensis* form heterogeneous populations of chromatin states – with beads on a string histone architecture,

and a thick fibrous structure containing TrmBL2 (Maruyama *et al.* 2011). *In vitro* reconstitutions of DNA with the NAP Alba and either histones or HTa (from *T. kodakarensis* and *Thermococcus acidophilum*, respectively) resulted in differing chromatin structures suggesting co-operation between different NAPs can fine tune the chromatin architecture (Maruyama *et al.* 2020). These studies show the potential for interaction between NAPs *in vivo* to alter chromatin state. Whether histone variants promote interaction with certain NAPs, and how they influence higher-order multi-complex structures, is a key avenue for further work.

This is particularly relevant for the prediction of chromatinisation in *M. stadtmanae* in different growth stages, where only histone-based components were considered. The abundancies of NAPs, Alba in particular, with differing DNA affinities (or which may outcompete histones for access to the genome) may also change over different growth conditions, and therefore add an additional layer of complexity to what is likely to occur *in vivo*. Whether *M. stadtmanae* cells do indeed exhibit these dynamic changes to chromatin in different growth phases will need to be determined.

The presence of histone variants with different structural properties and shared paralogy provides a tantalising hint at the diversity of histone-based chromatin in archaea. Further exploration of the questions above may shed light on this largely undiscovered field.

12. Methods

12.1.*In silico* modelling

12.1.1. Single mutation analysis of HMfB

The structure of HMfB as a hexamer complex with DNA was used as a starting structure to generate single amino acid mutations *in silico* using the FoldX suite (PDB: 5T5K, (Mattiroli *et al.* 2017)). The structure was relaxed using FoldX command RepairPDB, and all possible single amino acid mutations in each monomer were built using the FoldX command BuildModel (Schymkowitz *et al.* 2005). Each of the six monomers were mutated simultaneously. The relative change in Gibbs free energy (ΔΔG) upon mutation compared to the wild type energy-minimised structure was calculated for the system as a whole (stability – from the BuildModel output), as well as DNA binding and energies between dimers using FoldX (command=AnalyseComplex) (eq 1.). Monomers with chain IDs C and D form a tetramer interface with both dimers A, B, and E, F. The interface between chains A, B and E, F is indicative of oligomerisation i.e. between stacked dimers (Mattiroli *et al.* 2017, Henneman *et al.* 2018).

$$\Delta\Delta G = \Delta G_{model} - \Delta G_{HMfB} \quad \text{eq.1}$$

12.1.2. Generating tetrameric models of archaeal histone-DNA complexes (fast-mutational scanning)

To generate models of archaeal histone-DNA complexes for different species, the hexameric crystal structure of *M. fervidus* histone HMfB and DNA was used as a starting point (PDB 5t5k, (Mattiroli *et al.* 2017)). The concept for this method was developed by Jacob Swadling. One histone dimer and corresponding 30 bp DNA was removed from the hexamer to create a tetrameric complex. The resulting structure was repaired using FoldX (command=RepairPDB) to fix residues with high energy, VanderWaals clashes, or bad torsion angles (Schymkowitz *et al.* 2005). A list of substitutions needed to produce each histone sequence, from an alignment (See section 12.2.3), by mutating HMfB was produced. Substitutions at conserved positions in the core histone fold (positions 2-65 relative to HMfB) were modelled onto the HMfB tetramer using FoldX BuildModel (Schymkowitz *et al.* 2005). For each species, all combinations of

histone monomers encoded by that species were modelled as a tetramer. In this approach, each histone monomer was modelled onto each of the four monomer positions in the HMfB structure. For tetrameric structure of an archaea with 2 histones (A and B) the following models were created: AAAA, AABB, BBAA, ABAB, BABA, BBBB etc (see Figure 5.2).

So that fair comparisons could be drawn between models, those with deletions relative to the HFD were removed. Additionally, the tetramer structure was used as this allows both DNA binding and tetramerization to be probed without the assumption that histones assemble into longer oligomers (Henneman *et al.* 2018).

FoldX only minimises the side chains of mutated and neighbouring residues. To further relax potential high energy clashes which may not occur *in vivo*, each structure was energy minimised using AmberTools for 10,000 steps combined steepest-descent and conjugate gradient (Case *et al.* 2016). To avoid refolding, an all-atom minimisation was performed using the Amber forcefield ff14SB with a 2 kcal/mol/Å^2 harmonic restraint on backbone atoms (Maier *et al.* 2015).

To calculate the resulting DNA binding energies and interaction strength between dimers of the system, Gibbs free energy was used. Gibbs free energy (ΔG) defines that amount of reversible work a system can undergo. This is defined as the enthalpic contribution (ΔH) minus the entropic contribution (ΔS) at a given temperature (T) (eq 2.).

$$\Delta G = \Delta H - T\Delta S \qquad \text{eq.2}$$

The resulting change in Gibbs free energy (ΔG) of DNA binding and tetramerization strength for each model was calculated using the single trajectory MMPBSA method (Miller *et al.* 2012). The enthalpic contribution, ΔH, is decomposed into the gas phase energy and free energy of solvation (eq 3.). The AMBER ff14SB forcefield energy was used to calculate the gas phase energy (Maier *et al.* 2015) and Poisson-Boltzmann equations were used to approximate the free energy of solvation.

$$\Delta G = \left(E_{gas} + E_{solv}\right) - T\Delta S \qquad \text{eq.3}$$

Entropic calculations can be computationally expensive, and therefore difficult to do iterating over the high number of models in this dataset. Using normal node analysis to determine entropic contribution can also have a large margin of error. In this set-up, ΔΔS between models will be close to zero, as single mutations are unlikely to dramatically change the conformation of the histone in the complex. Indeed, structures of histone-DNA complexes are remarkably similar across species despite low sequence identity – e.g. crystal structures of *M. fervidus* histone HMfA (PDB 1hta (Decanniere *et al.* 2000)) and *Pyrococcus horikoshii OT3* histone HPhA (PDB 1ku5 (Li *et al.* 2003)) align with RMSD of 0.64 despite only 62% identity between the two (Berman *et al.* 2000, Li *et al.* 2003) – and between eukaryotes and archaea (Mattiroli *et al.* 2017). Therefore, the entropic contribution to ΔΔG was not included in the calculation. However, it is possible that, in some rare cases, a single residue mutation may alter the conformation of the histone enough as to significantly change the entropic contribution. This process likely would occur over longer timescales than probed here.

For models from all archaeal species, ΔΔG was calculated relative to the *M. fervidus* HMfB homotetramer (e.g., see eq 1.). When just considering *M. stadtmanae* models, ΔΔG is shown relative to the Msp_0769 homotetramer.

12.1.3. Generating models of *M. fervidus* histones HMfA/B as homodimeric and heterodimeric hexamer complexes, or as forced dimer tetramer complexes

The fast mutational scanning approach was used to create homology models of HMfA, HMB and HMfAB based on the HMfB hexamer structure. The HMfAB model has three dimers of AB in the same orientation (AB AB AB). To generate linker models, the linker sequence used by Marc *et al*. 2002 was added between pairs of monomers which dimerise in the structure ("ATAPDRRG" for HMfA, "ATAPDRRG" for HMfAB, "TAPDRRGA" for HMfB). One dimer (EF) was removed along with 30 bp DNA. The side chain residues were energy minimised using FoldX (Schymkowitz *et al.* 2005). The structures were energy minimised using AmberTools, and DNA binding strengths relative to HMfB were calculated (see section 12.1.2).

12.1.4. Random generation of *M. stadtmanae* histone models in exponential and stationary phase

The mean expression level for each *M. stadtmanae* histone in exponential and stationary phase was used as a weight for random sampling. 100,000 tetramers were sampled using the weighting to generate a set of models based on abundancy in each growth phase.

12.1.5. Molecular dynamics simulations

The Amber ff14SB (Maier *et al.* 2015) potentials were used to parameterise the histone DNA complexes in AmberTools16 tLeap (Case *et al.* 2016). 14Å TIP3P water was added to solvate the models and the overall negative charge arising from the DNA backbone was charge neutralised using NaCl (~0.18M). Each model was subsequently energy minimised for 10000 steps of combined steepest descent and conjugate gradient. The models were then heated to 300K over 20 ps using classical molecular dynamics (cMD) in the NVT ensemble. A Langevin thermostat (Davidchack *et al.* 2009) was used to regulate temperature. After heating, 100 ns of cMD in the isobaric/isothermal (NPT) ensemble was run. Pressure was maintained at a constant 1 atm using the Berendsen barostat (Berendsen *et al.* 1984). The Particle-Mesh Ewald summation was used to treat long-range electrostatic forces (Essmann *et al.* 1995). The GPU (CUDA) version 18.0.0 of PMEMD (Götz *et al.* 2012, Le Grand *et al.* 2013, Salomon-Ferrer *et al.* 2013) was used.

For *M. stadtmanae* and other species with capstones, tetrameric complex models built using the fast-mutational scanning approach from section 12.1.2 were used as initial starting structures. To create *M. stadtmanae* 'full models', any small insertions/substitutions/deletions not modelled in the tetrameric complex were added manually before energy minimising and running the simulation, as above. Images were rendered using VMD (Humphrey *et al.* 1996). Calculation of distance between dimers and RMSD from the starting structure was done using CPPTRAJ (Roe and Cheatham 2013). DNA binding and tetramerisation energies were calculated using the MMPBSA method when the models had stabilised (from frames 1500-5000) (Miller *et al.* 2012).

12.1.6. AlphaFold structure predictions

Structure predictions were performed using AlphaFold v2.0 (Jumper et al. 2021) collab notebook (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb) with jackhmmer as the MSA method and ranking models by PTMscore (Mirdita *et al.* 2022). The AlphaFold notebook was accessed on 18-08-21 for *T. kodakarensis* histones HTkA, HTkB, the bacteria-type singlet protein TK1040, and on 18-01-2022 for the prediction of heterodimer complexes containing HMfB and Msp_0383. Amber relax was used for each prediction, with five models per run. Models were generated for all homodimers and all combinations of histones as heterodimers. For each combination of histone proteins (i.e., HTkA+HTkA or HTkA+ TK1040) the top scoring structure is shown. Images were rendered using ChimeraX (Pettersen *et al.* 2021) and VMD (Humphrey *et al.* 1996). UCSF ChimeraX, was developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

12.2.Phylogenetic and sequence-based analysis

12.2.1. Alignment of protein sequences

Protein sequences were aligned using MAFFT-linsi (--localpair –maxiterate 1000) (Katoh and Standley 2013). Sequence alignments were visualised using Jalview (Waterhouse *et al.* 2009).

12.2.2. Phylogenetic tree building and visualisation of phylogenetic trees

Unless otherwise stated, RAxML-NG (Kozlov *et al.* 2019) was used to build 20 maximum likelihood phylogenetic trees with the model of evolution determined by ModelTest-NG (Darriba *et al.* 2020).

Figures containing phylogenetic trees were plotted using iTOL (Letunic and Bork 2019). Unless stated, the unrooted best maximum likelihood tree is shown. GeneSpy was used to plot the location of genes expressing orthologous proteins in the syntenic region of each histone for Thermococcales species (Garcia *et al.* 2019).

12.2.3. Archaeal and eukaryotic histone sequence diversity and similarity

A curated set of histones from a wide range of archaea was given to us by Simonetta Gribaldo (Adam *et al.* 2017). Sequences were filtered to include only those sequences with length between 60-80 amino acids, resulting in a final dataset consisting of 282 species.

Eukaryotic histone sequences were obtained by searching InterPro using the folds IPR007125 (Histone H2A/H2B/H3) and IPR035425 (CENP-T/Histone H4, histone fold) and IPR032454 (Histone H2A, C-terminal domain) (Mitchell *et al.* 2019) on 08-10-18. The resulting accession numbers were used to search Uniprot (Bateman *et al.* 2021) to obtain sequences and taxonomic information. Hits to bacteria and archaea were removed, and H3 and H4 sequences separated. H4 sequences were filtered to remove those under 95, or above 110 amino acids, H3 sequences over 145 or below 130 amino acids were also removed. At this stage, redundant sequences (those which were 100% identical to another sequence) were randomly removed.

Eukaryotic H3 and H4 sequences were added to the alignment of archaeal histones using MAFFT-linsi (--localpair –maxiterate 1000 —seed) (Katoh and Standley 2013). For the comparative analysis of eukaryotic and archaeal histones, positions where over 5% of the sequences had a gap were removed. This focused the analysis on conserved positions in the core histone fold.

At each residue (column) in the alignment, the Shannon diversity index (*H*) (Shannon and Weaver 1964) was used to calculate diversity. Calculations were performed across groups (archaea with one histone, archaea with multiple histones, eukaryotic H3, and eukaryotic H4). To compare relative diversity at residues across groups, Shannon diversity ratio was calculated for two groups (A and B) (eq.4).

$$Shannon\ diversity\ ratio = \frac{Shannon\ diversity\ index\ for\ A}{Shannon\ diversity\ index\ for\ B} \quad \text{eq.4}$$

To calculate similarity between groups (A and B), the Jaccard index formula was used (Jaccard 1908).

12.2.4. Recombination detection

The DNA sequence for genes encoding histones from Methanobrevibacter and Methanobacterium species were aligned. GARD (Kosakovsky Pond *et al.* 2006), PHI (Bruen *et al.* 2006) and RDP4 (Martin *et al.* 2015) were used to test for recombination events within these sequences. Standard settings were used for both PHI and GARD, with 100 bp window size and 1000 permutations for the PHI test. For analysis using RDP4, default settings (except for those highlighted below) were used with the following methods; RDP (window size 30bp), GENECONV, Chimaera (60 variable sites per window), MaxChi (70 variable sites per window), 3Seq, BootScan and SiScan. Primary scans were included for SiScan and BootScan. If $p < 0.05$, sequences were considered as having potential recombination events.

12.2.5. Identification of single and double HFD proteins across archaea

The taxid 2157 was used to download genomes, proteomes and GFF files for a balanced species set of archaea (Hocher *et al.* 2021) from GenBank (https://www.ncbi.nlm.nih.gov/assembly) on 21-05-2021. For genomes where there was no protein set available from GenBank, Prodigal v2.6.3 (Hyatt *et al.* 2010) was used to predict protein sequences from CDS sequences. Protein sequence datasets were searched using hmmsearch from HMMER v3.3.1 (hmmsearch --noali) (Eddy 2011, Finn *et al.* 2011) to identify histone proteins. Queries included the models CBFD_NFYB_HMF and DUF1931 from Pfam (Finn *et al.* 2014). To obtain bacteria-type singlet and doublet histones, jackhmmer searches were performed using the bacteria histones as a starting seed (Alva and Lupas 2019). The steps above were done by Antoine Hocher. Any proteins mis-identified as histones were manually removed. Protein sequences were aligned. The CDS sequence was mapped onto the protein alignment using PAL2NAL (Suyama *et al.* 2006). A large CDS-level phylogenetic tree of archaeal histones was built, and used to identify singlet and doublet histones, along with length information (see section 12.2.2, the best fit model of evolution was F81+G4). For the doublet histones, this included HMk from *Methanopyrus kandleri* (Slesarev *et al.* 1998, Fahrner *et al.* 2001) and the doublets from Halobacteria (Dulmage *et al.* 2015). Clades containing the doublet histone are shown on a tree of Archaea (Borrel *et al.* 2020). This was edited to include Hodarchaeota for Figure 2.6 (Liu *et al.* 2021).

To plot the distribution of singlet and doublet histones in archaea on the GTDB cladogram, the GTDB taxonomic tree was downloaded (Rinke *et al.* 2021, Parks *et al.* 2022) (https://gtdb.ecogenomic.org, accessed 18-02-2022). The balanced set of histones was supplemented with additional Asgard species. GFF files, genomes and protein sets for the taxid 1935183 were downloaded from GenBank on 15-06-21, and from primary sources (Liu *et al.* 2021). These were processed as above. The dataset was then subsetted by those present in GTDB, and organised by Class. The number of single HFD and double HFD proteins (identified by the phylogenetic tree) was calculated for each species. The GTDB tree was pruned to Class level, and is shown as a cladogram rooted at the split between DPANN and the Asgard/TACK/Euryarchaeota. A count of the number of histones for species within each Class was calculated.

12.2.6. Phylogenetic analysis of histones in Methanobacteriales

This work was done by Antoine Hocher. A large initial tree of histones was built from histones identified in 282 species used for tetrameric model construction and structural analysis and which were present in NCBI using hmmsearch (HMMer suite, http://hmmer.org) (Eddy 2011, Finn *et al.* 2011). Single domain hits to the model CBFD_NFYD_HMF from Pfam were obtained (Finn *et al.* 2014). Protein sequences were aligned and IQ-TREE2 was used to build a tree using automatic substitution model estimation, LG+R6 substitution model and with 1000 ultra-fast bootstraps (Minh *et al.* 2020). A subtree containing all *M. stadtmanae* histones was then considered including additional Methanosphaera genomes from the NCBI refseq database. These sequences were re-aligned and trees were built (see section 10.2.2, LG substitution model as determined by ModelTest-NG (Darriba *et al.* 2020) with 500 non-parametric bootstraps).

To build a species tree for the Methanobacteriales, hits to IF2a were identified for each species using the HMM model TIGR00491 and aligned. RAxML-NG was used to build the species tree with the model of evolution (LG substitution model) determined by ModelTest-NG (Darriba *et al.* 2020) and 200 bootstraps. This tree is in agreement with previously identified relationships in this order (Tokura *et al.* 2000).

12.2.7. Identification of histones in Thermococcales

To obtain a balanced set of Thermococcales histones, all genomes, GFF files and protein sets for Thermococci species were downloaded from GenBank (https://www.ncbi.nlm.nih.gov/assembly) searching NCBI with the taxid 183968 (accessed 27-05-2021). To balance this dataset, the GTDB species tree for Thermococci (Parks *et al.* 2022) (https://gtdb.ecogenomic.org, accessed 01-08-2021) was used to subset the dataset. Any genomes not represented in the GTDB dataset were removed. Two species branched far from the main set of Thermococci species using the GTDB tree but were annotated as Thermococci in NCBI taxonomy. These were not included in the analysis, leaving a total of 61 genomes. All remaining species were members of the Thermococcales.

Histone proteins were obtained for each species (processing datasets and HMM searches as described in section 12.2.5). The initial search for bacteria-type singlets hit some Thermococcales proteins with a high e-value. To determine whether these were bacteria-type singlets, we re-searched the dataset using an HMM model built from the initial set of Thermococcales bacteria-type singlets. This returned further bacteria-type singlet histones. At this point, some proteins were manually removed as they were incorrectly hit. *Pyrococcus yayanosii* histone AEH23914.1 had an incorrect ORF at the N-terminus, which was removed (sequence removed: MGAYRGVERPQNPKGLYISFALKENDKTFAGVVE).

12.2.8. Classification of Thermococcales histones into HTkA-like, HTkB-like, and bacteria-type singlets/doublets

Bacteria-type singlets and doublets were identified from HMM searches (see above), however there was some overlap between models (i.e., HMM models for different histone types, e.g., bacteria-type singlet and CBFD_NFYB_HMF, occasionally hit the same protein). Therefore, a protein-level tree of all Thermococcales histones was built (see section 12.2.2, LG+G4 model of evolution as determined by ModelTest-NG (Darriba *et al.* 2020)). Bacteria-type singlets and doublets separated clearly from one another, and from HMf-like histones (consisting of a singlet histone fold domain, and longer than 60 amino acids) on this tree. Bacteria-type histones could additionally be identified by length.

To identify HTkA/B-like histones, Thermococcales histones found to be HMf-like were aligned. To determine whether these histones could be separated into groups, the amino acid properties of each protein sequence was calculated using AAStats in Seqinr (Charif and Lobry 2007), and these properties were used to cluster the histone sequences using principal component analysis (PCA). Two main clusters were found, one containing HTkA and the other HTkB. Histones which clustered with either *T. kodakarensis* histone were labelled as HTkA- or HTkB-like. Phylogenetic trees were also used to separate the HTkA/B histones into groups (see section 12.2.2, LG+G4 model of evolution as determined by ModelTest-NG (Darriba *et al.* 2020)). Sequence logos were plotted in R using the package ggseqlogo showing bit score (Wagih 2017).

The distribution of each histone group was plotted onto the GTDB species tree.

12.2.9. Pairwise percentage sequence identity for HTkA/B-like histones and bacteria-type histones in Thermococcales

Thermococcales histones were separated into HTkA-like, HTkB-like, bacteria-type singlet, and bacteria-type doublet. Histones from species with either the bacteria-type singlet or bacteria-type doublet were aligned separately by group (i.e., HTkA-like, HTkB-like and bacteria-type singlets were aligned separately for all species with bacteria-type singlets). Pairwise sequence identity was calculated for each group using seqidentity from the R package bio3d (Grant *et al.* 2006).

12.2.10. Identification of HTkA- and HTkB-like histones in Thermococcales outgroup species, and comparison to *M. fervidus* histones HMfA/B

To determine whether HTkA- and HTkB-like histones are also present in other species, I downloaded datasets for species from the closest lineages; Theionarchaea, NM3 and Methanofastidiosales (Borrel *et al.* 2020) and processed as above (see section 10.2.5). Assemblies GCA_004212155.1 and GCA_011372835.1 from within the order Methanofastidiosales in GTDB were annotated as NM3 (Borrel *et al.* 2020).

The histones from outgroup species, and HMfA and HMfB protein sequences from *M. fervidus*, were aligned, separately, with HTkA/B-like histones from Thermococcales spp..

Trees were rebuilt (with the LG+G4 model of evolution) and PCA analysis performed as above (see section 12.2.7) with all sequences.

12.2.11. Identification of orthologs and genome order conservation in Thermococcales

For each species in the Thermococcales, orthologs were identified using BLAST to find reciprocal best hits for all genomes against *T. kodakarensis* (Altschul *et al.* 1990). Proteins were considered orthologs if they were within 20% length of one another and had a similarity score of >40% (Rocha 2006). The set of orthologs was used to plot histones in the genomic neighbourhood of histones. Genome order conservation (GOC) was calculated pairwise across genomes as in (Rocha 2006) (eq.5).

$$GOC = \frac{number\ of\ contiguous\ orthologs}{number\ of\ orthologs} \qquad \text{eq.5}$$

12.2.12. Expression level of histones

For comparison across species, exponential phase histone expression levels for *T. kodakarensis*, *T. onnurineus* and *M. smithii* were taken from primary publications and NCBI's Gene Expression Omnibus (GEO) (Barrett *et al.* 2013). The median value for *M. smithii* strain MsmPS across all replicates and conditions is shown (Hansen *et al.* 2011). *T. kodakarensis* is plotted as base mean (Jäger *et al.* 2014), and *T. onnurineus* as normalised mRNA (GSE85760 (Cho *et al.* 2017)). Expression levels for other species were obtained from (Müller *et al.* 2020) with the emission of *Methanocaldococcus jannaschii* and *Pyrococcus furiosus* which have high sequence identity across histones resulting in non-unique mapping. The expression level for *M. stadtmanae* histones in stationary and exponential phase were obtained from qRT-PCR and proteomics data gathered by Ruth Schmitz and Corinna Bang (Stevens *et al.* 2020).

Expression data for archaea in different environments were obtained from searching NCBI's Gene Expression Omnibus (GEO) (Barrett *et al.* 2013) and from primary publications (Boonyaratanakornkit *et al.* 2005, 2006, 2007, Cho *et al.* 2017, Liu and Lu 2018, Topçuoğlu *et al.* 2019, Sas-Chen *et al.* 2020). For Table 2.1, where histone levels were not given as $\log_2$(fold change) in expression between conditions, this was calculated.

For comparison of histone abundancy from transcriptomic and proteomic data in *T. kodakarensis*, proteome data from (Sas-Chen *et al.* 2020) was processed as in (Hocher *et al.* 2021), and abundance of proteins at 85ºC is shown. Transcriptomic data was downloaded from (Jäger *et al.* 2014) and was plotted as transcripts per million (TPM).

12.3.*In vitro* analysis of *M. stadtmanae* histones and heterologous expression in *E. coli*

12.3.1. Generation of *E. coli* strains expressing archaeal histones from *M. stadtmanae*

Rhamnose-inducible vector plasmids were designed to contain the histones from *M. stadtmanae* and a kanamycin resistance gene. Coding sequences for the two histones were codon-optimised for expression in bacteria by ATUM (Newark, CA). *E. coli* K12 MG1655 cells were transformed with the plasmid using heat shock. This work was done previously by Maria Rojec (in 2015).

E. coli BL21(DE3) cells (New England Biolabs, NEB) were also transformed with the pD861 plasmids containing Msp_0383 and Msp_0769, as described above.

E. coli K21 strains expressing Msp_0383 and Msp_0769 are further referred to as K21-msp0383 and K12-msp0769, respectively. *E. coli* BL21(DE3) strains expressing Msp_0383 and Msp_0769 are further referred to as BL21-msp0383 and BL21-msp0769, respectively.

Strains containing histones Msp_0769 and Msp_0383 were PCR verified using Phusion® High-Fidelity PCR Master Mix with HF buffer (Thermo Scientific) and specific primer pairs for each histone (Table 12.1). Primers were obtained from IDT DNA, Iowa.

Table 12.1 Primer sequences for *M. stadtmanae* recombinant histones.

Primer pair	**Sequence (5' -> 3')**	**Length (bp)**	**Tm**	**GC content**
Msp_0383 Forward	TCTGAATTGCCACTGACCC	19	51.5	52.6
Msp_0383 Reverse	CAAAAAACCCCTCAAGACCC	20	51.0	50.0
Msp_0769 Forward	ATTCGTCCACCACAATTCAG	20	50.3	45.0
Msp_0769 Reverse	ATCTTCGCATCTTCGCTAAC	20	50.4	45.0

The expected product length for Msp_0383 forward and reverse is 286 bp, and for Msp_0769 237 bp. Samples were run on 2% agarose 1 x Tris/Borate/EDTA (TBE) gel at 100 V for 45 minutes with a GeneRuler LowRange ladder (Thermo Scientific).

12.3.2. Bacterial growth

All strains were grown in LB medium with 50 ug/µL kanamycin at 180 rpm at 37ºC, with the exception of those used for expression tests, some of which had additional components (see section 12.3.3). When cells reached an optical density of 0.4 - 0.6, L-Rhamnose was added to induce histone expression. Cells were harvested by centrifugation after 4 hours of growth (at late exponential phase).

12.3.3. Expression tests

To determine the optimum conditions for expression of recombinant histones, expression tests were carried out for strains K12-msp0383, K12-msp0769, BL21-msp0383 and BL21-msp0769. K12 strains expressing histones were grown as previously described (section 12.3.2) at varying rhamnose concentrations (1 mM, 5 mM), supplemented with, or without, sorbitol (0 or 0.5 M sorbitol) and at varying temperatures (4 h at 37ºC, or 16-17 hours at room temperature). BL12(DE3) strains expressing histones were grown as previously described at different concentrations of L-rhamnose (1 mM, 5 mM).

12.3.4. Purification of archaeal histones

To obtain recombinant Msp_0383 and Msp_0769 from *E. coli*, protocols to purify histones from mesophilic archaea (Darcy *et al.* 1995, Li *et al.* 1998, Topping and Gloss 2004) and recombinant histones from eukaryotes (Klinker *et al.* 2014) were adapted.

12.3.4.1. Purification of recombinant *M. stadtmanae* histone Msp_0383 (capstone)

K21-msp0383 was grown as previously described (section 12.3.2) and expression of the histone was induced by addition of 5 mM Rhamnose. Cells harvested from late exponential phase were pelleted and stored at -80ºC until use (maximum 1 week).

Cell pellets were lysed and treated with DNaseI as previously described (Topping and Gloss 2004) by sonication at room temperature in 10 mL 50 mM Tris-HCl pH 7.5, 2 mM NaPi, 0.1 M NaCl, 0.1 mM PMSF. The cell suspension was brought to 5 mM $MgCl_2$, 0.1 mM $MnCl_2$. 20 µL DNase I (2000U/mL, NEB) and 20 µ L 20 mg/mL PureLink™ RNase A (Invitrogen) was added, and lysate was incubated at 37ºC 500 rpm for 1 h.

After clarification, the supernatant was brought to 70% ammonium sulphate at room temperature for 10 minutes with gentle stirring, then centrifuged at 4500 rcf for 15 min at room temperature (Topping and Gloss 2004, Duong-Ly and Gabelli 2014). The supernatant was dialysed into 1 L 20 mM NaCl, 25 mM Tris-HCl pH 8 for 1 h, then moved to fresh 1 L 20 mM NaCl, 25 mM Tris-HCl pH 8 overnight.

A HiTrap® Heparin High Performance 5 mL column (Cytiva, Sigma Aldrich) was equilibrated with 20 mM NaCl, 25 mM Tris-HCl pH 8. Dialysed supernatant was flowed through the column. Flow through was loaded into both anion (HiTrap Q HP, Cytiva, Sigma Aldrich)/cation (HiTrap SP HP, Cytiva, Sigma Aldrich) exchange columns joined together (with anion above cation) and equilibrated with 20 mM NaCl, 25 mM Tris-HCl pH 8 (Klinker *et al.* 2014). The flow through and fractions containing a protein of the correct molecular weight were collected, and dialysed into 10% glycerol, 0.5 M KCl, 25 mM Tris pH 7.5 for 1 h, then overnight. Purified recombinant Msp_0383 was stored at -80ºC.

12.3.4.2. Purification of recombinant *M. stadtmanae* histone Msp_0769 (non-capstone)

Msp_0769 was purified as Msp_0383 (see section 12.3.4.1) with the following exceptions.

Expression of Msp_0769 from the strain K12-msp0769 was induced with 1 mM Rhamnose. Cells were lysed on ice. After bringing the solution to 70% ammonium sulphate, dialysed samples were loaded onto a HiTrap® Heparin High Performance 5 mL column. A gradient of 20 mM NaCl to 2 M NaCl at 25 mM Tris-HCl pH 8 was run through the column. Fractions containing the histone were pooled and brought to ~ 20 mM NaCl and flowed through anion/cation exchange columns (as in section 12.3.4.1, but with cation above anion). The flow through, containing the highest concentration of the histone, was collected. As the volume of flow through was high, a vivaspin 3kDa device (Sartorius) was used to concentrate the sample before dialysing into 10% glycerol, 0.5 M KCl, 25 mM Tris pH 7.5. Purified proteins were stored at -80^{o}C.

12.3.4.3. Re-purification of *M. fervidus* histone HMfB

HMfB was previously purified by Maria in 2015 (Rojec 2020) using the protocol from (Starich *et al.* 1996). However, some contaminants were found in the stock solution. The protein stock was therefore re-purified. The previously purified HMfB stock solution was diluted to 0.25 M KCl, 25 mM Tris pH 7.5 and run through a HiTrap® Heparin High Performance 5 mL column with a gradient of 20 mM NaCl to 2 M KCl at 25 mM Tris pH 7.5. All proteins eluted together. Fractions containing HMfB were pooled and loaded onto both anion/cation exchange columns (see 12.3.4.2). A gradient was run as above. The flow through was found to contain the highest concentration of histone, and was collected. A vivaspin 3kDa was used to concentrate the flow through, and the purified protein was dialysed into 10% glycerol, 0.5 M KCl, 25 mM Tris pH 7.5. Purified proteins were stored at -80^{o}C.

12.3.5. Staining of protein gels with Coomassie blue

During protein purification steps, precast 4-20% Tris-Glycine TGX gels (BioRad Laboratories, California) were imaged before and after staining with colloidal Coomassie (Instant Blue, Sigma-Aldrich). This step was to add an additional layer of security that the low molecular weight band was the histone of interest. Under UV light, trihalo compounds in the pre-stained

4-20% TGX gels react with tryptophan residues in the proteins and produce fluorescence. As neither Msp_0383 nor Msp_0769 contain tryptophan in the protein sequence, a band would not be expected in the pre-stained gels. Coomassie blue staining after UV imaging then reveals all proteins, including those whose sequence does not contain tryptophan. Comparing these two gels, any bands which appear at the correct molecular weight in the post-stain but not pre-stain gel could be the histone proteins of interest.

12.3.6. Measuring protein concentration

For all purified proteins, protein concentration was estimated using absorbance at 205 nm. Absorbance at 205 nm (protein backbone) was used for each as HMfB and Msp_0769 have predicted 0 M^{-1} cm^{-1} molar absorptivity at 280 nm. The extinction co-efficient for each protein was calculated using Protein Parameter Calculator (Anthis and Clore 2013), (http://nickanthis.com/tools/a205.html accessed 04-08-2020).

12.3.7. Electrophoretic Mobility Shift Assay (EMSA)

Previous work has used gel shift assays with histone and linearised plasmid DNA to show differences in DNA binding (Sandman *et al.* 1994, Grayling *et al.* 1996, Higashibata *et al.* 1999, Marc *et al.* 2002). The EMSAs performed here were based on this previous work. Plasmid pBR322 (Thermo Scientific) was linearised using restriction enzyme HindIII-HF (NEB Biolabs). The concentration of DNA was measured using a nanodrop. Reactions were made up to 100 mM KCl, 50 mM Tris pH8. TMAO and KGlu were added at various concentrations if required. Reagents were added in the following order: water, Tris pH 8, salt, DNA, protein, and gently mixed before incubating at 25ºC for 25 minutes. Gel Loading Dye, Purple (6X), no SDS (NEB) was added.

Samples were run on a 1 % agarose, 1 x TAE gel at 4ºC for ~17 hours at 1 V/cm (20 V). Gels were stained using ethidium bromide (10 μL / 200 mL 1x TAE buffer) for 15 minutes and washed with water before imaging.

12.3.8. In vitro MNase digestion

In vitro MNase assays were performed as in (Mattiroli *et al.* 2017) with the following modifications.

Histones were incubated with 230 bp DNA derived from the 601-positioning sequence (Lowary and Widom 1998) with 100mM KCl.

To create renatured dimers, purified proteins were denatured and renatured. To denature the proteins, each was brought to 6 M urea. To renature, proteins were dialysed for ~16 hours into 10% glycerol, 0.5 M KCl, 25 mM Tris pH 7.5, plus an additional 2 h into fresh 10% glycerol, 0.5 M KCl, 25 mM Tris pH 7.5. Dialysing steps were performed at 4°C with gentle shaking. Renatured dimers were generated by denaturing and renaturing HMfB and Msp_0383 together. Mixed dimers were generated by incubating HMfB and Msp_0383 after separate renaturation.

After incubation, samples were digested for either 5 or 10 minutes with 10 U MNase (Thermofisher) at 37°C, 300 rpm. Samples were run on 3 % agarose 1 x TBE gel at 120 V for 35 minutes with 6 x purple loading dye.

12.3.9. MNase digestion of *E. coli* chromatin from cells expressing histones from *M. fervidus* or *M. stadtmanae*

MNase digestion of *E. coli* strains expressing archaeal histones was performed as previously described for both fixed (Rojec *et al.* 2019) and unfixed (Hocher *et al.* 2019) cells. For unfixed cells, mechanical lysis was done with grinding under liquid nitrogen.

DNA was purified using phenol chloroform extraction with incubation on ice at 4°C for 2 h to precipitate the DNA. Samples were run on 2.5 % agarose 1 x TBE gel at 100 V for 45 minutes, or run on a TapeStation (Agilent, CA).

12.3.10. Novobiocin sensitivity assay

Novobiocin sensitivity assays for *E. coli* strains expressing *M. stadtmanae* histones and HMf were performed as in (Rojec *et al.* 2019). In brief, overnight cultures were subcultured into 5 mL LB + 50 ug/mL kanamycin for 2 hours before expression was induced with 15 mM

Rhamnose for an additional 2 hours. 10 μL of each histone-expressing strain was added to wells of a Nunc96 plate containing 200 μL LB + 50 ug/mL kanamycin either with or without 10 ug/mL novobiocin (3 replicates per strain, per condition). Growth was measured every 30 minutes by monitoring absorbance at 600 nm at 37^{o}C over 15 hours.

13. References

Adam, P.S., Borrel, G., Brochier-Armanet, C., and Gribaldo, S., 2017. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *The ISME Journal*, 11 (11), 2407–2425.

Agalioti, T., Chen, G., and Thanos, D., 2002. Deciphering the transcriptional histone acetylation code for a human gene. *Cell*, 111 (3), 381–392.

Ahmad, K. and Henikoff, S., 2002. The Histone Variant H3.3 Marks Active Chromatin by Replication-Independent Nucleosome Assembly. *Molecular Cell*, 9, 1191–1200.

Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S., and Ishihama, A., 1999. Growth Phase-Dependent Variation in Protein Composition of the Escherichia coli Nucleoid. *Journal of Bacteriology*, 181 (20), 6361–6370.

Alpha-Bazin, B., Gorlas, A., Lagorce, A., Joulié, D., Boyer, J.-B., Dutertre, M., Gaillard, J.-C., Lopes, A., Zivanovic, Y., Dedieu, A., Confalonieri, F., and Armengaud, J., 2021. Lysine-specific acetylated proteome from the archaeon Thermococcus gammatolerans reveals the presence of acetylated histones. *Journal of Proteomics*, 232, 104044.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J., 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215 (3), 403–410.

Alva, V. and Lupas, A.N., 2019. Histones Predate the Split Between Bacteria and Archaea. *Bioinformatics*, 35 (14), 2349–2353.

Amatori, S., Tavolaro, S., Gambardella, S., and Fanelli, M., 2021. The dark side of histones: genomic organization and role of oncohistones in cancer. *Clinical Epigenetics*, 13 (1), 1–21.

Ammar, R., Torti, D., Tsui, K., Gebbia, M., Durbic, T., Bader, G.D., Giaever, G., and Nislow, C., 2012. Chromatin is an ancient innovation conserved between Archaea and Eukarya. *eLife*, 2012 (1).

Anthis, N.J. and Clore, G.M., 2013. Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Science*, 22 (6), 851–858.

Arents, G., Burlingame, R.W., Wang, B.C., Love, W.E., and Moudrianakis, E.N., 1991. The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *Proceedings of the National Academy of Sciences*, 88 (22), 10148–10152.

Arents, G. and Moudrianakis, E.N., 1995. The histone fold: A ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proceedings of the National*

Academy of Sciences of the United States of America, 92 (24), 11170–11174.

Armache, A., Yang, S., Martínez de Paz, A., Robbins, L.E., Durmaz, C., Cheong, J.Q., Ravishankar, A., Daman, A.W., Ahimovic, D.J., Klevorn, T., Yue, Y., Arslan, T., Lin, S., Panchenko, T., Hrit, J., Wang, M., Thudium, S., Garcia, B.A., Korb, E., Armache, K.J., Rothbart, S.B., Hake, S.B., Allis, C.D., Li, H., and Josefowicz, S.Z., 2020. Histone H3.3 phosphorylation amplifies stimulation-induced transcription. *Nature*, 583 (7818), 852–857.

Bailey, K.A., Chow, C.S., and Reeve, J.N., 1999. Histone stoichiometry and DNA circularization in archaeal nucleosomes. *Nucleic Acids Research*, 27 (2), 532–536.

Bailey, K.A., Marc, F., Sandman, K., and Reeve, J.N., 2002. Both DNA and histone fold sequences contribute to archaeal nucleosome stability. *Journal of Biological Chemistry*, 277 (11), 9293–9301.

Bailey, K.A., Pereira, S.L., Widom, J., and Reeve, J.N., 2000. Archaeal histone selection of nucleosome positioning sequences and the procaryotic origin of histone-dependent genome evolution. *Journal of Molecular Biology*, 303, 25–34.

Bang, C., Schilhabel, A., Weidenbach, K., Kopp, A., Goldmann, T., Gutsmann, T., and Schmitz, R.A., 2012. Effects of antimicrobial peptides on methanogenic archaea. *Antimicrobial Agents and Chemotherapy*, 56 (8), 4123–4130.

Bannister, A.J. and Kouzarides, T., 2011. Regulation of chromatin by histone modifications. *Cell Research*, 21 (3), 381–395.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S., and Soboleva, A., 2013. NCBI GEO: Archive for functional genomics data sets - Update. *Nucleic Acids Research*, 41 (D1), 991–995.

Baskakov, I. and Bolen, D.W., 1998. Forcing thermodynamically unfolded proteins to fold. *Journal of Biological Chemistry*, 273 (9), 4831–4834.

Bateman, A., Martin, M.J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Silva, A. Da, Denny, P., Dogan, T., Ebenezer, T.G., Fan, J., Castro, L.G., Garmiri, P., Georghiou, G., Gonzales, L., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Jokinen, P., Joshi, V., Jyothi, D., Lock, A., Lopez, R., Luciani, A., Luo, J., Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Menchi, M., Mishra, A., Moulang, K., Nightingale, A., Oliveira, C.S., Pundir, S., Qi, G., Raj, S., Rice, D., Lopez, M.R., Saidi, R., Sampson, J., Sawford, T., Speretta, E., Turner, E., Tyagi, N., Vasudev,

P., Volynkin, V., Warner, K., Watkins, X., Zaru, R., Zellner, H., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.C., Bolleman, J., Boutet, E., Breuza, L., Casals-Casas, C., de Castro, E., Echioukh, K.C., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Estreicher, A., Famiglietti, M.L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Keller, G., Kerhornou, A., Lara, V., Le Mercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T.B., Paesano, S., Pedruzzi, I., Pilbout, S., Pourcel, L., Pozzato, M., Pruess, M., Rivoire, C., Sigrist, C., Sonesson, K., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Ross, K., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.S., and Zhang, J., 2021. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49 (D1), D480–D489.

Becker, P.B. and Hörz, W., 2002. Atp-dependent nucleosome remodeling. *Annual Review of Biochemistry*, 71, 247–273.

Berendsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F., Dinola, A., and Haak, J.R., 1984. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81 (8), 3684–3690.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E., 2000. The Protein Data Bank. *Nucleic Acids Research*, 28 (1), 235–242.

Betts, H.C., Puttick, M.N., Clark, J.W., Williams, T.A., Donoghue, P.C.J., and Pisani, D., 2018. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nature Ecology and Evolution*, 2 (10), 1556–1562.

Bheda, P., Swatkoski, S., Fiedler, K.L., Boeke, J.D., Cotter, R.J., and Wolberger, C., 2012. Biotinylation of lysine method identifies acetylated histone H3 lysine 79 in Saccharomyces cerevisiae as a substrate for Sir2. *Proceedings of the National Academy of Sciences*, 109 (16), E916–E925.

Bjellqvist, B., Basse, B., Olsen, E., and Celis, J.E., 1994. Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis*, 15 (1), 529–539.

Blais Lecours, P., Marsolais, D., Cormier, Y., Berberi, M., Haché, C., Bourdages, R., and Duchaine, C., 2014. Increased Prevalence of Methanosphaera stadtmanae in

Inflammatory Bowel Diseases. *PLoS ONE*, 9 (2), e87734.

Boonyaratanakornkit, B., Córdova, J., Park, C.B., and Clark, D.S., 2006. Pressure affects transcription profiles of Methanocaldococcus jannaschii despite the absence of barophilic growth under gas-transfer limitation. *Environmental Microbiology*, 8 (11), 2031–2035.

Boonyaratanakornkit, B.B., Miao, L.Y., and Clark, D.S., 2007. Transcriptional responses of the deep-sea hyperthermophile Methanocaldococcus jannaschii under shifting extremes of temperature and pressure. *Extremophiles*, 11 (3), 495–503.

Boonyaratanakornkit, B.B., Simpson, A.J., Whitehead, T.A., Fraser, C.M., El-Sayed, N.M.A., and Clark, D.S., 2005. Transcriptional profiling of the hyperthermophilic methanarchaeon Methanococcus jannaschii in response to lethal heat and non-lethal cold shock. *Environmental Microbiology*, 7 (6), 789–797.

Borrel, G., Brugère, J.-F., Gribaldo, S., Schmitz, R.A., and Moissl-Eichinger, C., 2020. The host-associated archaeome. *Nature Reviews Microbiology*, 18 (11), 622–636.

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

Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P., 2008. Mesophilic crenarchaeota: Proposal for a third archaeal phylum, the Thaumarchaeota. *Nature Reviews Microbiology*, 6 (3), 245–252.

Brochier-Armanet, C., Forterre, P., and Gribaldo, S., 2011. Phylogeny and evolution of the Archaea: One hundred genomes later. *Current Opinion in Microbiology*, 14 (3), 274–281.

Bruen, T.C., Philippe, H., and Bryant, D., 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics*, 172 (4), 2665–2681.

Brunk, C.F. and Martin, W.F., 2019. Archaeal Histone Contributions to the Origin of Eukaryotes. *Trends in Microbiology*, 27 (8), 703–714.

Buschbeck, M. and Hake, S.B., 2017. Variants of core histones and their roles in cell fate decisions, development and cancer. *Nature Reviews Molecular Cell Biology*, 18 (5), 299–314.

Cajili, M.K.M. and Prieto, E.I., 2022. Interplay between Alba and Cren7 Regulates Chromatin Compaction in Sulfolobus solfataricus. *Biomolecules*, 12 (4), 481.

Case, D.A., Betz, R.M., Cerutti, D.S., Cheatham III, T.E., Darden, T.A., Duke, R.E., Giese, T.J., Gohlke, H., Goetz, A.W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J.,

Kovalenko, A., Lee, T.S., LeGrand, S., Li, P., Lin, C., Luchko, T., Luo, R., Madej, B., Mermelstein, D., Merz, K.M., Monard, G., Nguyen, H., Nguyen, H.T., Omelyan, I., Onufriev, A., Roe, D.R., Roitberg, A., Sagui, C., Simmerling, C.L., Botello-Smith, W.M., Swails, J., Walker, R.C., Wang, J., Wolf, R.M., Wu, X., Xiao, L., and Kollman, P.A., 2016. AMBER 2016. *University of California, San Francisco*.

Di Cerbo, V., Mohn, F., Ryan, D.P., Montellier, E., Kacem, S., Tropberger, P., Kallis, E., Holzner, M., Hoerner, L., Feldmann, A., Richter, F.M., Bannister, A.J., Mittler, G., Michaelis, J., Khochbin, S., Feil, R., Schuebeler, D., Owen-Hughes, T., Daujat, S., and Schneider, R., 2014. Acetylation of histone H3 at lysine 64 regulates nucleosome dynamics and facilitates transcription. *eLife*, 3 (3), e01632.

Chakravarthy, S. and Luger, K., 2006. The histone variant macro-H2A preferentially forms 'hybrid nucleosomes'. *Journal of Biological Chemistry*, 281 (35), 25522–25531.

Charif, D. and Lobry, J.R., 2007. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. *In*: U. Bastolla, M. Porto, H.E. Roman, and M. Vendruscolo, eds. *Structural approaches to sequence evolution: Molecules, networks, populations*. New York: Springer Verlag, 207–232.

Cho, S., Kim, M.-S., Jeong, Y., Lee, B.-R., Lee, J.-H., Kang, S.G., and Cho, B.-K., 2017. Genome-wide primary transcriptome analysis of H2-producing archaeon Thermococcus onnurineus NA1. *Scientific Reports*, 7 (1), 43044.

Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L., 2017. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology*, 18 (7), 407–422.

Cockram, C., Thierry, A., Gorlas, A., Lestini, R., and Koszul, R., 2021. Euryarchaeal genomes are folded into SMC-dependent loops and domains, but lack transcription-mediated compartmentalization. *Molecular Cell*, 81 (3), 459-472.e10.

Corujo, D. and Buschbeck, M., 2018. Post-Translational Modifications of H2A Histone Variants and Their Role in Cancer. *Cancers*, 10 (3), 59.

Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J., and Meyer, B.J., 2015. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature*, 523 (7559), 240–244.

Čuboňová, L., Katano, M., Kanai, T., Atomi, H., Reeve, J.N., and Santangelo, T.J., 2012. An archaeal histone is required for transformation of Thermococcus kodakarensis. *Journal of Bacteriology*, 194 (24), 6864–6874.

Dai, J., Hyland, E.M., Yuan, D.S., Huang, H., Bader, J.S., and Boeke, J.D., 2008. Probing nucleosome function: A highly versatile library of synthetic histone H3 and H4 mutants. *Cell*, 134 (6), 1066–1078.

Dalal, Y., Wang, H., Lindsay, S., and Henikoff, S., 2007. Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells. *PLoS Biology*, 5 (8), 1798–1809.

Darcy, T.J., Sandman, K., and Reeve, J.N., 1995. Methanobacterium formicicum, a mesophilic methanogen, contains three HFo histones. *Journal of Bacteriology*, 177 (3), 858–860.

Darling, A.E., Mau, B., and Perna, N.T., 2010. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE*, 5 (6), e11147.

Darriba, Di., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B., and Flouri, T., 2020. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular Biology and Evolution*, 37 (1), 291–294.

Davidchack, R.L., Handel, R., and Tretyakov, M. V., 2009. Langevin thermostat for rigid body dynamics. *The Journal of Chemical Physics*, 130 (23), 234101.

DeBry, R.W., 1998. Comparative analysis of evolution in a rodent histone H2a pseudogene. *Journal of Molecular Evolution*, 46 (3), 355–360.

Decanniere, K., Babu, A.M., Sandman, K., Reeve, J.N., and Heinemann, U., 2000. Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon Methanothermus fervidus. *Journal of Molecular Biology*, 303 (1), 35–47.

Dillon, S.C. and Dorman, C.J., 2010. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology*, 8 (3), 185–195.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485 (7398), 376–380.

Dong, P., Tu, X., Chu, P.Y., Lü, P., Zhu, N., Grierson, D., Du, B., Li, P., and Zhong, S., 2017. 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. *Molecular Plant*, 10 (12), 1497–1509.

Dorman, C.J., 2013. Genome architecture and global gene regulation in bacteria: Making progress towards a unified model? *Nature Reviews Microbiology*, 11 (5), 349–355.

Draizen, E.J., Shaytan, A.K., Mariño-Ramírez, L., Talbert, P.B., Landsman, D., and Panchenko, A.R., 2016. HistoneDB 2.0: a histone database with variants—an integrated resource to explore histones and their variants. *Database*, 2016, baw014.

Driessen, R.P.C. and Dame, R.T., 2011. Nucleoid-associated proteins in Crenarchaea.

Biochemical Society Transactions, 39 (1), 116–121.

Drinnenberg, I.A., deYoung, D., Henikoff, S., and Malik, H.S. ing., 2014. Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *eLife*, 3, 1–19.

Dulmage, K.A., 2015. Large-Scale Effectors of Gene Expression and New Models of Cell Division in the Haloarchaea. Duke University.

Dulmage, K.A., Todor, H., and Schmid, A.K., 2015. Growth-phase-specific modulation of cell morphology and gene expression by an archaeal histone protein. *mBio*, 6 (5), e00649-15.

Duong-Ly, K.C. and Gabelli, S.B., 2014. *Salting out of proteins using ammonium sulfate precipitation*. 1st ed. Methods in Enzymology. Elsevier Inc.

Eddy, S.R., 2011. Accelerated Profile HMM Searches. *PLoS Computational Biology*, 7 (10), e1002195.

Eickbush, T.H. and Moudrianakis, E.N., 1978. The histone core complex: an octamer assembled by two sets of protein-protein interactions. *Biochemistry*, 17 (23), 4955–4964.

Elkins, J.G., Podar, M., Graham, D.E., Makarova, K.S., Wolf, Y., Randau, L., Hedlund, B.P., Brochier-Armanet, C., Kunin, V., Anderson, I., Lapidus, A., Goltsman, E., Barry, K., Koonin, E. V., Hugenholtz, P., Kyrpides, N., Wanner, G., Richardson, P., Keller, M., and Stetter, K.O., 2008. A korarchaeal genome reveals insights into the evolution of the Archaea. *Proceedings of the National Academy of Sciences of the United States of America*, 105 (23), 8102–8107.

Eme, L., Sharpe, S.C., Brown, M.W., and Roger, A.J., 2014. On the Age of Eukaryotes: Evaluating Evidence from Fossils and Molecular Clocks. *Cold Spring Harbor Perspectives in Biology*, 6 (8), 1–16.

Eme, L., Spang, A., Lombard, J., Stairs, C.W., and Ettema, T.J.G., 2017. Archaea and the origin of eukaryotes. *Nature Reviews Microbiology*, 15 (12), 711–723.

Esin, A., Bergendahl, L.T., Savolainen, V., Marsh, J.A., and Warnecke, T., 2018. The genetic basis and evolution of red blood cell sickling in deer. *Nature Ecology and Evolution*, 2 (2), 367–376.

Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G., 1995. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, 103 (19), 8577–8593.

Fahrner, R.L., Cascio, D., Lake, J.A., and Slesarev, A., 2001. An ancestral nuclear protein assembly: Crystal structure of the Methanopyrus kandleri histone. *Protein Science*, 10,

2002–2007.

Feng, S., Cokus, S.J., Schubert, V., Zhai, J., Pellegrini, M., and Jacobsen, S.E., 2014. Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis. *Molecular Cell*, 55 (5), 694–707.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., and Punta, M., 2014. Pfam: The protein families database. *Nucleic Acids Research*, 42 (D1), 222–230.

Finn, R.D., Clements, J., and Eddy, S.R., 2011. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research*, 39 (SUPPL. 2), 29–37.

Forbes, A.J., Patrie, S.M., Taylor, G.K., Kim, Y. Bin, Jiang, L., and Kelleher, N.L., 2004. Targeted analysis and discovery of posttranslational modifications in proteins from methanogenic archaea by top-down MS. *Proceedings of the National Academy of Sciences of the United States of America*, 101 (9), 2678–2683.

Fricke, W.F., Seedorf, H., Henne, A., Krüer, M., Liesegang, H., Hedderich, R., Gottschalk, G., and Thauer, R.K., 2006. The genome sequence of Methanosphaera stadtmanae reveals why this human intestinal archaeon is restricted to methanol and H2 for methane formation and ATP synthesis. *Journal of Bacteriology*, 188 (2), 642–658.

Friedrich-Jahn, U., Aigner, J., Längst, G., Reeve, J.N., and Huber, H., 2009. Nanoarchaeal origin of histone H3? *Journal of Bacteriology*, 191 (3), 1092–1096.

Galtier, N., 2003. Gene conversion drives GC content evolution in mammalian histones, 19 (2), 65–68.

Garcia, B.A., Hake, S.B., Diaz, R.L., Kauer, M., Morris, S.A., Recht, J., Shabanowitz, J., Mishra, N., Strahl, B.D., Allis, C.D., and Hunt, D.F., 2007. Organismal differences in post-translational modifications in histones H3 and H4. *Journal of Biological Chemistry*, 282 (10), 7641–7655.

Garcia, P.S., Jauffrit, F., Grangeasse, C., and Brochier-Armanet, C., 2019. GeneSpy, a user-friendly and flexible genomic context visualizer. *Bioinformatics*, 35 (2), 329–331.

Gates, L.A., Shi, J., Rohira, A.D., Feng, Q., Zhu, B., Bedford, M.T., Sagum, C.A., Jung, S.Y., Qin, J., Tsai, M.J., Tsai, S.Y., Li, W., Foulds, C.E., and O'Malley, B.W., 2017. Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. *Journal of Biological Chemistry*, 292 (35), 14456–14472.

Giaimo, B.D., Ferrante, F., Herchenröther, A., Hake, S.B., and Borggrefe, T., 2019. The histone variant H2A.Z in gene regulation. *Epigenetics and Chromatin*, 12 (1), 1–22.

Gornik, S.G., Ford, K.L., Mulhern, T.D., Bacic, A., McFadden, G.I., and Waller, R.F., 2012.

Loss of nucleosomal DNA condensation coincides with appearance of a novel nuclear protein in dinoflagellates. *Current Biology*, 22 (24), 2303–2312.

Götz, A.W., Williamson, M.J., Xu, D., Poole, D., Le Grand, S., and Walker, R.C., 2012. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. generalized born. *Journal of Chemical Theory and Computation*, 8 (5), 1542–1555.

Le Grand, S., Götz, A.W., and Walker, R.C., 2013. SPFP: Speed without compromise - A mixed precision model for GPU accelerated molecular dynamics simulations. *Computer Physics Communications*, 184 (2), 374–380.

Grant, B.J., Rodrigues, A.P.C., ElSawy, K.M., McCammon, J.A., and Caves, L.S.D., 2006. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*, 22 (21), 2695–2696.

Grau-Bové, X., Navarrete, C., Chiva, C., Pribasnig, T., Antó, M., Torruella, G., Galindo, L.J., Lang, B.F., Moreira, D., López-Garcia, P., Ruiz-Trillo, I., Schleper, C., Sabidó, E., and Sebé-Pedrós, A., 2022. A phylogenetic and proteomic reconstruction of eukaryotic chromatin evolution. *Nature Ecology and Evolution*, 6 (7), 1007–1023.

Grayling, R.A., Bailey, K.A., and Reeve, J.N., 1997. DNA binding and nuclease protection by the HMf histones from the hyperthermophilic archaeon Methanothermus fervidus. *Extremophiles*, 1 (2), 79–88.

Grayling, R.A., Sandman, K., and Reeve, J.N., 1996. Histones and chromatin structure in hyperthermophilic Archaea. *FEMS Microbiology Reviews*, 18 (2–3), 203–213.

Guy, L. and Ettema, T.J.G., 2011. The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends in Microbiology*, 19 (12), 580–587.

Han, W., Li, X., and Fu, X., 2011. The macro domain protein family: Structure, functions, and their potential therapeutic implications. *Mutation Research - Reviews in Mutation Research*, 727 (3), 86–103.

Hansen, E.E., Lozupone, C.A., Rey, F.E., Wu, M., Guruge, J.L., Narra, A., Goodfellow, J., Zaneveld, J.R., McDonald, D.T., Goodrich, J.A., Heath, A.C., Knight, R., and Gordon, J.I., 2011. Pan-genome of the dominant human gut-associated archaeon, Methanobrevibacter smithii , studied in twins. *Proceedings of the National Academy of Sciences*, 108 (supplement_1), 4599–4606.

Heinicke, I., Müller, J., Pittelkow, M., and Klein, A., 2004. Mutational analysis of genes encoding chromatin proteins in the archaeon Methanococcus voltae indicates their involvement in the regulation of gene expression. *Molecular Genetics and Genomics*, 272 (1), 76–87.

Henikoff, S. and Smith, M.M., 2015. Histone Variants and Epigenetics. *Cold Spring Harbor Perspectives in Biology*, 7 (1), a019364.

Henneman, B., 2019. Histone-DNA assemblies in archaea: shaping the genome on the edge of life. Leiden University.

Henneman, B., Brouwer, T.B., Erkelens, A.M., Kuijntjes, G.J., Van Emmerik, C., Van Der Valk, R.A., Timmer, M., Kirolos, N.C.S., Van Ingen, H., Van Noort, J., and Dame, R.T., 2021. Mechanical and structural properties of archaeal hypernucleosomes. *Nucleic Acids Research*, 49 (8), 4338–4349.

Henneman, B., van Emmerik, C., van Ingen, H., and Dame, R.T., 2018. Structure and function of archaeal histones. *PLoS Genetics*, 14 (9), e1007582.

Higashibata, H., Fujiwara, S., Takagi, M., and Imanaka, T., 1999. Analysis of DNA compaction profile and intracellular contents of archaeal histones from Pyrococcus kodakaraensis KOD1. *Biochemical and Biophysical Research Communications*, 258 (2), 416–424.

Higashibata, H., Siddiqui, M.A., Takagi, M., Imanaka, T., and Fujiwara, S., 2003. Surface histidine residue of archaeal histone affects DNA compaction and thermostability. *FEMS Microbiology Letters*, 224 (1), 17–22.

Hocher, A., Borrel, G., Fadhlaoui, K., Brugère, J.F., Gribaldo, S., and Warnecke, T., 2021. Growth temperature is the principal driver of chromatinization in archaea. *BioRxiv*.

Hocher, A., Rojec, M., Swadling, J.B., Esin, A., and Warnecke, T., 2019. The dna-binding protein hta from thermoplasma acidophilum is an archaeal histone analog. *eLife*, 8, 1–24.

Hörz, W. and Altenburger, W., 1981. Sequence specific cleavage of DNA by micrococcal nuclease. *Nucleic Acids Research*, 9 (12), 2643–2658.

Hou, C., Li, L., Qin, Z.S., and Corces, V.G., 2012. Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains. *Molecular Cell*, 48 (3), 471–484.

Hsieh, T.-H.S., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., and Rando, O.J., 2015. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell*, 162 (1), 108–119.

Humphrey, W., Dalke, A., and Schulten, K., 1996. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14 (1), 33–38.

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J., 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC*

Bioinformatics, 11, 119.

Jaccard, P., 1908. Nouvelles recherches sur la distribution florale: Bulletin de la Societé Vandoise des Sciences Naturelles 44. *La Presse médicale*, 44, 223–270.

Jackson, J.D. and Gorovsky, M.A., 2000. Histone H2A.Z has a conserved function that is distinct from that of the major H2A sequence variants. *Nucleic Acids Research*, 28 (19), 3811–3816.

Jäger, D., Förstner, K.U., Sharma, C.M., Santangelo, T.J., and Reeve, J.N., 2014. Primary transcriptome map of the hyperthermophilic archaeon Thermococcus kodakarensis. *BMC Genomics*, 15 (1), 684.

Jelinska, C., Conroy, M.J., Craven, C.J., Hounslow, A.M., Bullough, P.A., Waltho, J.P., Taylor, G.L., and White, M.F., 2005. Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA Packaging. *Structure*, 13 (7), 963–971.

Jensen, M.W., Matlock, S.A., Reinheimer, C.H., Lawlor, C.J., Reinheimer, T.A., and Gorrell, A., 2015. Potassium stress growth characteristics and energetics in the haloarchaeon Haloarcula marismortui. *Extremophiles*, 19 (2), 315–325.

Jin, C. and Felsenfeld, G., 2007. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes and Development*, 21 (12), 1519–1529.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., and Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596 (August), 583–589.

Kaaij, L.J.T., van der Weide, R.H., Ketting, R.F., and de Wit, E., 2018. Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development. *Cell Reports*, 24 (1), 1-10.e4.

Kamakaka, R.T. and Biggins, S., 2005. Histone variants: Deviants? *Genes and Development*, 19 (3), 295–310.

Kanai, T., Takedomi, S., Fujiwara, S., Atomi, H., and Imanaka, T., 2010. Identification of the Phr-dependent heat shock regulon in the hyperthermophilic archaeon, Thermococcus kodakaraensis. *Journal of Biochemistry*, 147 (3), 361–370.

Katoh, K. and Standley, D.M., 2013. MAFFT multiple sequence alignment software version

7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30 (4), 772–780.

Keese, A.M., Schut, G.J., Ouhammouch, M., Adams, M.W.W., and Thomm, M., 2010. Genome-wide identification of targets for the archaeal heat shock regulator Phr by cell-free transcription of genomic DNA. *Journal of Bacteriology*, 192 (5), 1292–1298.

Khan, A.M., Miotto, O., Nascimento, E.J.M., Srinivasan, K.N., Heiny, A.T., Zhang, G.L., Marques, E.T., Tan, T.W., Brusic, V., Salmon, J., and August, J.T., 2008. Conservation and Variability of Dengue Virus Proteins: Implications for Vaccine Design. *PLoS Neglected Tropical Diseases*, 2 (8), e272.

Klinker, H., Haas, C., Harrer, N., Becker, P.B., and Mueller-Planitz, F., 2014. Rapid Purification of Recombinant Histones. *PLoS ONE*, 9 (8), e104029.

Kornberg, R.D., 1974. Chromatin Structure : A Repeating Unit of Histones and DNA. *Science (New York, N.Y.)*, 184 (4139), 868–871.

Kornberg, R.D., 1977. Structure of chromatin. *Annual review of biochemistry*, 46, 931–954.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., and Frost, S.D.W., 2006. GARD: A genetic algorithm for recombination detection. *Bioinformatics*, 22 (24), 3096–3098.

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A., 2019. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35 (21), 4453–4455.

Kozlowski, M., Corujo, D., Hothorn, M., Guberovic, I., Mandemaker, I.K., Blessing, C., Sporn, J., Gutierrez-Triana, A., Smith, R., Portmann, T., Treier, M., Scheffzek, K., Huet, S., Timinszky, G., Buschbeck, M., and Ladurner, A.G., 2018. MacroH2A histone variants limit chromatin plasticity through two distinct mechanisms. *EMBO reports*, 19 (10), 1–13.

Krassovsky, K., Henikoff, J.G., and Henikoff, S., 2012. Tripartite organization of centromeric chromatin in budding yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 109 (1), 243–248.

Kumar, S.V. and Wigge, P.A., 2010. H2A.Z-Containing Nucleosomes Mediate the Thermosensory Response in Arabidopsis. *Cell*, 140 (1), 136–147.

Lassalle, F., Périan, S., Bataillon, T., Nesme, X., Duret, L., and Daubin, V., 2015. GC-Content Evolution in Bacterial Genomes: The Biased Gene Conversion Hypothesis Expands. *PLOS Genetics*, 11 (2), e1004941.

Laurens, N., Driessen, R.P.C., Heller, I., Vorselen, D., Noom, M.C., Hol, F.J.H., White, M.F.,

Dame, R.T., and Wuite, G.J.L., 2012. Alba shapes the archaeal genome using a delicate balance of bridging and stiffening the DNA. *Nature Communications*, 3 (1), 1328.

Laursen, S.P., Bowerman, S., and Luger, K., 2021. Archaea: The Final Frontier of Chromatin. *Journal of Molecular Biology*, 433 (6), 166791.

Le, T.B.K., Imakaev, M. V., Mirny, L.A., and Laub, M.T., 2013. High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. *Science*, 342 (6159), 731–734.

Lei, B., Capella, M., Montgomery, S.A., Borg, M., Osakabe, A., Goiser, M., Muhammad, A., Braun, S., and Berger, F., 2021. A Synthetic Approach to Reconstruct the Evolutionary and Functional Innovations of the Plant Histone Variant H2A.W. *Current Biology*, 31 (1), 182-191.e5.

Letunic, I. and Bork, P., 2019. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, 47 (W1), 256–259.

Li, B., Carey, M., and Workman, J.L., 2007. The Role of Chromatin during Transcription. *Cell*, 128 (4), 707–719.

Li, T., Sun, F., Ji, X., Feng, Y., and Rao, Z., 2003. Structure based hyperthermostability of archaeal histone HPhA from Pyrococcus horikoshii. *Journal of Molecular Biology*, 325 (5), 1031–1037.

Li, W.T., Grayling, R.A., Sandman, K., Edmondson, S., Shriver, J.W., and Reeve, J.N., 1998. Thermodynamic stability of archaeal histones. *Biochemistry*, 37 (30), 10563–10572.

Li, W.T., Sandman, K., Pereira, S.L., and Reeve, J.N., 2000. MJ1647, an open reading frame in the genome of the hyperthermophile methanococcus jannaschii, encodes a very thermostable archaeal histone with a C-terminal extension. *Extremophiles*, 4 (1), 43–51.

Li, W.T., Shriver, J.W., and Reeve, J.N., 2000. Mutational analysis of differences in thermostability between histones from mesophilic and hyperthermophilic archaea. *Journal of Bacteriology*, 182 (3), 812–817.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., and Dekker, J., 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, 326 (5950), 289–293.

Lioy, V.S., Cournac, A., Marbouty, M., Duigou, S., Mozziconacci, J., Espéli, O., Boccard, F., and Koszul, R., 2018. Multiscale Structuring of the E. coli Chromosome by Nucleoid-Associated and Condensin Proteins. *Cell*, 172 (4), 771-783.e18.

Liu, P. and Lu, Y., 2018. Concerted Metabolic Shifts Give New Insights Into the Syntrophic

Mechanism Between Propionate-Fermenting Pelotomaculum thermopropionicum and Hydrogenotrophic Methanocella conradii. *Frontiers in Microbiology*, 9, 1551.

Liu, Y., Makarova, K.S., Huang, W.-C., Wolf, Y.I., Nikolskaya, A.N., Zhang, X., Cai, M., Zhang, C.-J., Xu, W., Luo, Z., Cheng, L., Koonin, E. V., and Li, M., 2021. Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature*, 593 (7860), 553–557.

Liu, Y. and Whitman, W.B., 2008. Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Annals of the New York Academy of Sciences*, 1125, 171–189.

Long, M., Sun, X., Shi, W., Yanru, A., Leung, S.T.C., Ding, D., Cheema, M.S., MacPherson, N., Nelson, C.J., Ausio, J., Yan, Y., and Ishibashi, T., 2019. A novel histone H4 variant H4G regulates rDNA transcription in breast cancer. *Nucleic acids research*, 47 (16), 8399–8409.

Lowary, P.T. and Widom, J., 1998. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of Molecular Biology*, 276 (1), 19–42.

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

Luijsterburg, M.S., White, M.F., Van Driel, R., and Th. Dame, R., 2008. The major architects of chromatin: Architectural proteins in bacteria, archaea and eukaryotes. *Critical Reviews in Biochemistry and Molecular Biology*, 43 (6), 393–418.

Lukeš, J. and Maslov, D.A., 2000. Unexpectedly high variability of the histone H4 gene in Leishmania. *Parasitology Research*, 86 (3), 259–261.

Luo, Z., Hu, T., Jiang, H., Wang, R., Xu, Q., Zhang, S., Cao, J., and Song, X., 2020. Rearrangement of macronucleus chromosomes correspond to TAD-like structures of micronucleus chromosomes in Tetrahymena thermophila. *Genome Research*, 30 (3), 406–414.

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., and Ravikesavan, R., 2013. Gene duplication as a major force in evolution. *Journal of Genetics*, 92 (1), 155–161.

Mah, L.J., El-Osta, A., and Karagiannis, T.C., 2010. γh2AX: A sensitive molecular marker of DNA damage and repair. *Leukemia*, 24 (4), 679–686.

Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., and Simmerling, C., 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11 (8), 3696–3713.

Malik, H.S. and Henikoff, S., 2003. Phylogenomics of the nucleosome. *Nature Structural*

Biology, 10 (11), 882–891.

Marbouty, M., Le Gall, A., Cattoni, D.I., Cournac, A., Koh, A., Fiche, J.B., Mozziconacci, J., Murray, H., Koszul, R., and Nollmann, M., 2015. Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging. *Molecular Cell*, 59 (4), 588–602.

Marc, F., Sandman, K., Lurz, R., and Reeve, J.N., 2002. Archaeal Histone Tetramerization Determines DNA Affinity and the Direction of DNA Supercoiling. *Journal of Biological Chemistry*, 277 (34), 30879–30886.

Marinov, G.K. and Lynch, M., 2016. Diversity and divergence of dinoflagellate histone proteins. *G3: Genes, Genomes, Genetics*, 6 (2), 397–422.

Martin, D.D., Ciulla, R.A., and Roberts, M.F., 1999. Osmoadaptation in archaea. *Applied and Environmental Microbiology*, 65 (5), 1815–1825.

Martin, D.P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B., 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, 1 (1), 1–5.

Martire, S. and Banaszynski, L.A., 2020. The roles of histone variants in fine-tuning chromatin organization and function. *Nature Reviews Molecular Cell Biology*, 21 (9), 522–541.

Maruyama, H., Harwood, J.C., Moore, K.M., Paszkiewicz, K., Durley, S.C., Fukushima, H., Atomi, H., Takeyasu, K., and Kent, N.A., 2013. An alternative beads-on-a-string chromatin architecture in Thermococcus kodakarensis. *EMBO Reports*, 14 (8), 711–717.

Maruyama, H., Nambu, T., Mashimo, C., Okinaga, T., and Takeyasu, K., 2021. Single-Molecule/Cell Analyses Reveal Principles of Genome-Folding Mechanisms in the Three Domains of Life. *International Journal of Molecular Sciences*, 22 (24), 13432.

Maruyama, H., Prieto, E.I., Nambu, T., Mashimo, C., Kashiwagi, K., Okinaga, T., Atomi, H., and Takeyasu, K., 2020. Different Proteins Mediate Step-Wise Chromosome Architectures in Thermoplasma acidophilum and Pyrobaculum calidifontis. *Frontiers in Microbiology*, 11, 1247.

Maruyama, H., Shin, M., Oda, T., Matsumi, R., Ohniwa, R.L., Itoh, T., Shirahige, K., Imanaka, T., Atomi, H., Yoshimura, S.H., and Takeyasu, K., 2011. Histone and TK0471/TrmBL2 form a novel heterogeneous genome architecture in the hyperthermophilic archaeon Thermococcus kodakarensis. *Molecular Biology of the Cell*, 22 (3), 386–398.

Marzluff, W.F. and Duronio, R.J., 2002. Histone mRNA expression: Multiple levels of cell cycle regulation and important developmental consequences. *Current Opinion in Cell*

Biology, 14 (6), 692–699.

Mattiroli, F., Bhattacharyya, S., Dyer, P.N., White, A.E., Sandman, K., Burkhart, B.W., Byrne, K.R., Lee, T., Ahn, N.G., Santangelo, T.J., Reeve, J.N., and Luger, K., 2017. Structure of histone-based chromatin in Archaea. *Science*, 357 (6351), 609–612.

Mattiroli, F., D'Arcy, S., and Luger, K., 2015. The right place at the right time: chaperoning core histone variants. *EMBO reports*, 16 (11), 1454–1466.

McKnight, S.L. and Miller, O.L., 1976. Ultrastructural patterns of RNA synthesis during early embryogenesis of Drosophila melanogaster. *Cell*, 8 (2), 305–319.

Miller, B.R., McGee, T.D., Swails, J.M., Homeyer, N., Gohlke, H., and Roitberg, A.E., 2012. MMPBSA.py: An efficient program for end-state free energy calculations. *Journal of Chemical Theory and Computation*, 8 (9), 3314–3321.

Miller, T.L. and Wolin, M.J., 1985. Methanosphaera stadtmaniae gen. nov., sp. nov.: a species that forms methane by reducing methanol with hydrogen. *Archives of Microbiology*, 141 (2), 116–122.

Miller, T.L., Wolin, M.J., De Macario, E.C., and Macario, A.J.L., 1982. Isolation of Methanobrevibacter smithii from human feces. *Applied and Environmental Microbiology*, 43 (1), 227–232.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., Lanfear, R., and Teeling, E., 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37 (5), 1530–1534.

Mirdita, M., Schütze, K., Moriwaki, Y., and Heo, L., 2022. ColabFold : making protein folding accessible to all. *Nature Methods*, 19, 679–682.

Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.Y., El-Gebali, S., Fraser, M.I., Gough, J., Haft, D.R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Nuka, G., Orengo, C., Pandurangan, A.P., Paysan-Lafosse, T., Pesseat, S., Potter, S.C., Qureshi, M.A., Rawlings, N.D., Redaschi, N., Richardson, L.J., Rivoire, C., Salazar, G.A., Sangrador-Vegas, A., Sigrist, C.J.A., Sillitoe, I., Sutton, G.G., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Yong, S.Y., and Finn, R.D., 2019. InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research*, 47 (D1), D351–D360.

Mito, Y., Henikoff, J.G., and Henikoff, S., 2005. Genome-scale profiling of histone H3.3 replacement patterns. *Nature Genetics*, 37 (10), 1090–1097.

Mizuguchi, T., Fudenberg, G., Mehta, S., Belton, J.-M., Taneja, N., Folco, H.D., FitzGerald, P., Dekker, J., Mirny, L., Barrowman, J., and Grewal, S.I.S., 2014. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. *Nature*, 516 (7531), 432–435.

Molaro, A. and Drinnenberg, I.A., 2018. Studying the Evolution of Histone Variants Using Phylogeny. *Methods in Molecular Biology*, 1832, 273–291.

Molaro, A., Young, J.M., and Malik, H.S., 2018. Evolutionary origins and diversification of testis-specific short histone H2A variants in mammals. *Genome Research*, 28 (4), 460–473.

Müller, J.B., Geyer, P.E., Colaço, A.R., Treit, P. V., Strauss, M.T., Oroshi, M., Doll, S., Virreira Winter, S., Bader, J.M., Köhler, N., Theis, F., Santos, A., and Mann, M., 2020. The proteome landscape of the kingdoms of life. *Nature*, 582 (7813), 592–596.

Nacev, B.A., Feng, L., Bagert, J.D., Lemiesz, A.E., Gao, J., Soshnev, A.A., Kundra, R., Schultz, N., Muir, T.W., and Allis, C.D., 2019. The expanding landscape of 'oncohistone' mutations in human cancers. *Nature*, 567 (7749), 473–478.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I.S., 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science*, 292 (5514), 110–113.

Nalabothula, N., Xi, L., Bhattacharyya, S., Widom, J., Wang, J.P., Reeve, J.N., Santangelo, T.J., and Fondufe-Mittendorf, Y.N., 2013. Archaeal nucleosome positioning in vivo and in vitro is directed by primary sequence motifs. *BMC Genomics*, 14, 391.

Nasir, A., Mughal, F., and Caetano-Anollés, G., 2021. The tree of life describes a tripartite cellular world. *BioEssays*, 43 (6), 1–11.

Van Noort, J., Verbrugge, S., Goosen, N., Dekker, C., and Dame, R.T., 2004. Dual architectural roles of HU: Formation of flexible hinges and rigid filaments. *Proceedings of the National Academy of Sciences of the United States of America*, 101 (18), 6969–6974.

Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., Chee, G.J., Hattori, M., Kanai, A., Atomi, H., Takai, K., and Takami, H., 2011. Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Research*, 39 (8), 3204–3223.

Nusinow, D.A., Sharp, J.A., Morris, A., Salas, S., Plath, K., and Panning, B., 2007. The Histone Domain of macroH2A1 Contains Several Dispersed Elements that Are Each Sufficient to Direct Enrichment on the Inactive X Chromosome. *Journal of Molecular*

Biology, 371 (1), 11–18.

Van Oss, S.B. and Carvunis, A.R., 2019. De novo gene birth. *PLoS Genetics*, 15 (5), 1–23.

Palmer, D.K., O'Day, K., Trong, H.L.E., Charbonneau, H., and Margolis, R.L., 1991. Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proceedings of the National Academy of Sciences of the United States of America*, 88 (9), 3734–3738.

Palmer, D.K., O'Day, K., Wener, M.H., Andrews, B.S., and Margolis, R.L., 1987. A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *Journal of Cell Biology*, 104 (4), 805–815.

Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., and Hugenholtz, P., 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research*, 50 (D1), D785–D794.

Patwal, I., Trinh, H., Golden, A., and Flaus, A., 2021. Histone sequence variation in divergent eukaryotes facilitates diversity in chromatin packaging. *bioRxiv*.

Pereira, S.L., Grayling, R. a, Lurz, R., and Reeve, J.N., 1997. Archaeal nucleosomes. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 12633–12637.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E., 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, 30 (1), 70–82.

Postberg, J., Forcob, S., Chang, W.-J., and Lipps, H.J., 2010. The evolutionary history of histone H3 suggests a deep eukaryotic root of chromatin modifying mechanisms. *BMC Evolutionary Biology*, 10 (1), 259.

Raman, P., Rominger, M.C., Young, J.M., Molaro, A., Tsukiyama, T., and Malik, H.S., 2022. Novel Classes and Evolutionary Turnover of Histone H2B Variants in the Mammalian Germline. *Molecular Biology and Evolution*, 39 (2), 1–19.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtier, K., Ponting, C.P., Allis, C.D., and Jenuwein, T., 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406 (6796), 593–599.

Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K., and Bonner, W., 2002. Histone H2A variants H2AX and H2AZ. *Current Opinion in Genetics & Development*, 12 (2), 162–169.

Reeve, J.N., Bailey, K.A., Li, W., Marc, F., Sandman, K., and Soares, D.J., 2004. Archaeal histones: structures, stability and DNA binding. *Biochemical Society Transactions*, 32 (2), 227–230.

Rinke, C., Chuvochina, M., Mussig, A.J., Chaumeil, P.A., Davín, A.A., Waite, D.W., Whitman, W.B., Parks, D.H., and Hugenholtz, P., 2021. A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nature Microbiology*, 6 (7), 946–959.

Rivera, C., Gurard-Levin, Z.A., Almouzni, G., and Loyola, A., 2014. Histone lysine methylation and chromatin replication. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1839 (12), 1433–1439.

Rizzo, P.J. and Noodén, L.D., 1972. Chromosomal proteins in the dinoflagellate alga Gyrodinium cohnii. *Science*, 176 (4036), 796–797.

Rocha, E.P.C., 2006. Inference and analysis of the relative stability of bacterial chromosomes. *Molecular Biology and Evolution*, 23 (3), 513–522.

Roe, D.R. and Cheatham, T.E., 2013. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, 9 (7), 3084–3095.

Rojec, M., 2020. The impact of archaeal histones on a naïve prokaryote. Imperial College London.

Rojec, M., Hocher, A., Stevens, K.M., Merkenschlager, M., and Warnecke, T., 2019. Chromatinization of Escherichia coli with archaeal histones. *eLife*, 8, e49038.

Rouvière-Yaniv, J. and Kjeldgaard, N.O., 1979. Native Escherichia coli HU protein is a heterotypic dimer. *FEBS Letters*, 106 (2), 297–300.

Sakrikar, S., Hackley, R.K., Martinez-Pastor, M., Darnell, C.L., Vreugdenhil, A., and Schmid, A.K., 2022. Comparative analysis of genome-wide protein-DNA interactions across domains of life reveals unique binding patterns for hypersaline archaeal histones. *bioRxiv*.

Sakrikar, S. and Schmid, A.K., 2021. An archaeal histone-like protein regulates gene expression in response to salt stress. *Nucleic acids research*, 49 (22), 12732–12743.

Salomon-Ferrer, R., Götz, A.W., Poole, D., Le Grand, S., and Walker, R.C., 2013. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. *Journal of Chemical Theory and Computation*, 9 (9), 3878–3888.

Sanders, T.J., Lammers, M., Marshall, C.J., Walker, J.E., Lynch, E.R., and Santangelo, T.J., 2019. TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin.

Molecular Microbiology, 111 (3), 784–797.

Sanders, T.J., Marshall, C.J., and Santangelo, T.J., 2019. The Role of Archaeal Chromatin in Transcription. *Journal of Molecular Biology*, 431 (20), 4103–4115.

Sanders, T.J., Ullah, F., Gehring, A.M., Burkhart, B.W., Vickerman, R.L., Fernando, S., Gardner, A.F., Ben-Hur, A., and Santangelo, T.J., 2021. Extended Archaeal Histone-Based Chromatin Structure Regulates Global Gene Expression in Thermococcus kodakarensis. *Frontiers in Microbiology*, 12, 1071.

Sandman, K., Grayling, R.A., Dobrinski, B., Lurz, R., and Reeve, J.N., 1994. Growth-phase-dependent synthesis of histones in the archaeon Methanothermus fervidus. *Proceedings of the National Academy of Sciences*, 91 (26), 12624–12628.

Sandman, K., Krzycki, J.A., Dobrinski, B., Lurz, R., and Reeve, J.N., 1990. HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon Methanothermus fervidus, is most closely related to histones. *Proceedings of the National Academy of Sciences*, 87 (15), 5788–5791.

Sandman, K., Louvel, H., Samson, R.Y., Pereira, S.L., and Reeve, J.N., 2008. Archaeal chromatin proteins histone HMtB and Alba have lost DNA-binding ability in laboratory strains of Methanothermobacter thermautotrophicus. *Extremophiles*, 12 (6), 811–817.

Sandman, K., Soares, D., and Reeve, J.N., 2001. Molecular components of the archaeal nucleosome. *Biochimie*, 83 (2), 277–281.

Sapp, J. and Fox, G.E., 2013. The Singular Quest for a Universal Tree of Life. *Microbiology and Molecular Biology Reviews*, 77 (4), 541–550.

Sas-Chen, A., Thomas, J.M., Matzov, D., Taoka, M., Nance, K.D., Nir, R., Bryson, K.M., Shachar, R., Liman, G.L.S., Burkhart, B.W., Gamage, S.T., Nobe, Y., Briney, C.A., Levy, M.J., Fuchs, R.T., Robb, G.B., Hartmann, J., Sharma, S., Lin, Q., Florens, L., Washburn, M.P., Isobe, T., Santangelo, T.J., Shalev-Benami, M., Meier, J.L., and Schwartz, S., 2020. Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. *Nature*, 583 (7817), 638–643.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L., 2005. The FoldX web server: An online force field. *Nucleic Acids Research*, 33 (Web Server issue), W382–W388.

Scienski, K., Fay, J.C., and Conant, G.C., 2015. Patterns of gene conversion in duplicated yeast histones suggest strong selection on a coadapted macromolecular complex. *Genome Biology and Evolution*, 7 (12), 3249–3258.

Scully, R. and Xie, A., 2013. Double strand break repair functions of histone H2AX.

Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 750, 5–14.

Seitz, K.W., Lazar, C.S., Hinrichs, K.U., Teske, A.P., and Baker, B.J., 2016. Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. *ISME Journal*, 10 (7), 1696–1705.

Sela, I., Wolf, Y.I., and Koonin, E. V., 2016. Theory of prokaryotic genome evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 113 (41), 11399–11407.

Senaratne, A.P., Cortes-Silva, N., and Drinnenberg, I.A., 2022. Evolution of holocentric chromosomes: Drivers, diversity, and deterrents. *Seminars in Cell & Developmental Biology*.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G., 2012. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, 148 (3), 458–472.

Shannon, C.E., 1948. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27 (3), 379–423.

Shannon, C.E. and Weaver, W., 1964. *The Mathematical Theory of Communication*. The University of Illinois Press. Urbana, IL: The University of Illinois Press.

Shin, J., Santangelo, T.J., Xie, Y., Reeve, J.N., and Kelman, Z., 2007. Archaeal Minichromosome Maintenance (MCM) Helicase Can Unwind DNA Bound by Archaeal Histones and Transcription Factors. *Journal of Biological Chemistry*, 282 (7), 4908–4915.

Sing, S., Chu, D., and Roy, S., 2022. Nematode histone H2A variant evolution reveals diverse histories of retention and loss, and evidence for conserved core-like variants. *BioRxiv*.

Skoko, D., Yoo, D., Bai, H., Schnurr, B., Yan, J., McLeod, S.M., Marko, J.F., and Johnson, R.C., 2006. Mechanism of Chromosome Compaction and Looping by the Escherichia coli Nucleoid Protein Fis. *Journal of Molecular Biology*, 364 (4), 777–798.

Slesarev, A.I., Belova, G.I., Kozyavkin, S.A., and Lake, J.A., 1998. Evidence for an early prokaryotic origin of histones H2A and H4 prior to the emergence of eukaryotes. *Nucleic Acids Research*, 26 (2), 427–430.

Soares, D., Dahlke, I., Li, W.T., Sandman, K., Hethke, C., Thomm, M., and Reeve, J.N., 1998. Archaeal histone stability, DNA binding, and transcription inhibition above 90°C. *Extremophiles*, 2 (2), 75–81.

Soares, D.J., Marc, F., and Reeve, J.N., 2003. Conserved eukaryotic histone-fold residues substituted into an archaeal histone increase DNA affinity but reduce complex flexibility. *Journal of bacteriology*, 185 (11), 3453–7.

Soares, D.J., Sandman, K., and Reeve, J.N., 2000. Mutational analysis of archaeal histone-DNA interactions. *Journal of Molecular Biology*, 297 (1), 39–47.

Soo, V.W.C. and Warnecke, T., 2021. Slaying the last unicorn: discovery of histones in the microalga Nanochlorum eucaryotum. *Royal Society Open Science*, 8 (2), rsos.202023.

Spaans, S.K., van der Oost, J., and Kengen, S.W.M., 2015. The chromosome copy number of the hyperthermophilic archaeon Thermococcus kodakarensis KOD1. *Extremophiles*, 19 (4), 741–750.

Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., Van Eijk, R., Schleper, C., Guy, L., and Ettema, T.J.G., 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, 521 (7551), 173–179.

Spellerberg, I.F. and Fedor, P.J., 2003. A tribute to Claude Shannon (1916-2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon-Wiener' Index. *Global Ecology and Biogeography*, 12 (3), 177–179.

Starich, M.R., Sandman, K., Reeve, J.N., and Summers, M.F., 1996. NMR structure of HMfB from the hyperthermophile, Methanothermus fervidus, confirms that this archaeal protein is a histone. *Journal of Molecular Biology*, 255 (1), 187–203.

Stetter, K.O., Thomm, M., Winter, J., Wildgruber, G., Huber, H., Zillig, W., Jané-Covic, D., König, H., Palm, P., and Wunderl, S., 1981. Methanothermus fervidus, sp. nov., a novel extremely thermophilic methanogen isolated from an Icelandic hot spring. *Zentralblatt für Bakteriologie Mikrobiologie und Hygiene: I. Abt. Originale C: Allgemeine, angewandte und ökologische Mikrobiologie*, 2 (2), 166–178.

Stevens, K.M., Hocher, A., and Warnecke, T., 2022. Deep Conservation of Histone Variants in Thermococcales Archaea. *Genome Biology and Evolution*, 14 (1), 1–9.

Stevens, K.M., Swadling, J.B., Hocher, A., Bang, C., Gribaldo, S., Schmitz, R.A., and Warnecke, T., 2020. Histone variants in archaea and the evolution of combinatorial chromatin complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 117 (52), 33384–33395.

Stevens, K.M. and Warnecke, T., 2022. Histone variants in archaea – An undiscovered country. *Seminars in Cell & Developmental Biology*.

Sun, F. and Huang, L., 2016. A euryarchaeal histone modulates strand displacement synthesis by replicative DNA polymerases. *Science China Life Sciences*, 59 (7), 709–716.

Suyama, M., Torrents, D., and Bork, P., 2006. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, 34 (WEB. SERV. ISS.), 609–612.

Taboada, B., Estrada, K., Ciria, R., and Merino, E., 2018. Operon-mapper: A web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics*, 34 (23), 4118–4120.

Takahashi, K., Murakami, S., Chikashige, Y., Funabiki, H., Niwa, O., and Yanagida, M., 1992. A low copy number central sequence with strict symmetry and unusual chromatin structure in fission yeast centromere. *Molecular Biology of the Cell*, 3 (7), 819–835.

Takayanagi, S., Morimura, S., Kusaoke, H., Yokoyama, Y., Kano, K., and Shioda, M., 1992. Chromosomal structure of the halophilic archaebacterium Halobacterium salinarium. *Journal of Bacteriology*, 174 (22), 7207–7216.

Takemata, N. and Bell, S.D., 2020. Emerging views of genome organization in Archaea. *Journal of Cell Science*, 133 (10), 1–7.

Takemata, N. and Bell, S.D., 2021. Multi-scale architecture of archaeal chromosomes. *Molecular Cell*, 81 (3), 473–487.

Takemata, N., Samson, R.Y., and Bell, S.D., 2019. Physical and Functional Compartmentalization of Archaeal Chromosomes. *Cell*, 179 (1), 165–179.

Talbert, P.B., Ahmad, K., Almouzni, G., Ausió, J., Berger, F., Bhalla, P.L., Bonner, W.M., Cande, W.Z., Chadwick, B.P., Chan, S.W.L., Cross, G.A.M., Cui, L., Dimitrov, S.I., Doenecke, D., Eirin-López, J.M., Gorovsky, M.A., Hake, S.B., Hamkalo, B.A., Holec, S., Jacobsen, S.E., Kamieniarz, K., Khochbin, S., Ladurner, A.G., Landsman, D., Latham, J.A., Loppin, B., Malik, H.S., Marzluff, W.F., Pehrson, J.R., Postberg, J., Schneider, R., Singh, M.B., Smith, M.M., Thompson, E., Torres-Padilla, M.-E., Tremethick, D.J., Turner, B.M., Waterborg, J.H., Wollmann, H., Yelagandula, R., Zhu, B., and Henikoff, S., 2012. A unified phylogeny-based nomenclature for histone variants. *Epigenetics & Chromatin*, 5 (1), 7.

Talbert, P.B. and Henikoff, S., 2010. Histone variants - ancient wrap artists of the epigenome. *Nature Reviews Molecular Cell Biology*, 11 (4), 264–275.

Talbert, P.B. and Henikoff, S., 2014. Environmental responses mediated by histone variants. *Trends in Cell Biology*, 24 (11), 642–650.

Tessarz, P. and Kouzarides, T., 2014. Histone core modifications regulating nucleosome structure and dynamics. *Nature Reviews Molecular Cell Biology*, 15 (11), 703–708.

Thatcher, T.H. and Gorovsky, M.A., 1994. Phylogenetic analysis of the core histones H2A,

H2B, H3, and H4. *Nucleic Acids Research*, 22 (2), 174–179.

Thatcher, T.H., Macgaffey, J., Bowen, J., Horowitz, S., Shapiro, D.L., and Gorovsky, M.A., 1994. Independent evolutionary origin of histone H3.3-like variants of animals and tetrahymena. *Nucleic Acids Research*, 22 (2), 180–186.

Thombre, R.S., Shinde, V.D., Oke, R.S., Dhar, S.K., and Shouche, Y.S., 2016. Biology and survival of extremely halophilic archaeon Haloarcula marismortui RR12 isolated from Mumbai salterns, India in response to salinity stress. *Scientific Reports*, 6 (April), 6–15.

Tokura, M., Ohkuma, M., and Kudo, T., 2000. Molecular phylogeny of methanogens associated with flagellated protists in the gut and with the gut epithelium of termites. *FEMS Microbiology Ecology*, 33 (3), 233–240.

Tomschik, M., Karymov, M.A., Zlatanova, J., and Leuba, S.H., 2001. The Archaeal Histone-Fold Protein HMf Organizes DNA into Bona Fide Chromatin Fibers. *Structure*, 9 (12), 1201–1211.

Topçuoğlu, B.D., Meydan, C., Nguyen, T.B., Lang, S.Q., and Holden, J.F., 2019. Growth Kinetics, Carbon Isotope Fractionation, and Gene Expression in the Hyperthermophile Methanocaldococcus jannaschii during Hydrogen-Limited Growth and Interspecies Hydrogen Transfer. *Applied and Environmental Microbiology*, 85 (9), e00180-19.

Topping, T.B. and Gloss, L.M., 2004. Stability and Folding Mechanism of Mesophilic, Thermophilic and Hyperthermophilic Archael Histones: The Importance of Folding Intermediates. *Journal of Molecular Biology*, 342 (1), 247–260.

Treangen, T.J. and Rocha, E.P.C., 2011. Horizontal Transfer, Not Duplication, Drives the Expansion of Protein Families in Prokaryotes. *PLoS Genetics*, 7 (1), e1001284.

Tropberger, P., Pott, S., Keller, C., Kamieniarz-Gdula, K., Caron, M., Richter, F., Li, G., Mittler, G., Liu, E.T., Bühler, M., Margueron, R., and Schneider, R., 2013. Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. *Cell*, 152 (4), 859–872.

Verma, S.C., Qian, Z., and Adhya, S.L., 2019. Architecture of the Escherichia coli nucleoid. *PLOS Genetics*, 15 (12), e1008456.

Wagih, O., 2017. ggseqlogo: A versatile R package for drawing sequence logos. *Bioinformatics*, 33 (22), 3645–3647.

Wang, S., Su, J., Beliveau, B.J., Bintu, B., Moffitt, J.R., Wu, C., and Zhuang, X., 2016. Spatial organization of chromatin domains and compartments in single chromosomes. *Science*, 353 (6299), 598–602.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh,

T.Y., Peng, W., Zhang, M.Q., and Zhao, K., 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics*, 40 (7), 897–903.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J., 2009. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25 (9), 1189–1191.

Weidenbach, K., Glöer, J., Ehlers, C., Sandman, K., Reeve, J.N., and Schmitz, R.A., 2008. Deletion of the archaeal histone in Methanosarcina mazei Gö1 results in reduced growth and genomic transcription. *Molecular Microbiology*, 67 (3), 662–671.

Weng, L., Feng, Y., Ji, X., Cao, S., Kosugi, Y., and Matsui, I., 2004. Recombinant expression and characterization of an extremely hyperthermophilic archaeal histone from Pyrococcus horikoshii OT3. *Protein Expression and Purification*, 33 (1), 145–152.

Wilhelm, C., Wild, A., Eisenbeis, G., and Zahn, R., 1982. Nanochlorum Eucaryotum: A Very Reduced Coccoid Species of Marine Chlorophyceae. *Zeitschrift fur Naturforschung - Section C Journal of Biosciences*, 37 (1–2), 107–114.

Wilkinson, S.P., Ouhammouch, M., and Geiduschek, E.P., 2010. Transcriptional activation in the context of repression mediated by archaeal histones. *Proceedings of the National Academy of Sciences*, 107 (15), 6777–6781.

Williams, T.A., Foster, P.G., Cox, C.J., and Embley, T.M., 2013. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature*, 504 (7479), 231–236.

Woese, C.R. and Fox, G.E., 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America*, 74 (11), 5088–5090.

Woese, C.R., Kandler, O., and Wheelis, M.L., 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America*, 87 (12), 4576–4579.

Wolfe, J.M. and Fournier, G.P., 2018. Horizontal gene transfer constrains the timing of methanogen evolution. *Nature Ecology and Evolution*, 2 (5), 897–903.

Wood, W.B., 1966. Host specificity of DNA produced by Escherichia coli: Bacterial mutations affecting the restriction and modification of DNA. *Journal of Molecular Biology*, 16 (1), 118–133.

Woodcock, C.L.F., Safer, J.P., and Stanchfield, J.E., 1976. Structural repeating units in chromatin. *Experimental Cell Research*, 97 (1), 101–110.

Xie, R., Wang, Y., Huang, D., Hou, J., Li, L., Hu, H., Zhao, X., and Wang, F., 2022. Expanding Asgard members in the domain of Archaea sheds new light on the origin of

eukaryotes. *Science China Life Sciences*, 65 (4), 818–829.

Xie, Y. and Reeve, J.N., 2004. Transcription by an archaeal RNA polymerase is slowed but not blocked by an archaeal nucleosome. *Journal of Bacteriology*, 186 (11), 3492–3498.

Xu, F., Zhang, K., and Grunstein, M., 2005. Acetylation in histone H3 globular domain regulates gene expression in yeast. *Cell*, 121 (3), 375–385.

Yuan, J., Pu, M., Zhang, Z., and Lou, Z., 2009. Histone H3-K56 acetylation is important for genomic stability in mammals. *Cell Cycle*, 8 (11), 1747–1753.

Zahn, R.K., 1984. A green alga with minimal eukaryotic features:Nanochlorum eucaryotum. *Origins of Life*, 13 (3–4), 289–303.

Zander, A., Willkomm, S., Ofer, S., van Wolferen, M., Egert, L., Buchmeier, S., Stöckl, S., Tinnefeld, P., Schneider, S., Klingl, A., Albers, S.-V., Werner, F., and Grohmann, D., 2017. Guide-independent DNA cleavage by archaeal Argonaute from Methanocaldococcus jannaschii. *Nature Microbiology*, 2 (6), 17034.

Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, Di., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., Stott, M.B., Nunoura, T., Banfield, J.F., Schramm, A., Baker, B.J., Spang, A., and Ettema, T.J.G., 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*, 541 (7637), 353–358.

Zhao, Y. and Garcia, B.A., 2015. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harbor Perspectives in Biology*, 7 (9), a025064.

Zhu, W., Sandman, K., Lee, G.E., Reeve, J.N., and Summers, M.F., 1998. NMR Structure and Comparison of the Archaeal Histone HFoB from the Mesophile Methanobacterium formicicum with HMfB from the Hyperthermophile Methanothermus fervidus. *Biochemistry*, 37 (30), 10573–10580.

Zlatanova, J. and Thakar, A., 2008. H2A.Z: View from the Top. *Structure*, 16 (2), 166–179.

14. Appendix

Table 14.1 Genomes with candidate capstones.
This table was adapted from (Stevens *et al.* 2020).

Species	Amino acid property	Amino acid at position 49 relative to HMfB in alignment
archaeon GW2011-AR5	hydrophobic	F
candidate division MSBL1 archaeon SCGC-AAA259D14	negative	E
candidate division MSBL1 archaeon SCGC-AAA259E17	hydrophobic	A
candidate division MSBL1 archaeon SCGC-AAA259E22	negative	E
candidate division MSBL1 archaeon SCGC-AAA259O05	hydrophobic	A
candidate division MSBL1 archaeon SCGC-AAA261G05	hydrophobic	A
candidate division MSBL1 archaeon SCGC-AAA382F02	hydrophobic	A
candidate division MSBL1 archaeon SCGC-AAA833F18	hydrophobic	A
candidate division MSBL1 archaeon SCGC-AAA833K04	hydrophobic	A
Candidatus Caldiarchaeum subterraneum	hydrophobic	Y
Candidatus Haloredivivus sp. G17	hydrophobic	F
Candidatus Heimdallarchaeota archaeon LC_3	hydrophobic	M
Candidatus Nanosalina sp. J07AB43	negative	D
Candidatus Nanosalinarum sp. J07AB56	negative	D
DHVEG-6 Guayamas35	hydrophobic	Y
Hadesarchaea archaeon DG-33	hydrophobic	I
Hadesarchaea archaeon DG-33-1	hydrophobic	I
Hadesarchaea archaeon YNP_N21	hydrophobic	A
Methanobrevibacter boviskoreani JH1	hydrophobic	I
Methanobrevibacter millerae	hydrophobic	L
Methanobrevibacter olleyae	hydrophobic	I/L
Methanobrevibacter oralis JMR01	hydrophobic	I
Methanobrevibacter ruminantium M1	hydrophobic	I/M
Methanobrevibacter smithii ATCC 35061	hydrophobic	I
Methanobrevibacter sp. AbM4	hydrophobic	I
Methanobrevibacter wolinii SH	hydrophobic	L
Methanocaldococcus villosus KIN24-T80	negative	E
Methanosphaera stadtmanae DSM 3091	negative	E
miscellaneous Crenarchaeota group-15 archaeon DG-45	negative	E
Nanohaloarchaea archaeon B1-Br10_U2g1	negative	D
Nanohaloarchaea archaeon B1-Br10_U2g21	negative/hydrophobic	E/Y
Nanohaloarchaea archaeon B1-Br10_U2g29	negative/hydrophobic	E/Y
Nanohaloarchaea archaeon PLBr10_U2g16	negative	E
Nanohaloarchaea archaeon PLBr10_U2g19	negative/negative	D/E
Nanohaloarchaea archaeon PLBr10_U2g27	negative/hydrophobic	E/Y

Table 14.2 HMfB single amino acid mutants characterised in gel shift assays.
Gel shift results are from (Soares *et al*. 2000). This table was adapted from (Stevens *et al.* 2020).

Mutation	ΔΔG DNA binding (kcal/mol)	ΔΔG stability of the protein (kcal/mol)	Gel shift result from Soares *et al.* (2000)
M1G	-0.229	2.373	Same gel shift as WT
E2D	-1.5081	0.725	Decreased gel shift
E2K	-1.8864	-17.952	Same gel shift
L3C	-0.866	12.1636	Did not accumulate in E. coli
L3I	-0.4955	0.234	Decreased gel shift
P4C	-0.7506	15.609	Same gel shift
P4S	3.5532	14.2371	Same gel shift
I5V	-1.0162	3.758	Same gel shift
A6P	30.558863	-3.104	Did not accumulate in E. coli
P7A	1.4649	13.8786	Same gel shift
I8V	0.0469	4.354	Same gel shift
G9D	0.9748	0.818	Did not accumulate in E. coli
R10S	1.41	0.315	Did not accumulate in E. coli
R10K	1.8425	-6.518	Decreased gel shift
R10G	1.4124	0.416	Did not accumulate in E. coli
I11L	-0.7244	-4.562	Same gel shift
K13E	-0.2283	-4.133	Did not accumulate in E. coli
K13Q	-0.4862	-1.443	Did not accumulate in E. coli
K13T	-0.9661	4.569	Did not accumulate in E. coli
K13R	-2.0568	-1.117	Same gel shift
D14E	-0.3723	-5.831	Did not accumulate in E. coli
D14N	-0.493	-9.57	Same gel shift
D14K	-1.24	-16.484	Decreased gel shift
D14H	-3.0588	-14.311	Did not accumulate in E. coli
A15S	0	7.3696	Same gel shift
A15G	0	7.3371	Did not accumulate in E. coli
E18D	1.2779	0.034	Same gel shift
E18P	0.4728	-5.265	Same gel shift
E18K	-1.4242	3.089	Did not accumulate in E. coli
R19I	4.3091	-3.551	Did not accumulate in E. coli
R19Q	5.2768	-2.048	Did not accumulate in E. coli
R19S	9.9084	1.502	Did not accumulate in E. coli
R19G	10.5705	5.8858	Decreased gel shi ft
R19K	5.6178	-1.358	Decreased gel shift
V20C	-0.4124	9.6386	Did not accumulate in E. coli
V20I	0.3076	-6.426	Same gel shift
V20D	-1.2659	19.0681	Did not accumulate in E. coli
S21A	1.9393	1.831	Decreased gel shift
S21C	-1.3203	7.5993	Same gel shift
S21T	1.5381	-1.531	Decreased gel shift
R25K	5.5543	-6.126	Same gel shift
I26E	0.3859	5.25	Decreased gel shift
T27A	-0.0002	3.174	Same gel shift
T27C	0	6.5756	Same gel shift
I31V	0.0005	6.4488	Same gel shift
I31C	-0.0019	15.9234	Same gel shift
L32C	0	19.2103	Did not accumulate in E. coli
M35C	0	10.6139	Same gel shift
R37K	-0.1871	-8.6	Same gel shift
R37L	0.024	-4.509	Decreased gel shift
R37Q	-0.014	2.771	Decreased gel shift
D38E	0	-3.054	Same gel shift
I39V	0	1.634	Same gel shift
I39C	0	11.2149	Same gel shift
A43I	0.0061	32.1119	Did not accumulate in E. coli
S41C	0.012	0.715	Same gel shift
I44V	0	3.902	Same gel shift
L46I	0.0158	6.3608	Decreased gel shift
L46V	0.0079	10.9782	Decreased gel shift
L46F	-0.0037	4.347	Same gel shift
L46S	0.0021	20.4698	Did not accumulate in E. coli
R48K	0.3128	-0.278	Same gel shift
H49A	0.1917	1.99	Decreased gel shift
H49D	0.5219	11.2511	Same gel shift
G51K	-0.9888	8.6454	Did not accumulate in E. coli
G51A	-0.677	11.4035	Did not accumulate in E. coli
R52K	-0.252	1.341	Decreased gel shift
R52A	1.4069	5.407	Did not accumulate in E. coli
R52H	4.0269	5.6672	Did not accumulate in E. coli

R52Q	-0.5464	3.51	Did not accumulate in E. coli
K53R	1.6919	-1.096	Decreased gel shift
K53E	3.2454	1.266	Did not accumulate in E. coli
K53T	2.1058	2.286	Did not accumulate in E. coli
T54A	3.5459	2.269	Did not accumulate in E. coli
T54C	2.7237	1.158	Did not accumulate in E. coli
T54K	4.9682	1.889	Did not accumulate in E. coli
T54R	7.0644	4.545	Did not accumulate in E. coli
T54S	2.2691	4.55	Did not accumulate in E. coli
T54V	-1.0832	-2.153	Did not accumulate in E. coli
T54Y	19.2299	19.3918	Did not accumulate in E. coli
I55V	0.662	4.123	Same gel shift
I55L	0.2032	-0.262	Decreased gel shift
I55T	0.6911	17.92	Decreased gel shift
I55M	-0.2553	4.661	Decreased gel shift
I55C	0.285	16.4504	Did not accumulate in E. coli
K56R	-1.1339	0.905	Same gel shift
K56T	-1.529	2.93	Did not accumulate in E. coli
K56E	2.0138	-1.73	Did not accumulate in E. coli
K56I	-0.6459	4.121	Did not accumulate in E. coli
K56Q	-1.3688	0.836	Did not accumulate in E. coli
E58S	-1.7313	5.415	Did not accumulate in E. coli
D59A	1.3352	9.3873	Did not accumulate in E. coli
D59E	-1.5257	5.7071	Did not accumulate in E. coli
D59N	-0.1062	13.877	Did not accumulate in E. coli
I60V	0	4.055	Same gel shift
E61A	0.006	1.404	Same gel shift
E61V	-0.0047	3.984	Same gel shift
E61Q	0.0064	-1.091	Same gel shift
E61K	0.0129	-2.887	Decreased gel shift
E61R	-0.0027	-2.443	Decreased gel shift
L62I	0.0099	5.129	Decreased gel shift
L62V	0.007	7.5351	Decreased gel shift
L62Y	-0.096	3.525	Same gel shift
L62M	0.0173	-2.076	Did not accumulate in E. coli
V64R	-0.0057	4.251	Decreased gel shift
R65K	-0.0062	2.925	Same gel shift
R66M	0.0066	-1.184	Same gel shift

Table 14.3 HMfB and *S. cerevisiae* H3/H4 sequences and residue numbering for each position in the alignment used for diversity analysis.
This table was adapted from (Stevens *et al.* 2020).

***M. fervidus* HMfB residue number**	***M. fervidus* HMfB residue**	***S. cerevisiae* H3 residue number (P61830)**	***S. cerevisiae* H3 residue**	***S. cerevisiae* H4 residue number (P02309)**	***S. cerevisiae* H4 residue**
2	E	60	E	29	G
3	L	63	I	30	I
4	P	64	R	31	T
5	I	65	K	32	K
6	A	66	L	33	P
7	P	67	P	34	A
8	I	68	F	35	I
9	G	69	Q	36	R
10	R	70	R	37	R
11	I	71	L	38	L
12	I	72	V	39	A
13	K	73	R	40	R
14	D	74	E	41	R
15	A	75	I	42	G
16	G	76	A	43	G
17	A	77	Q	44	V
18	E	78	D	45	K
19	R	84	R	46	R
20	V	85	F	47	I
21	S	86	Q	48	S
22	D	87	S	49	G
23	D	88	S	50	L
24	A	89	A	51	I
25	R	90	I	52	Y
26	I	91	G	53	E
27	T	92	A	54	E
28	L	93	L	55	V
29	A	94	Q	56	R
30	K	95	E	57	A
31	I	96	S	58	V
32	L	97	V	59	L
33	E	98	E	60	K
34	E	99	A	61	S
35	M	100	Y	62	F
36	G	101	L	63	L
37	R	102	V	64	E
38	D	103	S	65	S
39	I	104	L	66	V
40	A	105	F	67	I
41	S	106	E	68	R
42	E	107	D	69	D
43	A	108	T	70	S
44	I	109	N	71	V
45	K	110	L	72	T
46	L	111	A	73	Y
47	A	112	A	74	T
48	R	113	I	75	E
49	H	114	H	76	H
50	A	115	A	77	A
51	G	116	K	78	K
52	R	117	R	79	R
53	K	118	V	80	K
54	T	119	T	81	T
55	I	120	I	82	V
56	K	121	Q	83	T
57	A	122	K	84	S
58	E	123	K	85	L
59	D	124	D	86	D
60	I	125	I	87	V
61	E	126	K	88	V
62	L	127	L	89	Y
63	A	128	A	90	A
64	V	129	R	91	L
65	R	130	R	92	K

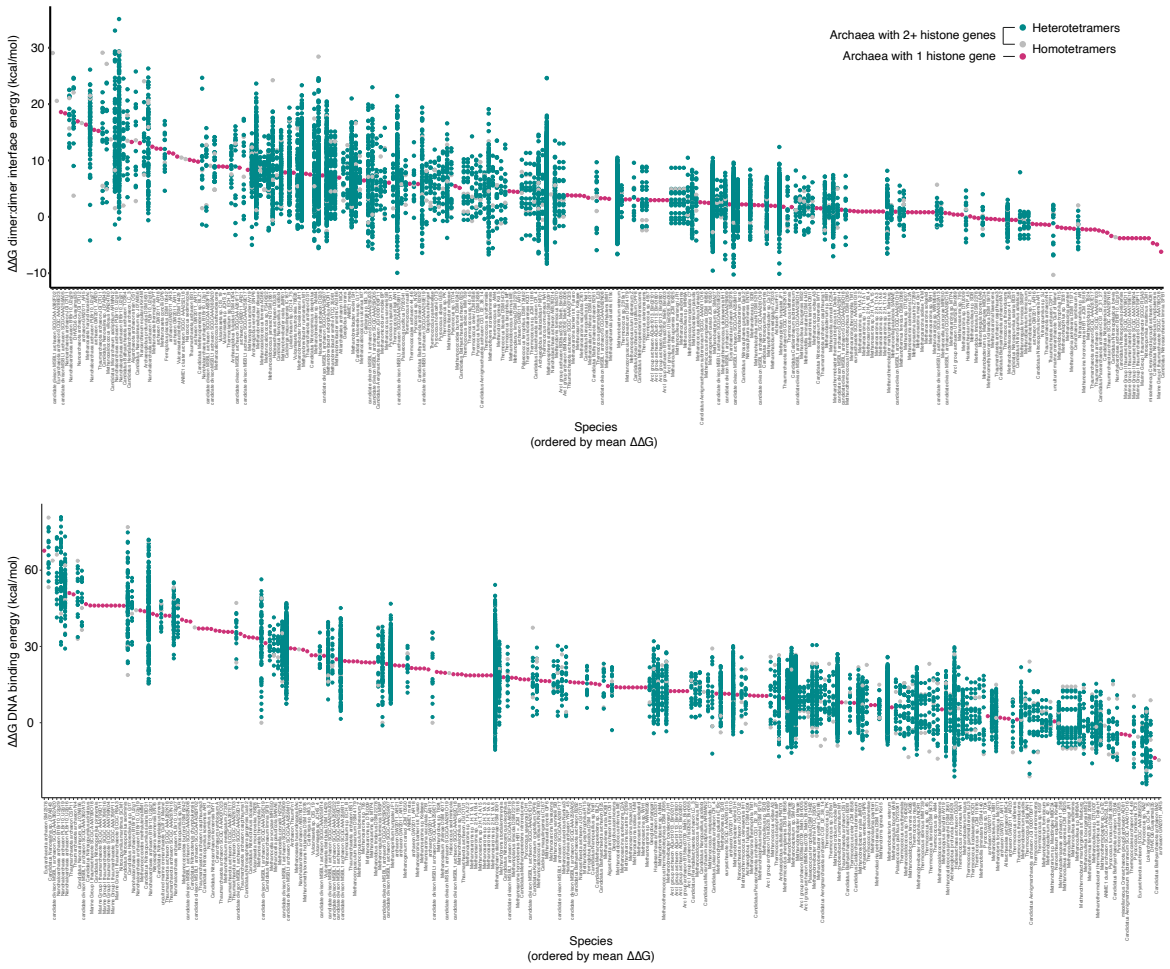


Figure 14.1 DNA binding strength and tetramerization strength for archaeal tetrameric complexes with species names annotated.
Individual complexes are represented by a single point, and ordered by mean energies. Energies are shown relative to HMfB. This figure was published in (Stevens *et al.* 2020).



Figure 14.2 Distribution of substitutions across histones in archaea.
Positions which differ from the canonical residue at each site in the alignment are shown. For each species, the number of histones with each substitution is shown as a heatmap. This figure was published in (Stevens *et al.* 2020).

Histone variants in archaea and the evolution of combinatorial chromatin complexity

Kathryn M. Stevens[a,b], Jacob B. Swadling[a,b], Antoine Hocher[a,b], Corinna Bang[c,d], Simonetta Gribaldo[e], Ruth A. Schmitz[c], and Tobias Warnecke[a,b,1]

[a]Molecular Systems Group, Quantitative Biology Section, Medical Research Council London Institute of Medical Sciences, London W12 0NN, United Kingdom; [b]Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London W12 0NN, United Kingdom; [c]Institute for General Microbiology, University of Kiel, 24118 Kiel, Germany; [d]Institute of Clinical Molecular Biology, University of Kiel, 24105 Kiel, Germany; and [e]Department of Microbiology, Unit "Evolutionary Biology of the Microbial Cell," Institut Pasteur, 75015 Paris, France

Edited by W. Ford Doolittle, Dalhousie University, Halifax, NS, Canada, and approved October 28, 2020 (received for review April 14, 2020)

Nucleosomes in eukaryotes act as platforms for the dynamic integration of epigenetic information. Posttranslational modifications are reversibly added or removed and core histones exchanged for paralogous variants, in concert with changing demands on transcription and genome accessibility. Histones are also common in archaea. Their role in genome regulation, however, and the capacity of individual paralogs to assemble into histone–DNA complexes with distinct properties remain poorly understood. Here, we combine structural modeling with phylogenetic analysis to shed light on archaeal histone paralogs, their evolutionary history, and capacity to generate combinatorial chromatin states through hetero-oligomeric assembly. Focusing on the human commensal *Methanosphaera stadtmanae* as a model archaeal system, we show that the heteromeric complexes that can be assembled from its seven histone paralogs vary substantially in DNA binding affinity and tetramer stability. Using molecular dynamics simulations, we go on to identify unique paralogs in *M. stadtmanae* and *Methanobrevibacter smithii* that are characterized by unstable interfaces between dimers. We propose that these paralogs act as capstones that prevent stable tetramer formation and extension into longer oligomers characteristic of model archaeal histones. Importantly, we provide evidence from phylogeny and genome architecture that these capstones, as well as other paralogs in the Methanobacteriales, have been maintained for hundreds of millions of years following ancient duplication events. Taken together, our findings indicate that at least some archaeal histone paralogs have evolved to play distinct and conserved functional roles, reminiscent of eukaryotic histone variants. We conclude that combinatorially complex histone-based chromatin is not restricted to eukaryotes and likely predates their emergence.

histone variants | chromatin | archaea | evolution

Cells dynamically regulate access to genomic information in response to upstream signals. This may involve wholesale remodeling of chromatin, for example during spermatogenesis where histones are largely replaced by protamines. Other changes in chromatin state constitute less radical tweaks to preexisting chromatin architecture. In eukaryotes, the nucleosome provides the principal platform for such tweaks, prominently via posttranslational modifications (PTMs) but also through the exchange of core histones for paralogous variants (1). Like PTMs, histone variants can alter nucleosome dynamics or affect the recruitment of *trans* factors to reinforce existing chromatin states, establish new ones, or poise chromatin for future change. In many cases, such paralog exchange is regulated and adaptive. For example, in humans, de novo deposition of one histone variant (H2A.X) and eviction of another (H2A.Z) facilitate repair of ultraviolet-induced double-strand breaks (2).

Histones are not restricted to eukaryotes but are also common in archaea, where they assemble into tetramers that are structurally very similar to the $(H3\text{-}H4)_2$ tetramers at the core of eukaryotic nucleosomes (Fig. 1*A*) (3–5). In some archaea, including the model species *Methanothermus fervidus* and *Thermococcus kodakarensis*, additional histone dimers can be tagged onto this tetramer to yield oligomers of increasing length that wrap correspondingly more DNA (3, 6–9). Almost all archaeal histones lack tails and PTMs have yet to be reported. Many archaea do, however, encode multiple histone paralogs (8, 10) that can flexibly homo- and heterodimerize in some species and—in principle—generate chromatin states of considerable combinatorial complexity.

Prior studies in a handful of model species found that archaeal histone paralogs can differ in their expression through the growth cycle, DNA binding affinity, and oligomerization potential, and specific effects on growth and transcription were evident when different paralogs from the same archaeon were deleted (8, 11–14). Yet how the properties of different histone paralogs combine within a single cell to generate dynamic, responsive chromatin states and whether archaeal histone paralogs play conserved roles akin to eukaryotic histone variants remain unknown.

Here, we shed light on the evolution of archaeal histone paralogs and their capacity to generate diverse chromatin states through multimeric assembly. Combining in silico fast mutational scanning with molecular dynamics (MD) simulations and evolutionary analysis, we show that histone paralogs can generate substantial diversity when it comes to key structural properties of the histone–DNA complex. Using *Methanosphaera stadtmanae*—which encodes an unusually large number of histone paralogs (seven)—as a case study, we show that chromatin state space in this multihistone

Significance

Chromatin in eukaryotes is built around histone–DNA complexes, which act as platforms for the integration of regulatory information. Different layers of information are integrated in a combinatorial fashion, for example by exchanging core histones for variants with different properties. We show that histone variants also exist in archaea. In particular, we identify unique archaeal variants that act as capstones, preventing extension of histone–DNA complexes. Importantly, we show that some archaeal histone variants are ancient and have been maintained as distinct units for hundreds of millions of years. Our work suggests that complex combinatorial chromatin that uses histones as its building blocks exists outside eukaryotes and that the ancestor of eukaryotes might have already had complex chromatin.

Author contributions: K.M.S., J.B.S., A.H., and T.W. designed research; K.M.S., A.H., and C.B. performed research; J.B.S., S.G., and R.A.S. contributed new reagents/analytic tools; K.M.S., J.B.S., A.H., and T.W. analyzed data; and K.M.S. and T.W. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

[1]To whom correspondence may be addressed. Email: tobias.warnecke@lms.mrc.ac.uk.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2007056117/-/DCSupplemental.

First published December 7, 2020.

Downloaded from https://www.pnas.org by Imperial College London Library on June 20, 2022 from IP address 193.60.222.2.

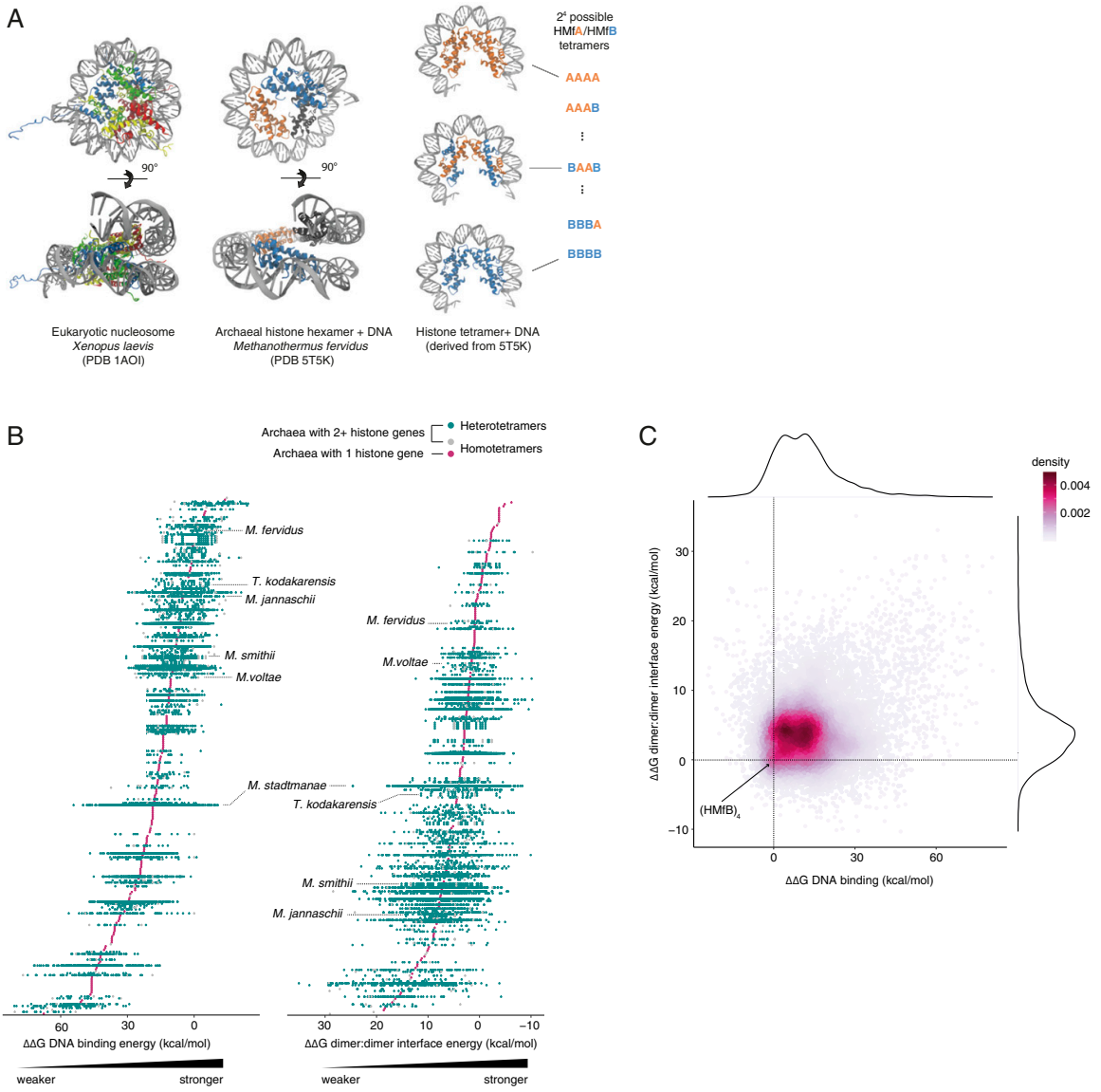


Fig. 1. Structural diversity of archaeal histone tetramers. (*A*) Crystal structures of the octameric eukaryotic nucleosome (PDB 1AOI), the hexameric archaeal nucleosome (PDB 5T5K), and the same structure with one dimer removed to yield the tetrameric complex, alongside a schematic showing the different combinations of histones in homo- and heterotetrameric models built for two histones (e.g., *M. fervidus* HMfA and HMfB). (*B*) DNA binding strength and tetramerization strength (dimer:dimer interface energy) for every possible tetrameric histone complex within each species of archaea in our sample. Each point, grouped by species, represents an individual complex. Species are ordered by mean interaction energy across tetramers. Species labels are provided in *SI Appendix*, Fig. S1. (*C*) Relationship between DNA binding and tetramerization strength for each tetrameric model. Most complexes have slightly weaker tetramerization strength and DNA binding than HMfB. ΔΔG is given relative to the HMfB homotetramer for all plots.

system is large but dense and can be traversed smoothly by altering the dosage of individual paralogs. At the same time, we highlight the potential for more radical change: We describe the widespread existence of capstones—histones that are predicted to prevent further oligomer extension. Importantly, we show that capstones (and other paralogs) in the Methanobacteriales are related by vertical descent, providing evidence for long-term maintenance of functionally distinct paralogs akin to eukaryotic histone variants. Finally, we trace divergent paralog properties to individual amino acid residues and show that paralog diversification has been driven by substitutions at structurally sensitive sites. We propose that paralog exchange might constitute a major mechanism of chromatin state change in archaea, a mechanism that was complemented—and arguably superseded—in eukaryotes by the proliferation of post-translational modifications. Our results suggest that the last common ancestor of eukaryotes, which emerged from within

Downloaded from https://www.pnas.org by Imperial College London Library on June 20, 2022 from IP address 193.60.222.2.

the Archaea (15, 16), might have already possessed histone-based chromatin of considerable combinatorial complexity, with implications for the contribution of histones to the establishment of eukaryotes (17).

Results

Heteromeric Histone–DNA Complexes Exhibit Large Differences in DNA Binding Affinity and Stability across Archaea. Our current knowledge of functional differences among archaeal histone paralogs is limited, especially for archaea with more than two histone genes, where functional diversity might be greatest. Many of these archaea remain genetically inaccessible and/or difficult to culture, preempting detailed experimental characterization. This includes archaea from the Asgard clade, the closest known relatives of eukaryotes (15, 16). To shed light on the functional diversity of histone paralogs in archaea, we therefore combined structural modeling approaches with evolutionary analysis.

First, using the hexameric crystal structure of HMfB from *M. fervidus* as a template, we built models of tetrameric histone complexes bound to DNA for 282 diverse archaea (*Methods*). Tetramers constitute the minimal oligomeric unit capable of wrapping DNA and have been observed in a range of archaea in vivo (3, 9, 18, 19). For archaea with more than one histone gene, we modeled all possible tetrameric combinations (n^4, where n is the number of paralogs; Fig. 1*A*), excluding only histones with large insertions, deletions, or terminal extensions (tails) and those with deletions in the core histone fold (*Methods*). This resulted in 349 homo-oligomeric and 15,905 hetero-oligomeric complexes in total. We then considered Gibbs free energy changes (ΔG) at the DNA–protein interface (a measure of DNA binding affinity) and at the interface between the two histone dimers (a measure of tetramer stability; *Methods*). Across our diverse sample of archaea, we observe substantial apparent variability in DNA binding affinity and tetramer stability (Fig. 1 *B* and *C*; *SI Appendix*, Fig. S1; and Dataset S1). Effective differences between species might, however, be less pronounced than they appear. We model tetrameric complexes under standardized conditions (*Methods*), yet archaea differ widely with regard to growth temperature, pH, the concentration of organic and inorganic solutes, and other factors that can influence protein–protein and protein–DNA interactions in vivo. As attempts to systematically control for such potential confounders are plagued by incomplete information, we focus first on comparisons within species, where different heteromeric complexes can be compared more fairly. In particular, we consider *M. stadtmanae*, a mesophilic methanogen that inhabits the human gut, as a case study.

***M. stadtmanae* as a Case Study for Combinatorially Complex Chromatin.** *M. stadtmanae* DSM3091 encodes seven nonidentical histone genes, located around the chromosome as apparent single-gene operons (*SI Appendix*, Fig. S2). The 7^4 (= 2,401) tetrameric histone–DNA complexes we built from these histones span the largest DNA affinity range (ΔΔG of −10.47 to 54.39 kcal/mol relative to HMfB) and the fourth largest tetramer stability range (ΔΔG of −9.33 to 23.63 kcal/mol) in our sample (Figs. 1 *B* and *C* and 2*A*), providing an excellent model system to interrogate the capacity of an individual archaeal cell to generate different chromatin states by altering the composition of histone–DNA complexes via paralog exchange.

We find that tetrameric combinations are not randomly distributed across this state space but occupy partially distinct areas based on which paralog dominates the complex (Fig. 2*B*). Homotetramers are found toward the edges while the intervening space is densely populated (Fig. 2 *A* and *B*). Complexes that are intermediate in terms of paralog dosage tend to have intermediate properties, enabling smooth transitions in chromatin state space, from one extreme to another (Fig. 2 *B* and *C*). Paralogs in this system therefore provide the capacity for graded control of chromatin state through changes in relative paralog dosage, as well as for more radical transitions (below).

In Vivo Expression of Histone Paralogs in *M. stadtmanae*. Is the capacity for graded control of chromatin state used dynamically in vivo? And what areas of chromatin state space are actually explored? To begin to address the latter question, we quantified the relative abundance of histone paralogs in exponential and stationary phase *M. stadtmanae* cells using label-free mass spectrometry and RT-qPCR (*Methods* and *SI Appendix*, Fig. S3). Protein abundance varies over a 27-fold range between paralogs but expression levels of individual paralogs are well correlated in exponential and stationary phase (Fig. 2*D* and *SI Appendix*, Fig. S3). Intriguingly, relative paralog abundance in exponential phase exhibits a strong correlation with tetramer stability: Paralogs that are inferred to form more stable homotetramers are more abundant (rho = −0.82, $P = 0.034$; Fig. 2*D*). This is also the case (based on previously determined relative transcript/protein abundance) in *Methanobrevibacter smithii*, another member of the order Methanobacteriales, but not, for example, in the hyperthermophile *T. kodakarensis* (*SI Appendix*, Fig. S4).

To mimic the relative abundance of different complexes in the cell and better approximate actual vis-à-vis theoretical chromatin state space in vivo, we generated 100,000 tetrameric complexes in silico, with individual histones recruited into each complex at random based on their relative abundance at the protein level. Assuming that histones dimerize randomly, we find that the center of gravity in chromatin state space shifts toward complexes that are on average less stable, exhibit lower DNA binding affinity (Fig. 2*E* and *SI Appendix*, Fig. S5), and therefore likely give rise to fewer stable higher-order oligomers. This shift is driven by the up-regulation of two histones, Msp_0168 and Msp_0518 (Fig. 2*D* and *SI Appendix*, Fig. S3), which we infer exhibit relatively low DNA binding affinity and tetramer stability as homotetramers. Thus, we predict that stationary phase should—other things equal—be characterized by more open histone-based chromatin. This again contrasts with prior observations in *M. fervidus*, where expression of HMfB—capable of greater DNA compaction—increases in stationary phase relative to HMfA, the second paralog in *M. fervidus* (11). We return to this difference below.

MD Simulations of *M. stadtmanae* Homotetramers. To gain more detailed insights into the extremes of *M. stadtmanae* chromatin state space, we carried out extensive MD simulations on all its homotetrameric histone–DNA complexes. DNA binding affinities inferred from these simulations correlate well with results obtained from fast mutational scanning (rho = 0.96, $P < 0.001$; *SI Appendix*, Fig. S6), providing further validation that the fast mutational scanning approach captures salient properties of the histone–DNA complex. The correlation is less tight for tetramerization energies (rho = 0.5, $P = 0.26$), suggestive of dynamic behavior uniquely captured by MD. Indeed, we find that Msp_0383, the paralog with the lowest predicted DNA binding affinity and tetramer stability, exhibits much more extreme spatial displacement from the starting point of the crystal structure than the other histones (Fig. 3*A*). Further analysis of trajectories over the 100-ns simulation revealed that the Msp_0383 homotetramer displays an unstable dimer:dimer interface—unlike the other homomeric complexes, which reach an approximate equilibrium after <20 ns. While the two Msp_0383 dimers remain individually bound to DNA, they are refractory to tetramerization (Fig. 3*A* and Movies S1 and S2). Thus, our modeling predicts that Msp_0383 assembles into histone–DNA complexes that are structurally distinct from classic tetrameric complexes observed for *M. fervidus,* other model archaea, and the remaining *M. stadtmanae* paralogs.

Msp_0383 has a negatively charged glutamic acid (E) at position 49 whereas the other paralogs (and most histones across

Downloaded from https://www.pnas.org by Imperial College London Library on June 20, 2022 from IP address 193.60.222.2.

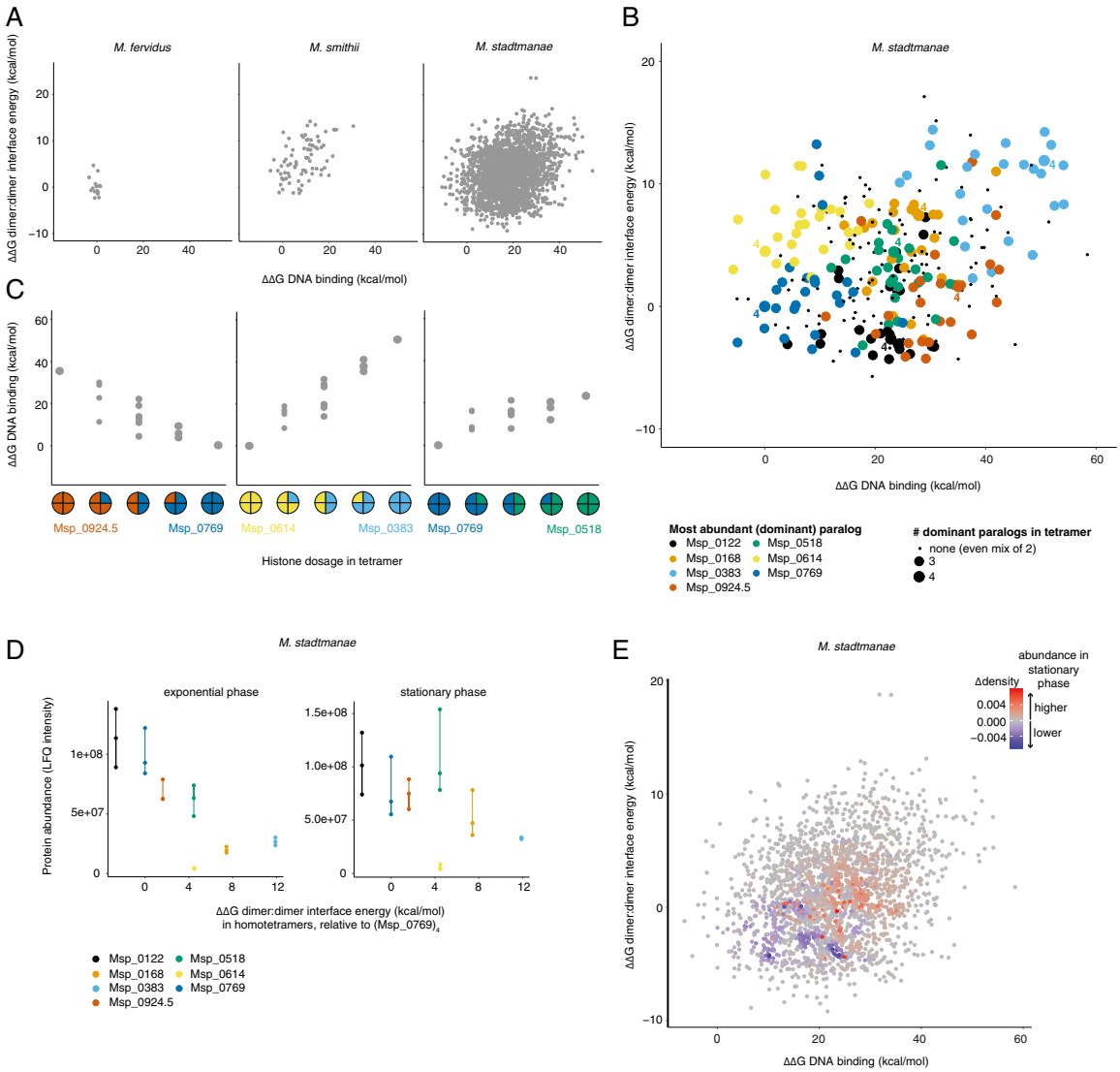


Fig. 2. Structural diversity of *M. stadtmanae* tetrameric histone–DNA complexes. (*A*) Chromatin state space defined by DNA binding and tetramerization strength for models built using fast mutational scanning for *M. fervidus* (two histone paralogs), *M. smithii* (three histone paralogs), and *M. stadtmanae* (seven histone paralogs). ΔΔG is given relative to the HMfB homotetramer. (*B*) Chromatin state space defined by DNA binding and tetramerization strength for *M. stadtmanae* histone complexes containing two or fewer histone paralogs. Points are colored by the dominant paralog in the complex (three or four out of four monomers in the tetramer). Homotetramers are labeled ("4"). (*C*) Examples of DNA binding strength varying gradually with paralog dosage. (*D*) Tetramerization strength of *M. stadtmanae* homotetramers compared to empirically determined paralog abundance in exponential and stationary phase. Each data point represents protein abundance measured in one biological replicate. (*E*) Relative change in the abundance of different tetrameric complexes in stationary versus exponential phase, as predicted by sampling 100,000 tetrameric complexes based on relative protein abundance (mean LFQ intensity) in exponential and stationary phase. Increased abundance of complexes in stationary phase is shown in red and decreased abundance in blue. ΔΔG is given relative to Msp_0769 for *B*–*E*.

archaea) have a positively charged histidine (H, Fig. 3*A*) (note that, throughout this article, we number residues based on positional orthology to HMfB; the raw residue number in *M. stadtmanae* is 50). Residue 49 is close to the interface between dimers and mutations at this site were previously shown to impact tetramer formation of HMfB in vitro (18). To test whether amino acid identity at this site is sufficient to account for the repulsive effects observed, we in silico-substituted E for H in all histones of the tetramer and subjected the resulting complex to the same simulation protocol. We find that this substitution alone is enough to significantly reduce the distance between dimers, with Msp_0383^{E49H} exhibiting dynamics that are intermediate between Msp_0383 and the other paralogs (Fig. 3*B*). These results suggest that Msp_0383 functions as a capstone, preventing tetramerization and, when tagged onto an existing complex, further oligomerization.

Downloaded from https://www.pnas.org by Imperial College London Library on June 20, 2022 from IP address 193.60.222.2.

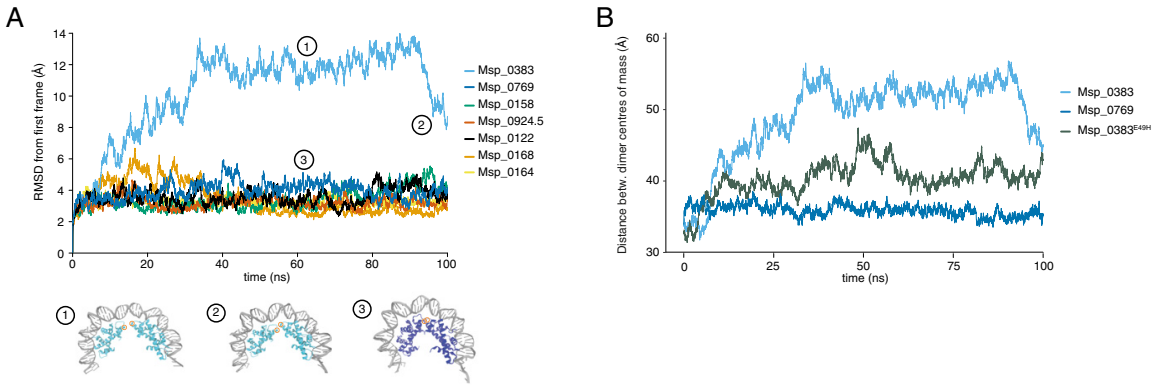


Fig. 3. MD simulations of homotetrameric histone models from *M. stadtmanae*. (*A*) Root-mean-square deviation (rmsd) over the course of the simulation, beginning from the crystal structure into which *M. stadtmanae* histones have been substituted (*Methods*). Example structures at different frames are shown for Msp_0383 and Msp_0769 with the position of residue 49 highlighted in orange. (*B*) Distance between the centers of mass of dimers in homotetrameric models of Msp_0383, Msp_0769, and the mutant Msp_0383^{E49H}.

Such potential capstones are not unique to *M. stadtmanae* but are also present in other members of the Methanobacteriales, as demonstrated by comparative MD simulations of *M. smithii* homotetramers, which also reveal a single, lowly expressed histone (Msm_1260) associated with much-reduced tetramer stability (*SI Appendix*, Figs. S4 and S7).

Phylogenetic Analysis Reveals Long-Term Persistence of Archaeal Histone Variants. Some eukaryotic histone variants are ancient and have persisted through multiple rounds of speciation as recognizable, distinct paralogs, often with conserved function and dedicated chaperones that can discriminate between them (20). Notably, this includes H2A.Z, which emerged at the base of eukaryotes. Other variants, like macroH2A, are restricted to certain clades and therefore evolved more recently. Yet others, like H2A.X, appear polyphyletic in origin, pointing to repeated independent emergence of functionally analogous variants (5, 20). What is the situation for archaeal histones? Are there persistent, recognizable paralogs of ancient origin? Or is most diversification relatively recent and lineage specific?

Phylogenetic analysis of histones across archaea is complicated by the fact that histones are short (<70 amino acids) and timescales are large, leading to poorly supported nodes in a global phylogeny of archaeal histones (*Methods* and Datasets S2–S7). We therefore focused our analysis on the Methanobacteriales, which include *Methanosphaera*, *Methanobrevibacter*, and *Methanobacterium* spp. as well as *M. fervidus* (Fig. 4*A*). Alongside abundant lineage-specific duplication events (lighter taxon labels in Fig. 4 *B* and *C*), we find several cases of longer-term paralog maintenance, indicated by the existence of multiple groups of sequences that each recapitulate the species phylogeny (Fig. 4 *B* and *C* and *SI Appendix*, Fig. S8). For example, branching patterns—as well as conserved synteny—of histones in *Methanobacterium* strongly suggest two ancient gene duplication events that preceded the divergence of this genus (Fig. 4*B*). Importantly, synteny analysis also reveals maintenance of paralogs between *Methanobrevibacter* and *Methanobacterium* (groups 1 and 3 in Fig. 4 *B* and *C*), indicating that these originated from even more ancient duplications, dating back to the ancestor of these two genera. As synteny breaks down further, making confident assignments becomes harder. Closer inspection of local gene neighborhoods, however, suggests that there might be even deeper conservation of recognizable paralogs all the way out to *M. fervidus*, where *hmfB* (*hmfA*) is flanked upstream (downstream) by *trpS* (*radB*), whose relative position is conserved in *Methanobrevibacter* and *Methanobacterium* spp. (Fig. 4 *B* and *C*). We found no evidence for gene conversion between paralogous histones in this clade (*Methods*).

To put the timescale of paralog origin into context, we note that the lineages leading to *M. stadtmanae* and *M. smithii* split an estimated ∼1.3 Gya, while the wider Methanobacteriales are thought to have emerged as a clade ∼1.6 Gya (21). At least some archaeal histone variants have therefore been maintained for hundreds of millions of years of evolution, rendering them comparable in age to the oldest known eukaryotic histone variants, which date back to the last common ancestor of eukaryotes, roughly 1.2 to 2 Gya (22, 23).

Regarding capstones, we find evidence for shared vertical descent of the *M. stadtmanae* and *M. smithii* capstones (Fig. 5). At the same time, we note that histones with negatively charged/hydrophobic amino acids at residues 49 are also present in multiple independent lineages outside the Methanobacteriales, including members of the Hadesarchaea and Nanohaloarchaeota (*SI Appendix*, Fig. S7 and Table S1). Additional MD simulations for histones from each of these clades show unstable dimer:-dimer interfaces like those we observed for *M. stadtmanae* and *M. smithii* capstones (*SI Appendix*, Fig. S7). In as much as amino acid identity at residue 49 can be used as a diagnostic marker, this suggests that capstone functionality has evolved multiple times independently.

Single Amino Acid Changes Underpin Functional Differences between Paralogs. The case of *M. stadtmanae* Msp_0383 illustrates that substitutions of individual amino acids can have strong effects on histone properties and, ultimately, chromatin state. This is also true in eukaryotes (1, 24, 25). H3.3 and H3.1, for example, differ in only four amino acids (three of which are located in the histone fold domain), but are recognized by different chaperones, deposited at defined locations along the genome, and make distinct, nonredundant contributions to genome function, notably during gametogenesis (20, 26, 27).

To understand how specific amino acid changes underpin the functional diversification of archaeal histone paralogs, we integrated structural modeling and evolutionary analysis. First, we used the FoldX forcefield (*Methods*) to in silico mutate each amino acid in the model histone HMfB from *M. fervidus* to every other possible amino acid to identify sites particularly sensitive to change. We then compared these predicted effects to previous in vitro work on HMfB, which had identified residues that, when

Downloaded from https://www.pnas.org by Imperial College London Library on June 20, 2022 from IP address 193.60.222.2.

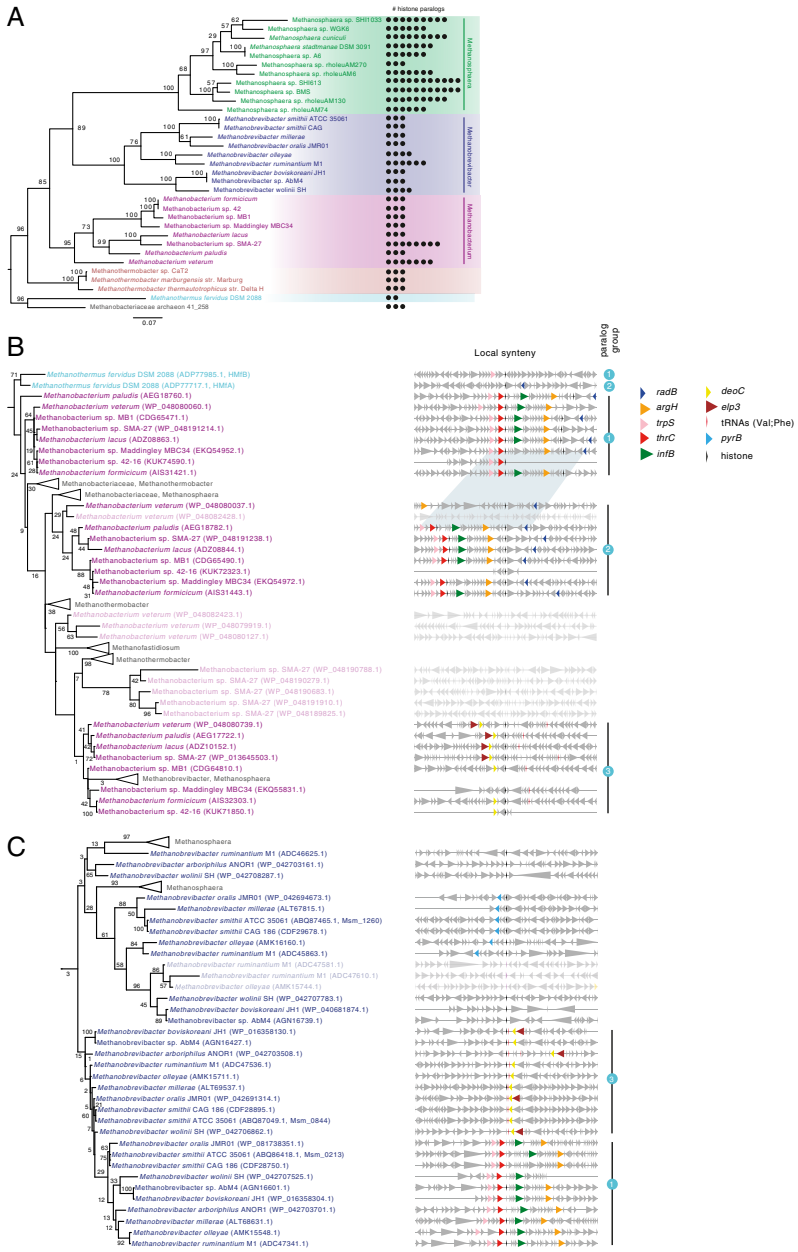


Fig. 4. Phylogenetic analysis of histone paralogs in the Methanobacteriales. (*A*) Maximum-likelihood reference phylogeny (species tree) of the order Methanobacteriales, using IF-2a as a representative, vertically inherited gene. Bootstrap values are shown as a percentage out of 200 nonparametric bootstraps. The tree is rooted with *M. fervidus* and Methanobacteriaceae archaeon 41_258 as the outgroup. The number of histone paralogs in a given genome is mapped on the right-hand side. (*B* and *C*) Examples of recent duplications (lighter taxon labels) and long-term maintenance of paralogs in the genera *Methanobacterium* (*B*) and *Methanobrevibacter* (*C*), as supported by tree topology and conserved synteny. Note the clustering of proteins according to shared synteny rather than by species. Examples of paralog groups (1/2/3) are highlighted. Shared synteny across the *Methanobacterium*/*Methanobrevibacter* divide supports paralogous relationships (groups 1 and 3). Even deeper paralogy is suggested by the fact that local spatial association of *trpS* (group 1) and *radB* (group 2) with histones extends to *M. fervidus*. Longer gaps in the synteny blocks, evident for individual genomes, are the result of incomplete genome annotations. Genes are automatically color coded based on similarities in functional annotation (*Methods*). The trees are rooted with reference to a wider phylogeny of archaeal histones (*Methods*) and bootstrap values shown as a percentage out of 500 nonparametric bootstraps. In both *B* and *C* some sequences from other Methanobacteriales have been collapsed for clarity. The scale bar represents the average number of substitutions per site. Full trees and the underlying alignments are provided as Datasets S2–S7.

mutated, affect DNA binding, the direction of DNA supercoiling, rigidity of the histone–DNA complex, thermostabilization, oligomer formation, and the ability of the histone to accumulate in *Escherichia coli*, a proxy for folding stability (3, 18, 28–30). We find that predicted and observed effects are highly concordant (Fig. 6 *A* and *B* and *SI Appendix*, Table S2). For example, our fast mutational scanning identifies the four residues (46/49/59/62; Fig. 6*D*) previously highlighted as critical for stable tetramerization (8, 18) and we predict, more often than not, whether, in previous gel shift assays (30), a specific mutation had led to increased (stronger DNA binding) or decreased (weaker DNA binding) mobility (Fig. 6*B*). This high degree of congruence provides additional validation for our modeling approach. It also increases our confidence in predictions of structural sensitivity for residues that have not been experimentally interrogated. For example, residues 21 and 50, for which no experimental data are available, show large deviations in DNA binding affinity and tetramerization strength, respectively, when mutated (Fig. 6*D*).

Next, we asked how this comprehensive landscape of possible effects compares with substitutions that actually occurred during the evolution of archaea. Do structurally sensitive sites remain largely conserved across paralogs? Or are changes at key sites, like those we observe for Msp_0383, relatively commonplace? To answer this question in a pan-archaeal manner, we took a

Downloaded from https://www.pnas.org by Imperial College London Library on June 20, 2022 from IP address 193.60.222.2.

nonphylogenetic approach. We aligned the 506 archaeal histone proteins in our sample (*Methods*) and then split them into two groups, depending on whether they come from a genome that encodes only a single histone gene or from a genome that encodes two or more paralogs. Our objective here was to identify residues along the histone fold that have become more diverse in multiparalog systems, where relaxed constraint or positive selection could drive diversification following duplication. Fig. 7 shows the amino acid diversity ratio H_M/H_s for each residue, where H_M and H_s are Shannon diversity indexes calculated for a given residue (column in the alignment) across multihistone and single-histone genes, respectively (*Methods*). The average Shannon ratio will be affected by phylogenetic sampling, the number of histones in each group, and other factors and is therefore relatively uninformative. What is informative, however, are deviations from this average at specific residues.

Strikingly, diversification in species with multiple paralogs is strongly associated with structurally critical residues (Fig. 7*A*). This includes the capstone residue 49, but also several residues that make large contributions to DNA binding (10/19/53/54; Figs. 6 *C* and *D* and 7*A*), concentrated in the loop regions of the histone fold, and loop 2 in particular. Perhaps the most egregious example is residue 19 in loop 1, which is perfectly conserved as an arginine in single-histone archaea ($H_s = 0$) but accommodates eight different amino acids across the multihistone archaea in our sample. This suggests a significant change in the evolutionary regime at this site once more than two histones are present in the system. Given the strong deviation from the baseline diversity ratio, we think that positive selection is likely implicated in the diversification process rather than relaxed constraint alone.

Although we do not explore this extensively here, we note that residue-level diversification has some phylogenetic structure. Some residues exhibit a narrow phyletic pattern (*SI Appendix*, Fig. S9). Notably, this includes residue 54, which forms a conserved interaction with residue 19 (3). Diversification at this residue is confined almost entirely to the Asgard clade and excluding this clade from the analysis dramatically reduces diversity at residue 54 (*SI Appendix*, Fig. S10).

Diversification of Eukaryotic versus Archaeal Histone Folds. The residues involved in archaeal histone tetramerization are also important for interactions at the interface of two H3 molecules from neighboring H3:H4 dimers (31, 32). How, then, do archaeal histone paralogs compare to eukaryotic histone variants? Did diversification of the core histone fold follow a similar path? To address this question, we first added eukaryotic H3 and H4 sequences to our preexisting alignment of archaeal histones (*Methods* and *SI Appendix*, Table S3). We then calculated Shannon diversity indexes for H3 (H_3) and H4 (H_4) proteins found across eukaryotes and compared H_4/H_s and H_3/H_s to H_M/H_s. We find that diversification dynamics across the histone fold follow a similar pattern in multihistone genes and H3 (rho = 0.40, $P = 0.0081$) and to a lesser extent also H4 (rho = 0.24, $P = 0.063$). Residues 2 and 33, which are involved in intramonomer interactions, are not diverse in H3, H4, or archaeal histones. Substitutions at these positions may prevent the formation of the tertiary histone fold structure and are therefore selected against. Conversely, residues around the loop 2 region in particular experience accelerated diversification in both H3 and multihistone archaea relative to single-histone archaea. These similarities notwithstanding, several residues show conspicuous diversification in multiarchaeal histones but not H3/4, and vice versa. This includes residues 49 (high H_M/H_s, low $H_{3/4}/H_s$), 59 (high H_M/H_s, low H_4/H_s), and 43 (low H_M/H_s, high H_4/H_s). In addition, even residues with high diversity ratios in both eukaryotes and archaea only partially explore the same part of sequence space and tend to evolve toward different sets of amino acids (*SI Appendix*, Fig. S10). Our results therefore suggest that histone variants from archaea and eukaryotes independently focused their exploration of structural–functional space on structurally sensitive sites in the loop 2 region but also highlight significant lineage-specific constraints on histone evolvability.

Discussion

Prior observations—from variable expression along the growth cycle to differential phenotypic effects upon deletion (11, 13, 14)—pointed to functional diversity of archaeal histone paralogs. The observations we report here not only reinforce this notion but also demonstrate that some histone paralogs in archaea have been maintained as distinct functional units over long evolutionary timescales, akin to eukaryotic histone variants.

Our modeling results suggest that paralogs, by exploiting the combinatorial opportunities of histone oligomerization, can generate diverse chromatin states at the level of individual histone–DNA complexes and enable both subtle, graded dosage-driven transitions and more radical changes such as those associated with the expression of capstones. We explored one of these transitions in

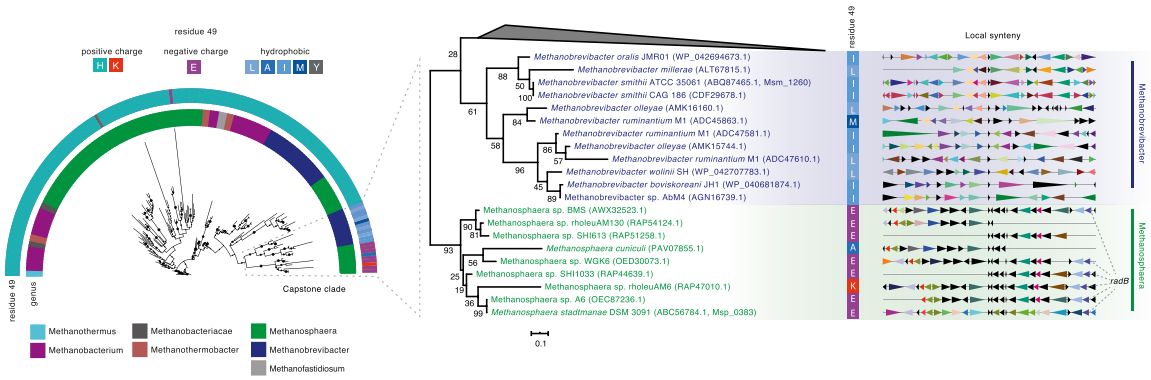


Fig. 5. Phylogenetic analysis of capstone histones. A maximum-likelihood tree including all Methanobacteriales genomes (Dataset S7) is displayed at *Left*, along with information on amino acid identity at residue 49. Histone paralogs with capstone properties (negatively charged or hydrophobic amino acids) in *Methanosphaera* and *Methanobrevibacter* spp. cluster to the exclusion of other histones found in these species. Local synteny in the vicinity of histone paralogs is shown at *Right*. Genes are automatically color coded based on similarities in functional annotation. The *radB* gene is highlighted to allow cross-referencing with Fig. 4 *B* and *C*. Bootstrap values are shown as a percentage out of 500 nonparametric bootstraps. The scale bar represents the average number of substitutions per site.

Downloaded from https://www.pnas.org by Imperial College London Library on June 20, 2022 from IP address 193.60.222.2.

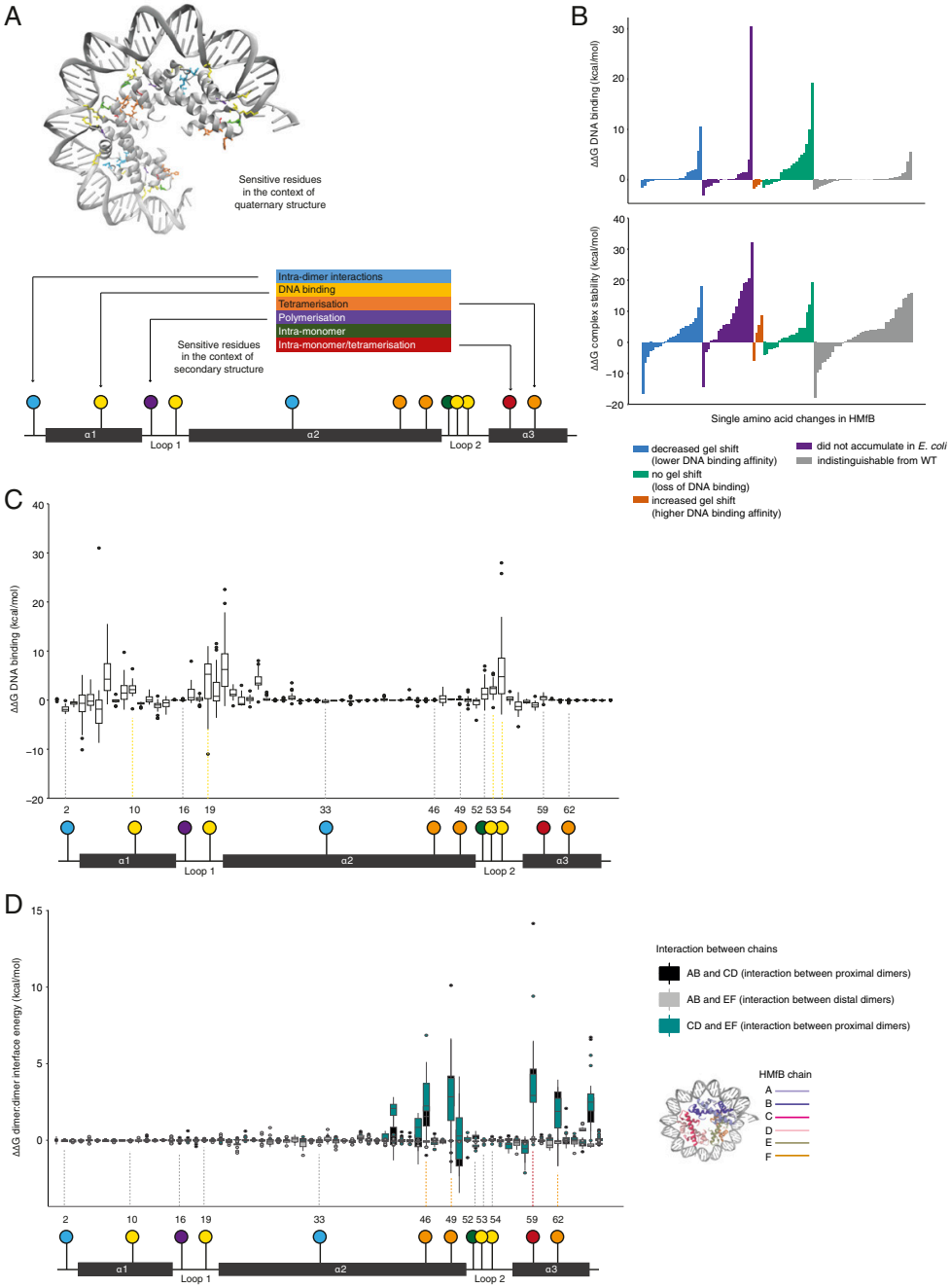


Fig. 6. Modeling the impact of single amino acid mutations on the $(HMfB)_6$–DNA complex. (*A*) Residues where mutations are known from previous experimental work (3, 18, 28–30) to affect monomer:monomer interactions, DNA binding, tetramerization, polymerization, and intramonomer interactions are highlighted on the quaternary and secondary structures. (*B*) FoldX-calculated changes in DNA binding affinity (*Top*) and stability (*Bottom*) for HMfB single amino acid mutants previously characterized qualitatively in gel shift experiments (30). Individual mutations are listed in *SI Appendix*, Table S2. (*C* and *D*) DNA binding (*C*) and tetramerization strength (*D*) for all possible single amino acid mutations of HMfB. The location of residues with previously known function is shown on the secondary structure beneath. For *D*, the resulting interaction energy between each dimer pair in the hexamer was calculated and the location of dimer pairs in the hexamer is shown. ΔΔG is quoted relative to the wild-type HMfB structure for all plots.

Downloaded from https://www.pnas.org by Imperial College London Library on June 20, 2022 from IP address 193.60.222.2.

M. stadtmanae, where the relative expression of different histone paralogs changes in stationary versus exponential phase. Based on our structural modeling and empirical protein abundance data, we predict that stationary phase in *M. stadtmanae* (as well as *M. smithii*) should be characterized by a larger fraction of less stable histone–DNA complexes. This is, arguably, unexpected given opposite trends inferred for *M. fervidus* and other hyperthermophiles and the general notion that stationary phase is associated with greater chromatin compaction. Experimental data will ultimately be required to determine whether this inferred difference is genuine or not. However, the discrepancy serves as a timely reminder to highlight the limitations of our modeling approach, which does not consider absolute histone titers, changes in intracellular conditions (e.g., in terms of solutes), and expression of other abundant architectural proteins (e.g., Alba) that will codetermine higher-order chromatin states. In this regard, our results should be considered a valuable starting point and incentive for further exploration rather than the final word, hewn in stone, on comparative chromatin complexity in archaea. Substantial further work, both in vitro and in vivo, will be required to elucidate why certain histone properties have been selected for in different lineages and how individual paralogs are deployed in physiological context.

It is also worth noting that we examined only a small branch of the archaeal tree in depth, did not consider archaeal histone with tails or large indels (8, 33), and did not explore interactions and combinatorial complexity beyond the tetramer level. Our estimates of archaeal capacity to generate different chromatin states are therefore likely conservative. In particular, tetramer models do not allow us to consider stacking interactions between nonadjacent dimers, which affect oligomerization propensity (3, 8). Substantial additional complexity might further emerge from the consideration of N-terminal tails, which are present in some Heimdallarchaea (3, 8), the closest known relatives of eukaryotes (16). Studying these archaea and their tails will be particularly important to understand what—in the context of histone-based chromatin—constitutes eukaryotic innovation, elaboration, or shared archaeal heritage. This includes the question of whether "deep paralogy" might exist between extant eukaryotic and archaeal histones, something our results do not directly imply.

Based on our current knowledge, we speculate that paralog-mediated structural change might play an outsize role in archaea compared to eukaryotes, where posttranslational modifications and interactions with other proteins are heavily involved in altering chromatin state in response to upstream signals. One of the key eukaryotic innovations might have been a switch from predominantly paralog-based generation of different chromatin states to using an octameric nucleosome as a platform for integrating epigenetic information. This innovation might also have enabled another: local specification. In eukaryotes, divergent regulatory states can be encoded along the same chromosome via targeted deposition of paralogs and histone marks by enzymes and chaperones that can interact with specific histones, DNA sequences, and/or other constituents of chromatin. At present, we have no evidence that the capacity for such local control exists in archaea. Current data support only a global, genome-wide role in reshaping chromatin state. It will be interesting in the future to determine whether complexes of different composition are indeed randomly distributed or show nonrandom patterns along archaeal chromosomes in a manner anticipating eukaryotic chromatin. To this end, we need to develop a better understanding of archaeal histone variants in physiological context. The specific functional roles of archaeal variants in the context of genome function remain entirely unknown, a glaring gap that can only be plugged by in vivo experiments. Our study provides ample incentive for further research to establish how archaeal paralogs are regulated, how they interact with other DNA-binding proteins to determine global and perhaps local chromatin states, and how paralogs contribute to adaptive responses in physiological context.

Methods

Alignment of Histones. A previously compiled set of archaeal histone proteins (10) was filtered to include only proteins between 60 and 80 amino acids in length with a single histone fold (*SI Appendix*, Fig. S11). For reference, HMfB is 69 amino acids long. Sequences were further filtered to randomly remove redundant entries (i.e., sequences 100% identical to another entry). This filtered set of histones from 282 species of archaea (139 with more than one histone, 143 with one histone) was aligned using MAFFT-linsi (-localpair -maxiterate 1000) (34). Eukaryotic H3 and H4 protein sequences were downloaded from InterPro (matching folds IPR007125, IPR035425, and IPR032454) (35), filtered for length (95 to 110 amino acids for H4, 130 to 145 amino acids for H3; *SI Appendix*, Fig. S11), and added to the archaeal histone alignment using MAFFT-linsi (-localpair -maxiterate 1000 -seed) after removing H2A/B sequences. Positions where more than 5% of sequences had a gap were removed from further analysis.

HMfB Single Mutants. We used the BuildModel command in FoldX (36) to introduce all possible single amino acid changes into the HMfB hexamer (Protein Data Bank [PDB] structure 5T5K). All six histone monomers in the structure were mutated simultaneously. FoldX refines structures by minimizing the energy of mutated side-chain residues and neighboring residues according to its empirically derived forcefield. The positions of nonadjacent residues and all peptide backbone atoms remain fixed. Although MD simulations are more rigorous to determine accurate binding affinities and allow us to sample the dynamics of the complex (below), FoldX allows us to sample, at high throughput, changes in energy associated with mutations at individual positions in the protein. We therefore refer to this approach as a fast mutational scanning technique. FoldX was used at the default temperature setting of 298 K. We calculated the relative change in Gibbs free energy ($\Delta\Delta G$) of the system, DNA binding, and tetramerization energies for each mutant using FoldX relative to the minimized HMfB hexamer structure.

The Gibbs free energy (Eq. **1**) is a thermodynamic quantity defined as the amount of reversible work a mechanical system can undergo, where ΔH is the enthalpic contribution and ΔS is the entropic contribution. By calculating the sum total of inter- and intramolecular forces, determined by the FoldX forcefield, we can calculate ΔG and predict the structural stability of the complex. By subtracting ΔG_{mutant} from $\Delta G_{wildtype}$ of HMfB we arrive at the relative change in Gibbs free energy, $\Delta\Delta G$ (Eq. **2**). The binding affinity can be determined by subtracting the energetic contribution from the DNA and histone from the complex (Eq. **3**). The same can be said for the histone tetramerization energy; subtracting dimer energies from the tetramer energy will leave us with the energetic contribution of tetramerization:

$$\Delta G = \Delta H - T\Delta S \quad [1]$$

$$\Delta\Delta G = \Delta G_{mutant} - \Delta G_{wildtype} \quad [2]$$

$$\Delta G_{bind} = \Delta G_{complex} - \Delta G_{DNA} - \Delta G_{histone}. \quad [3]$$

Tetramer Models of Archaeal Histones. The HMfB tetramer model was built by removing one histone dimer (chains E and F) and 30 bp of DNA from the 5T5K PDB structure. For each species, all possible combinations of histone monomers were modeled as a tetramer, with the following exceptions: To enable fair structural comparison, we analyzed only models where no histone carried a deletion in the core histone fold (HMfB residues 2 to 65) and considered only histones 60 to 80 amino acids in length. We focused on tetramers as this allows DNA binding and tetramerization strength to be calculated without assuming that histones assemble into longer oligomers. Substitutions at positions in the core histone fold were mapped onto the HMfB tetramer using the BuildModel function of FoldX (36). Structures were energy minimized for 10,000 steps of combined steepest descent and conjugate gradient using AmberTools. Unlike FoldX, which minimizes only mutated side-chain residues and their neighbors, we used an all-atom minimization (using AMBER ff14SB) but avoided any significant refolding by applying a 2-kcal·mol^{-1}·Å^{-2} harmonic restraint on backbone atoms.

Binding affinity and tetramerization energies were calculated using the single-trajectory molecular mechanics Poisson–Boltzmann surface area (MMPBSA) approach (37). In this method we decompose ΔH in Eq. **1** into the gas phase energy and the free energy of solvation (Eq. **4**). The gas phase energy was calculated as the total of energy from the AMBER ff14SB

Downloaded from https://www.pnas.org by Imperial College London Library on June 20, 2022 from IP address 193.60.222.2.

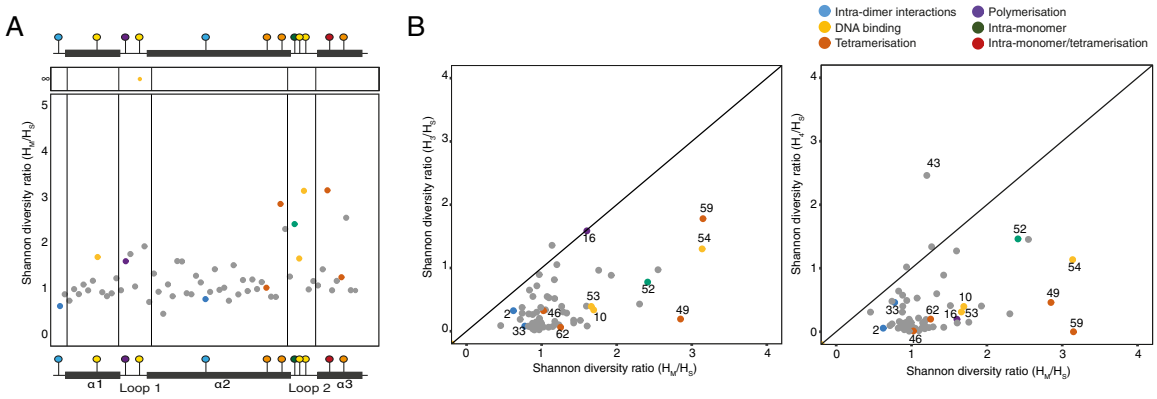


Fig. 7. Comparative analysis of sequence diversity in archaeal and eukaryotic histones. (*A*) Shannon diversity ratio (H_M/H_S) at each position in the core histone fold domain. Residues are colored by key function from previous mutational studies (Fig. 6). H_S for residue 19 is 0, so the Shannon ratio is undefined. (*B*) Shannon diversity ratios for H3 (H_3/H_S, *Top*) and H4 (H_4/H_S, *Bottom*) compared to H_M/H_S. Residues of particular interest are numbered.

forcefield (38) and the free energy of solvation was approximated using the Poisson–Boltzmann equations:

$$\Delta G = (E_{gas} + E_{solv.}) - T\Delta S. \quad [4]$$

In our present energy minimization scheme, single mutations will not significantly change the conformation of the histone and the relative change in entropy, $\Delta\Delta S$, will be close to zero. For this reason, we have not included the entropic contribution to the Gibbs free energy values. This is not to say that a single-residue mutation will never perturb the conformational landscape to an extent that would lead to a significant change in the entropic contribution to energy. Although rare, these processes may happen over longer timescales that are not accessible using our current methods.

$\Delta\Delta G$ was calculated relative to the Msp_0769 homotetramer for *M. stadtmanae* tetramer models and relative to the HMfB homotetramer in all other cases.

MD Simulations of *M. stadtmanae*. Complexes of homotetrameric histones with DNA were parameterized using the Amber ff14SB potentials for canonical proteins using tLeap in AmberTools. Residues present in the sequence but removed in the filtering stage after alignment were manually added to the "full model" homotetrameric structures generated by FoldX and the complexes were energy minimized as above. Models were solvated with 14 Å of transferable intermolecular potential with 3 points water and neutralized with NaCl (~0.18 M), countering the overall negative charge of the DNA backbone. Energy minimization was performed for 2,000 steps using combined steepest-descent and conjugate gradient methods. Following minimization, 20 ps of classical molecular dynamics (cMD) was performed in the constant temperature, constant volume (NVT) ensemble using a Langevin thermostat (39) to regulate the temperature as we heated it up from 0 to 300 K. Following the heat-up phase, we performed 100 ns of cMD in the isobaric/isothermal (constant temperature, constant pressure [NPT]) ensemble using the Berendsen barostat (40) to maintain constant pressure of 1 atm during the simulation. All simulations were performed using GPU (CUDA) version 18.0.0 of PMEMD (41–43) with long-range electrostatic forces treated with particle-mesh Ewald summation (44). MMPBSA calculations for DNA binding affinity and tetramerization strength were performed from frames 1,500—where the rmsd of each trajectory had started to equilibrate—to the end.

Phylogenetic and Evolutionary Analysis. To build an initial tree of archaeal histones, we queried all 282 species present in the structural analysis and available through NCBI with hmmsearch (HMMer suite, http://hmmer.org) and considered all single-domain hits against Pfam model CBFD_NFYD_HMF (PF00808, Pfam v.23) that were filtered out from the initial dataset. For reproducibility purposes, the Pfam gathering threshold was used as the thresholding option of hmmsearch (–cut_ga). Sequences were first aligned with MAFFT-linsi (using blosum30) and an initial tree inferred with IQ-TREE2 (automatic substitution model estimation: LG+R6 substitution model, 1,000 ultrafast bootstraps) (45). We then considered the minimal subtree containing all histones from *M. stadtmanae*. To extend the diversity of Methanosphaera histones, we downloaded additional available Methanosphaera genomes from the NCBI refseq database. All sequences were then realigned using MAFFT-linsi (using blosum62), and a maximum-likelihood tree was built using RAxML-ng (500 nonparametric bootstraps, LG substitution model) (46). Trees were visualized using iTol (47) and local synteny using Genespy (48). A reference species tree was built using RAxML-ng (LG substitution model, 200 bootstraps) based on a MAFFT-linsi alignment of IF-2a proteins (identified as hits against the TIGR00491 HMM model). This tree recapitulates previously inferred relationships among the Methanobacteriales (49).

To establish whether branching patterns might be confounded by gene conversion, we searched an alignment of all Methanobacteriales histones for signals of recombination/gene conversion using GARD (50), PHIpack (51), and RDP4 (52). GARD and PHIpack were run with standard settings using 1,000 permutations and a window size of 100 bp for PHIpack. RDP4 was also run with default settings using all available methods (RDP, GENECONV, Chimaera, MaxChi, 3Seq, BootScan, and SiScan). BootScan and SiScan primary scans were included. A window size of 30 was used for RDP. For MaxChi and Chimaera, the number of variable sites per window was 70 and 60, respectively. Not a single gene conversion event was supported by GARD or PHIpack or the RDP4 consensus. Tentative events called by individual methods as part of the RDP4 pipeline were investigated manually and confirmed not to affect phylogenetic inference. Overall, we found very little support for the hypothesis that gene conversion has played a major role during evolution of histones in this clade, consistent with the observed clustering of paralogs in a manner that recapitulates species phylogeny.

Diversity at a given residue (column in the alignment) and for a given group (e.g., archaea with multiple histone paralogs) was calculated using the Shannon diversity index (*H*). Subsequently, we computed diversity ratios for two groups (A and B) as

$$\text{Shannon diversity ratio} = \frac{\text{Shannon diversity index for A}}{\text{Shannon diversity index for B}}.$$

Similarity between histone groups in terms of the types of amino acids found at a given residue was calculated using the Jaccard index formula.

Histone Expression Levels for Different Species. For *M. smithii*, *T. kodakarensis*, and *Thermococcus onnurineus*, histone mRNA levels in exponential phase were obtained from NCBI's Gene Expression Omnibus (GEO) and primary publications. The relative expression of histones in *T. kodakarensis* (53) and *T. onnurineus* (GSE85760) (54) is plotted as base mean and normalized mRNA, respectively, in *SI Appendix*, Fig. S4. For *M. smithii*, we used the median value of histone expression across all replicates and conditions for strain MsmPS as determined in ref. 55. Expression levels for *Thermococcus litoralis*, *Methanothermobacter thermoautotrophicus*, *Methanothermobacter marburgensis*, *Ferroglobus placidus*, *Archaeoglobus fulgidus*, and *Archaeoglobus profundus* were taken from comparative quantitative proteomics data reported in ref. 56. We did not include archaea from the latter study where, due to high sequence identity among paralogs, intensities could not be uniquely assigned to a single paralog (*Methanococcus jannaschii*, *Pyrococcus furiosus*).

Downloaded from https://www.pnas.org by Imperial College London Library on June 20, 2022 from IP address 193.60.222.2.

***M. stadtmanae* Culture, qRT-PCR Analysis, and Proteomics.** *M. stadtmanae* DSM3091 was grown as previously described (57). Briefly, cultures were grown at 37 °C in 50 mL minimal medium under strict anaerobic conditions. Medium was reduced with Na_2S and cysteine (2 mM) and supplemented with 100 μg/mL ampicillin to prevent bacterial contamination. A 150-mM concentration of methanol and 1.5 atm H_2-CO_2 (80/20 [vol/vol]) served as carbon and energy source. Growth was monitored via turbidity at 600 nm (T_{600}) and stopped at exponential or stationary phase by short incubation on ice (15 min) and subsequent centrifugation of cultures (3,200 × *g* for 30 min at 4 °C). Resulting cell pellets were resuspended either in 500 μL 50 mM Tris containing RiboLock (Thermo Fisher Scientific) for RNA isolation or in 500 μL 50 mM triethylammonium bicarbonate buffer for proteomics until further processing.

For both isolation of RNA and proteins, *M. stadtmanae* cells were lysed in liquid nitrogen using a Mikro-Dismembrator S laboratory ball mill (Sartorius) for 3 min at 1,600 bpm.

For proteome analysis, cells were centrifuged after homogenization at 15,700 × *g* and 4 °C for 30 min and supernatant was used as cell-free protein extracts. RNA extraction and qRT-PCR assays were then performed as described earlier (58). mRNA expression levels of three biological replicates were calculated using the normalizing 2-ΔΔCt value. *Msp_16S* and *Msp_rpoB* were used as genes for normalization (59). Primers used are provided in *SI Appendix*, Table S4.

Cell-free protein extracts were run on a gel, low-molecular-weight section (<10 kDa) excised, and processed using a procedure adapted from ref. 60. Briefly, excised gel sections were further cut into cubes of ~2 × 2 mm and washed with 50 mM ammonium bicarbonate in 50% aqueous acetonitrile (ACN). Dehydration of gel sections was carried out with 100% ACN. Sections were then sequentially reduced and alkylated with 10 mM dithiothreitol and 55 mM iodoacetamide, respectively. Digestions were carried out by addition of 500 ng of trypsin per gel section, followed by incubation at 37 °C overnight. Gel digest supernatants were then dried completely by vacuum centrifugation. Following extraction of tryptic peptides from gel pieces, dried extracts were reconstituted in 1% aqueous ACN, 0.1% formic acid (FA). Desalting was performed using C18 reverse-phase solid-phase extraction spin tips (Glygen Corp.) following the manufacturer's recommendations and eluted tryptic peptides were then dried by vacuum centrifugation.

Desalted gel digests were solubilized in 20 μL of 0.1% aqueous trifluoroacetic acid (TFA) and clarified solutions transferred to autosampler vials for liquid chromatography–mass spectrometry analysis. Peptides were separated using an Ultimate 3000 RSLC nanoliquid chromatography system (Thermo Scientific) coupled to a LTQ Velos Orbitrap mass spectrometer (Thermo Scientific) via an EASY-Spray source. Sample aliquots (5.0 μL per injection) were loaded in technical duplicate onto a trapping column (Acclaim PepMap 100 C18, 100 μm × 2 cm) at 8 μL/min in 2% ACN, 0.1% TFA. Peptides were then eluted online to an analytical column (EASY-Spray PepMap C18, 75 μm × 25 cm) and peptides were separated using a stepped 90-min gradient: 4 to 25% buffer B for 60 min, 25 to 45% buffer B for 30 min. Buffer compositions were buffer A, 2% ACN, 0.1% FA; buffer B, 80% ACN, 0.1% FA. Eluted peptides were analyzed by the LTQ Velos operating in positive ion polarity using a data-dependent acquisition mode. Ions were selected for fragmentation from an initial MS1 survey scan at 15,000 resolution (at *m/z* 200), followed by Ion Trap collisional induced dissociation (CID) of the top 10 most abundant ions. MS1 and MS2 scan automatic gain control (AGC) targets were set to 1e6 and 1e4 for maximum injection times of 500 and 100 ms, respectively. A survey scan with *m/z* range of 350 to 1,500 was used, with a normalized collision energy (NCE) set to 35%, charge state rejection enabled for +1 ions, and a minimum threshold for triggering fragmentation of 500 counts.

The resulting data were processed using the MaxQuant software platform (v1.5.3.8), with database searches carried out by the in-built Andromeda search engine against the *M. stadtmanae* DSM3091 proteome as annotated in NCBI. A reverse decoy database search approach was used at a 1% false discovery rate (FDR) for peptide spectrum matches and protein identifications. Search parameters included maximum missed cleavages set to 2, fixed modification of cysteine carbamidomethylation and variable modifications of methionine oxidation, protein N-terminal and lysine acetylation, glutamine to pyro-glutamate conversion, and asparagine deamidation as well as lysine and arginine methylation.

Label-free quantification (LFQ) was enabled with a minimum ratio count of 2. The "match between runs" function was used with match and alignment time settings of 0.7 and 20 min, respectively.

Data Availability. All study data are included in this article and *SI Appendix*.

ACKNOWLEDGMENTS. We thank Samuel Bowerman, Remus Dame, and Karolin Luger for feedback on the manuscript; members of the Molecular Systems and DNA Replication groups for discussion; and Pierre Garcia for advice on phylogenetic tools. This work was funded by Medical Research Council core funding (to T.W.), a Medical Research Council studentship (to K.M.S.), a UK Research and Innovation Fellowship (to J.B.S.), a European Molecular Biology Organization Short-Term Fellowship 8472 (to A.H.), and Deutsche Forschungsgemeinschaft Grant SCHM1052/11-2 (to R.A.S.). S.G. acknowledges funding from the French National Agency for Research, Grant ArchEvol (ANR-16-CE02-0005-01). This project made use of time on UK Tier 2 Joint Academic Data Science Endeavour granted via the UK High-End Computing Consortium for Biomolecular Simulation supported by the Engineering and Physical Sciences Research Council (Grant EP/R029407/1).

1. S. Henikoff, M. M. Smith, Histone variants and epigenetics. *Cold Spring Harb. Perspect. Biol.* **7**, a019364 (2015).
2. S. Piquet *et al*., The histone chaperone FACT coordinates H2A.X-dependent signaling and repair of DNA damage. *Mol. Cell* **72**, 888–901.e7 (2018).
3. F. Mattiroli *et al*., Structure of histone-based chromatin in Archaea. *Science* **357**, 609–612 (2017).
4. K. Decanniere, A. M. Babu, K. Sandman, J. N. Reeve, U. Heinemann, Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon Methanothermus fervidus. *J. Mol. Biol.* **303**, 35–47 (2000).
5. H. S. Malik, S. Henikoff, Phylogenomics of the nucleosome. *Nat. Struct. Biol.* **10**, 882–891 (2003).
6. H. Maruyama *et al*., An alternative beads-on-a-string chromatin architecture in Thermococcus kodakarensis. *EMBO Rep.* **14**, 711–717 (2013).
7. N. Nalabothula *et al*., Archaeal nucleosome positioning in vivo and in vitro is directed by primary sequence motifs. *BMC Genomics* **14**, 391 (2013).
8. B. Henneman, C. van Emmerik, H. van Ingen, R. T. Dame, Structure and function of archaeal histones. *PLoS Genet.* **14**, e1007582 (2018).
9. M. Rojec, A. Hocher, K. M. Stevens, M. Merkenschlager, T. Warnecke, Chromatinization of *Escherichia coli* with archaeal histones. *eLife* **8**, 2407 (2019).
10. P. S. Adam, G. Borrel, C. Brochier-Armanet, S. Gribaldo, The growing tree of archaea: New perspectives on their diversity, evolution and ecology. *ISME J.* **11**, 2407–2425 (2017).
11. K. Sandman, R. A. Grayling, B. Dobrinski, R. Lurz, J. N. Reeve, Growth-phase-dependent synthesis of histones in the archaeon Methanothermus fervidus. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12624–12628 (1994).
12. K. A. Bailey, F. Marc, K. Sandman, J. N. Reeve, Both DNA and histone fold sequences contribute to archaeal nucleosome stability. *J. Biol. Chem.* **277**, 9293–9301 (2002).
13. I. Heinicke, J. Müller, M. Pittelkow, A. Klein, Mutational analysis of genes encoding chromatin proteins in the archaeon Methanococcus voltae indicates their involvement in the regulation of gene expression. *Mol. Genet. Genomics* **272**, 76–87 (2004).
14. L. Čuboňováa *et al*., An archaeal histone is required for transformation of Thermococcus kodakarensis. *J. Bacteriol.* **194**, 6864–6874 (2012).
15. L. Eme, A. Spang, J. Lombard, C. W. Stairs, T. J. G. Ettema, Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.* **15**, 711–723 (2017).
16. T. A. Williams, C. J. Cox, P. G. Foster, G. J. Szöllősi, T. M. Embley, Phylogenomics provides robust support for a two-domains tree of life. *Nat. Ecol. Evol.* **4**, 138–147 (2020).
17. C. F. Brunk, W. F. Martin, Archaeal histone contributions to the origin of eukaryotes. *Trends Microbiol.* **27**, 703–714 (2019).
18. F. Marc, K. Sandman, R. Lurz, J. N. Reeve, Archaeal histone tetramerization determines DNA affinity and the direction of DNA supercoiling. *J. Biol. Chem.* **277**, 30879–30886 (2002).
19. J. N. Reeve *et al*., Archaeal histones: Structures, stability and DNA binding. *Biochem. Soc. Trans.* **32**, 227–230 (2004).
20. P. B. Talbert, S. Henikoff, Histone variants–Ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.* **11**, 264–275 (2010).
21. J. M. Wolfe, G. P. Fournier, Horizontal gene transfer constrains the timing of methanogen evolution. *Nat. Ecol. Evol.* **2**, 897–903 (2018).
22. L. Eme, S. C. Sharpe, M. W. Brown, A. J. Roger, On the age of eukaryotes: Evaluating evidence from fossils and molecular clocks. *Cold Spring Harb. Perspect. Biol.* **6**, a016139 (2014).
23. H. C. Betts *et al*., Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat. Ecol. Evol.* **2**, 1556–1562 (2018).
24. I. Maze, K.-M. Noh, A. A. Soshnev, C. D. Allis, Every amino acid matters: Essential contributions of histone variants to mammalian development and disease. *Nat. Rev. Genet.* **15**, 259–271 (2014).
25. B. A. Nacev *et al*., The expanding landscape of 'oncohistone' mutations in human cancers. *Nature* **567**, 473–478 (2019).
26. D. Filipescu, S. Müller, G. Almouzni, Histone H3 variants and their chaperones during development and disease: Contributing to epigenetic control. *Annu. Rev. Cell Dev. Biol.* **30**, 615–646 (2014).
27. H. Wollmann *et al*., The histone H3 variant H3.3 regulates gene body DNA methylation in Arabidopsis thaliana. *Genome Biol.* **18**, 94 (2017).
28. H. Higashibata, M. A. Siddiqui, M. Takagi, T. Imanaka, S. Fujiwara, Surface histidine residue of archaeal histone affects DNA compaction and thermostability. *FEMS Microbiol. Lett.* **224**, 17–22 (2003).
29. D. J. Soares, F. Marc, J. N. Reeve, Conserved eukaryotic histone-fold residues substituted into an archaeal histone increase DNA affinity but reduce complex flexibility. *J. Bacteriol.* **185**, 3453–3457 (2003).

Downloaded from https://www.pnas.org by Imperial College London Library on June 20, 2022 from IP address 193.60.222.2.

30. D. J. Soares, K. Sandman, J. N. Reeve, Mutational analysis of archaeal histone-DNA interactions. *J. Mol. Biol.* **297**, 39–47 (2000).
31. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251–260 (1997).
32. J. Postberg, S. Forcob, W.-J. Chang, H. J. Lipps, The evolutionary history of histone H3 suggests a deep eukaryotic root of chromatin modifying mechanisms. *BMC Evol. Biol.* **10**, 259 (2010).
33. U. Friedrich-Jahn, J. Aigner, G. Längst, J. N. Reeve, H. Huber, Nanoarchaeal origin of histone H3? *J. Bacteriol.* **191**, 1092–1096 (2009).
34. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
35. A. L. Mitchell *et al*., InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **47**, D351–D360 (2019).
36. J. Schymkowitz *et al*, The FoldX web server: An online force field. *Nucleic Acids Res.* **33**, W382–W388 (2005).
37. B. R. Miller 3rd *et al*., MMPBSA.py: An efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* **8**, 3314–3321 (2012).
38. J. A. Maier *et al*., ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
39. R. L. Davidchack, R. Handel, M. V. Tretyakov, Langevin thermostat for rigid body dynamics. *J. Chem. Phys.* **130**, 234101 (2009).
40. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1998).
41. R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, R. C. Walker, Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* **9**, 3878–3888 (2013).
42. S. Le Grand, A. W. Götz, R. C. Walker, SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **184**, 374–380 (2013).
43. A. W. Götz *et al*., Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *J. Chem. Theory Comput.* **8**, 1542–1555 (2012).
44. U. Essmann *et al*., A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
45. B. Q. Minh *et al*., IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **46**, W537 (2020).
46. A. M. Kozlov, D. Darriba, T. Flouri, B. Morel, A. Stamatakis, RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
47. I. Letunic, P. Bork, Interactive tree of life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
48. P. S. Garcia, F. Jauffrit, C. Grangeasse, C. Brochier-Armanet, GeneSpy, a user-friendly and flexible genomic context visualizer. *Bioinformatics* **35**, 329–331 (2019).
49. M. Tokura, M. Ohkuma, T. Kudo, Molecular phylogeny of methanogens associated with flagellated protists in the gut and with the gut epithelium of termites. *FEMS Microbiol. Ecol.* **33**, 233–240 (2000).
50. S. L. Kosakovsky Pond, D. Posada, M. B. Gravenor, C. H. Woelk, S. D. W. Frost, GARD: A genetic algorithm for recombination detection. *Bioinformatics* **22**, 3096–3098 (2006).
51. T. C. Bruen, H. Philippe, D. Bryant, A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665–2681 (2006).
52. D. P. Martin, B. Murrell, M. Golden, A. Khoosal, B. Muhire, RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, vev003 (2015).
53. D. Jäger, K. U. Förstner, C. M. Sharma, T. J. Santangelo, J. N. Reeve, Primary transcriptome map of the hyperthermophilic archaeon Thermococcus kodakarensis. *BMC Genomics* **15**, 684 (2014).
54. S. Cho *et al*., Genome-wide primary transcriptome analysis of H 2 -producing archaeon Thermococcus onnurineus NA1. *Sci. Rep.* **7**, 1–12 (2017).
55. E. E. Hansen *et al*., Pan-genome of the dominant human gut-associated archaeon, Methanobrevibacter smithii, studied in twins. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4599–4606 (2011).
56. J. B. Müller *et al*., The proteome landscape of the kingdoms of life. *Nature* **582**, 592–596 (2020).
57. C. Bang *et al*., Effects of antimicrobial peptides on methanogenic archaea. *Antimicrob. Agents Chemother.* **56**, 4123–4130 (2012).
58. A. Buddeweg, K. Sharma, H. Urlaub, R. A. Schmitz, $sRNA_{41}$ affects ribosome binding sites within polycistronic mRNAs in Methanosarcina mazei Gö1. *Mol. Microbiol.* **107**, 595–609 (2018).
59. B. Dridi, M. Henry, A. El Khéchine, D. Raoult, M. Drancourt, High prevalence of Methanobrevibacter smithii and Methanosphaera stadtmanae detected in the human gut using an improved DNA detection protocol. *PLoS One* **4**, e7063 (2009).
60. A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen, M. Mann, In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860 (2006).

Downloaded from https://www.pnas.org by Imperial College London Library on June 20, 2022 from IP address 193.60.222.2.

GBE

Deep Conservation of Histone Variants in Thermococcales Archaea

Kathryn M. Stevens[1,2], Antoine Hocher[1,2], and Tobias Warnecke[1,2,*]

[1]Epigenetics Section, Medical Research Council London Institute of Medical Sciences, United Kingdom
[2]Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, United Kingdom

*Corresponding author: E-mail: tobias.warnecke@lms.mrc.ac.uk.

Accepted: 6 December 2021

Abstract

Histones are ubiquitous in eukaryotes where they assemble into nucleosomes, binding and wrapping DNA to form chromatin. One process to modify chromatin and regulate DNA accessibility is the replacement of histones in the nucleosome with paralogous variants. Histones are also present in archaea but whether and how histone variants contribute to the generation of different physiologically relevant chromatin states in these organisms remains largely unknown. Conservation of paralogs with distinct properties can provide *prima facie* evidence for defined functional roles. We recently revealed deep conservation of histone paralogs with different properties in the Methanobacteriales, but little is known experimentally about these histones. In contrast, the two histones of the model archaeon *Thermococcus kodakarensis*, HTkA and HTkB, have been examined in some depth, both in vitro and in vivo. HTkA and HTkB exhibit distinct DNA-binding behaviors and elicit unique transcriptional responses when deleted. Here, we consider the evolution of HTkA/B and their orthologs across the order Thermococcales. We find histones with signature HTkA- and HTkB-like properties to be present in almost all Thermococcales genomes. Phylogenetic analysis indicates the presence of one HTkA- and one HTkB-like histone in the ancestor of Thermococcales and long-term maintenance of these two paralogs throughout Thermococcales diversification. Our results support the notion that archaea and eukaryotes have convergently evolved histone variants that carry out distinct adaptive functions. Intriguingly, we also detect more highly diverged histone-fold proteins, related to those found in some bacteria, in several Thermococcales genomes. The functions of these bacteria-type histones remain unknown, but structural modeling suggests that they can form heterodimers with HTkA/B-like histones.

Key words: archaea, chromatin, histones, paralogs, Thermococcus.

Significance

Histones are key components of chromatin in eukaryotes and many archaea. In eukaryotes, histone variants exist, and play defined roles in cellular function and development. Some of these variants are highly conserved and can date back to the last common ancestor of eukaryotes. Archaea also frequently encode multiple sequence-divergent histones but whether these play distinct functional roles that are conserved through evolution remains largely unknown. Here, using phylogenetic tools, we establish the existence of histone variants with different properties in the order Thermococcales, which have been conserved for hundreds of millions of years. Our work provides additional evidence for histone variants in archaea and that these have evolved in parallel to eukaryotes to mediate flexible, adaptive chromatin states.

Introduction

The ability of eukaryotic cells to respond to environmental change and regulate transcription relies on dynamic control of DNA accessibility through chromatin alterations. This involves many different processes, including the addition/removal of histone modifications and the exchange of histone proteins for paralogous variants. Such variants can modify structural properties of the nucleosome or change how it

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/gbe/article/14/1/evab274/6459647 by Imperial College London Library user on 20 June 2022

interacts with its binding partners (Talbert and Henikoff 2010; Henikoff and Smith 2015; Martire and Banaszynski 2020). For example, macro-H2A has a large C-terminal domain and precipitates transcriptional repression (Bönisch and Hake 2012; Martire and Banaszynski 2020), whereas cenH3, a fast-evolving H3 variant, is specifically localized to centromeres and involved in chromosome segregation (Palmer et al. 1991; Talbert and Henikoff 2010). Importantly, significant functional changes can come from small differences in sequence. H3.3, for example, is deposited in a replication-independent manner in actively transcribed regions of the genome (Talbert and Henikoff 2010) and is important for mammalian development (Sitbon et al. 2020; Jang et al. 2015) but differs from its paralog H3.1 by only five amino acids.

Histones are not exclusive to eukaryotes. Archaeal histone proteins, first discovered in *Methanothermus fervidus* (Sandman et al. 1990; Starich et al. 1996), have since been identified in diverse archaeal lineages (Henneman et al. 2018; Hocher et al. 2021) and are often highly expressed (Hocher et al. 2021). Eukaryotic and archaeal histones share a conserved histone fold (HF) domain, form dimers and tetramers that are structurally very similar, and bind DNA nonspecifically, albeit with a preference for more bendable sequences (Bailey et al. 2000; Decanniere et al. 2000; Nalabothula et al. 2013; Mattiroli et al. 2017; Rojec et al. 2019). Unlike eukaryotic histones, almost all archaeal histones lack long terminal extensions ("tails") (Henneman et al. 2018). In at least some instances, archaeal histones have the capacity to form homo- as well as heterodimers and to assemble into long oligomeric structures on DNA (Mattiroli et al. 2017). These extended complexes can, in theory, consist of different histone paralogs, providing opportunities for chromatin state modulation through the exchange of histones with different properties (Stevens et al. 2020). In fact, many archaea encode two or more sequence-divergent histone paralogs, but whether these paralogs have defined functional roles akin to eukaryotic histone variants, and whether their expression and assembly change dynamically to mediate adaptive chromatin states, remains poorly understood.

What we do know from prior experimental work is that archaeal histone paralogs are more than mere copy number variants. The two histones of *M. fervidus* (HMfA, HMfB), for example, display differences in DNA-binding affinity (Bailey et al. 2002). Compared with HMfA, recombinant HMfB also induces more positive supercoiling upon binding to plasmid DNA and forms a more compact complex as inferred from gel-shift and tethered particle motion experiments (Sandman et al. 1994; Henneman et al. 2021). There are also differences between HMfA and HMfB in their relative expression during the growth cycle: in early exponential phase, HMfA is more highly expressed than HMfB, expression of which increases toward stationary phase to reach an almost equal ratio between the two (Sandman et al. 1994). The different properties of *M. fervidus* histones are consistent with the hypothesis that the two paralogs may have distinct functions, but whether the properties are physiologically relevant and affect organismal fitness remains to be addressed experimentally.

Recently, we considered this question using an evolutionary approach. We identified histone paralogs in the order Methanobacteriales that exhibit distinct structural properties and have been maintained over hundreds of millions of years (Stevens et al. 2020), indicative of the importance of each individual paralog for fitness. Structural modeling identified histone variants that prevent stable tetramerization and might act as *capstones* that limit further extension when incorporated into a histone oligomer, providing a potential pathway to dynamically alter chromatin state. Are the Methanobacteriales unique or are there other clades of archaea with histone paralogs that have been maintained over long periods of time? And do these paralogs also show conserved and distinct structural properties?

Here, we consider archaea in the order Thermococcales, which includes the model archaea *Pyrococcus furiosus* and *Pyrococcus abyssi* as well as *T. kodakarensis*, which has served as a model species for the in vivo study of archaeal histones. Thanks to the efforts of Santangelo and coworkers in particular, its two histones—HTkA (TK1413) and HTkB (TK2289)—are arguably the best characterized paralogs in vivo. Similar to HMfA and HMfB in *M. fervidus*, HTkA and HTkB can assemble into long oligomers both in vitro and in vivo (Maruyama et al. 2013; Mattiroli et al. 2017; Sanders et al. 2021; Bowerman et al. 2021). The two histones differ from one another at 11 out of 67 residues (84% identity) and have several distinct properties. HTkA is the more highly expressed paralog, at least in exponential phase, where it makes up 1.1% of the proteome compared with 0.66% for HTkB (Hocher et al. 2021). Together, they are abundant enough to coat the entire *T. kodakarensis* genome (Sanders et al. 2019). HTkB has been shown to bind to DNA more strongly than HTkA and to form more compact complexes, which show faster migration during agarose gel electrophoresis (Higashibata et al. 1999). Deletion of each histone individually results in overlapping but distinct perturbations of the transcriptome (Čuboňová et al. 2012; Sanders et al. 2021). Notably, HTkB-deficient cells exhibit reduced growth, possibly due to changes in the expression, not seen in strains lacking HTkA, of genes that encode translation factors and ribosomal proteins (Čuboňová et al. 2012). Deletion of *htkA* but not *htkB* leads to downregulation of hypothetical membrane proteins and prevents transformation of *T. kodakarensis*, suggesting HTkA alone plays a critical role in DNA uptake and/or integration (Čuboňová et al. 2012).

In this study, we show that histone paralogs with HTkA- and HTkB-like properties are present across the Thermococcales, including Thermococcus, Pyrococcus, and Palaeococcus spp. We use structural modeling to show that, in most Thermococcales,

Downloaded from https://academic.oup.com/gbe/article/14/1/evab274/6459647 by Imperial College London Library user on 20 June 2022

HTkB-like histones are predicted to exhibit stronger DNA binding than those with HTkA-like properties. Phylogenetic analysis reveals that HTkA-like histones share a common ancestor to the exclusion of HTkB-like histones and vice versa, suggesting that the last common ancestor of the Thermococcales already encoded an HTkA-like and an HTkB-like histone, each of which has been maintained throughout the diversification of this clade for (very) approximately 750 Myr (Wolfe and Fournier 2018). The long-term preservation of these two paralogs across the order Thermococcales supports the notion that HTkA/B in *T. kodakarensis* (and their orthologs in other Thermococcales) make unique contributions to genome function and fitness. These findings add further evidence that histone variants are widespread in archaea, evolving in parallel to those in eukaryotes.

Intriguingly, many Thermococcales archaea encode additional types of histone-fold proteins that are similar to histone-fold proteins found in some bacteria (Alva and Lupas 2019). One of these consists of an end-to-end duplication of the HF and is rarely found in archaea outside the Thermococcales. Their physiological roles remain unknown, but structural modeling suggest that they are able to heterodimerize with HTkA/B and might therefore further diversify histone-based chromatin states in these archaea.

Results and Discussion

Almost All Thermococcales Have Histone Paralogs with HTkA- and HTkB-Like Properties

To identify putative histone proteins across the Thermococcales, we scanned 61 predicted proteomes using HMM models and iterative jackhmmer searches (see Materials and Methods). Histones with a single histone-fold domain similar to archaeal HMf-like histones were found in all genomes (fig. 1c, supplementary table S1, Supplementary Material online). We also recovered putative histone-fold proteins similar to those found in some bacteria (Alva and Lupas 2019), which we discuss further below. A principal component analysis of the HMf-like archaeal histones based on their amino acid properties and isoelectric points (see Materials and Methods) suggests that histones can be assigned to one of two groups. One of these groups contains HTkA, the other HTkB (fig. 1a). This is consistent with a previous classification effort that also recovered two major groups of Thermococcales histones (Henneman 2019). Amino acid identities at several residues along the HF differ systematically between groups and are diagnostic of group membership. For example, tyrosine is always found at position 35 (Y35) in HTkA-like histones whereas HTkB-like histones have a positively charged lysine (K, 60 out of 62) or histidine (H, 2 out of 62). Similarly, glutamic acid at position 18 (E18) is present in 59 out of the 61 HTkA-like but none of the HTkB-like histones (fig. 1*b*).

For some of these residues, we know from prior in vitro studies—as well as from structural modeling—that amino acid identity can affect specific histone properties (Soares et al. 2000; Stevens et al. 2020). For example, substituting Y for K at residue 35 (the amino acids seen in HTkA- and HTkB-like histones, respectively), increases the stability of recombinant histone HFoB from *Methanobacterium formicicum* (Li et al. 2000). In addition, evidence from mass spectrometry indicates that K35 in HTkB from *T. kodakarensis* and *Thermococcus gammatolerans* is acetylated in vivo (Alpha-Bazin et al. 2021) although stoichiometry and functional significance of this modification remain to be determined. A tyrosine at the same position in HTkA removes the potential for acetylation. E18 in HMfB forms an intermolecular salt bridge with K53, which helps to stabilize the interaction between monomers in the histone dimer (Decanniere et al. 2000; Sandman et al. 2001). Mutating E18 to proline does not alter DNA binding (Soares et al. 2000), but loss of the intermonomer salt bridge may result in less rigid dimer structures. Finally, having leucine (L) or phenylalanine (F) at residue 46 has no obvious effect on DNA binding in HMfA/B, but the residue, located at the interface between dimers, is important for tetramer formation (Soares et al. 2000; Marc et al. 2002).

We considered how amino acid differences between HTkA- and HTkB-like histones affect two key aspects of the histone–DNA complex: DNA affinity and tetramerization strength, a proxy for tetramer stability. Using a structural modeling approach, we find that predicted DNA binding for HTkB-like paralogs is, in most cases, stronger than for HTkA-like paralogs (fig. 2, see Materials and Methods). This is in line with experimental findings that HTkB binds to DNA more tightly and forms a more compact complex with DNA than HTkA (Higashibata et al. 1999). In contrast, predicted tetramerization strength does not strongly discriminate HTkA- from HTkB-like histones (fig. 2). We also considered residues (K14, G17, K26, E30, E34, Q48, E58, K61, K65) that were previously suggested to be important for stacking interactions for either HTkA or HTkB, in the context of longer oligomers (Mattiroli et al. 2017; Henneman et al. 2018, 2021). None of these residues differ substantially between HTkA- and HTkB-like histones, with the exception of Q48 (HTkB) (Henneman et al. 2018). We note that Q48 is relatively uncommon in HTkB-like histones and that *T. kodakarensis* HTkB may therefore form more stable oligomeric complexes than the HTkB-like histones of most of its cousins.

HTkA- and HTkB-Like Histones Form Ancient Paralogous Groups

Almost all Thermococcales have both an HTkA-like and an HTkB-like histone (fig. 1c). This is consistent with (but not sufficient to demonstrate) ancient paralogy. To unravel the evolutionary history of Thermococcales histones, we used RaxML-NG (Kozlov et al. 2019) to build phylogenetic trees of all 123 HMf-like histones found across the 61 genomes in our analysis (see Materials and Methods). We find that

Downloaded from https://academic.oup.com/gbe/article/14/1/evab274/6459647 by Imperial College London Library user on 20 June 2022

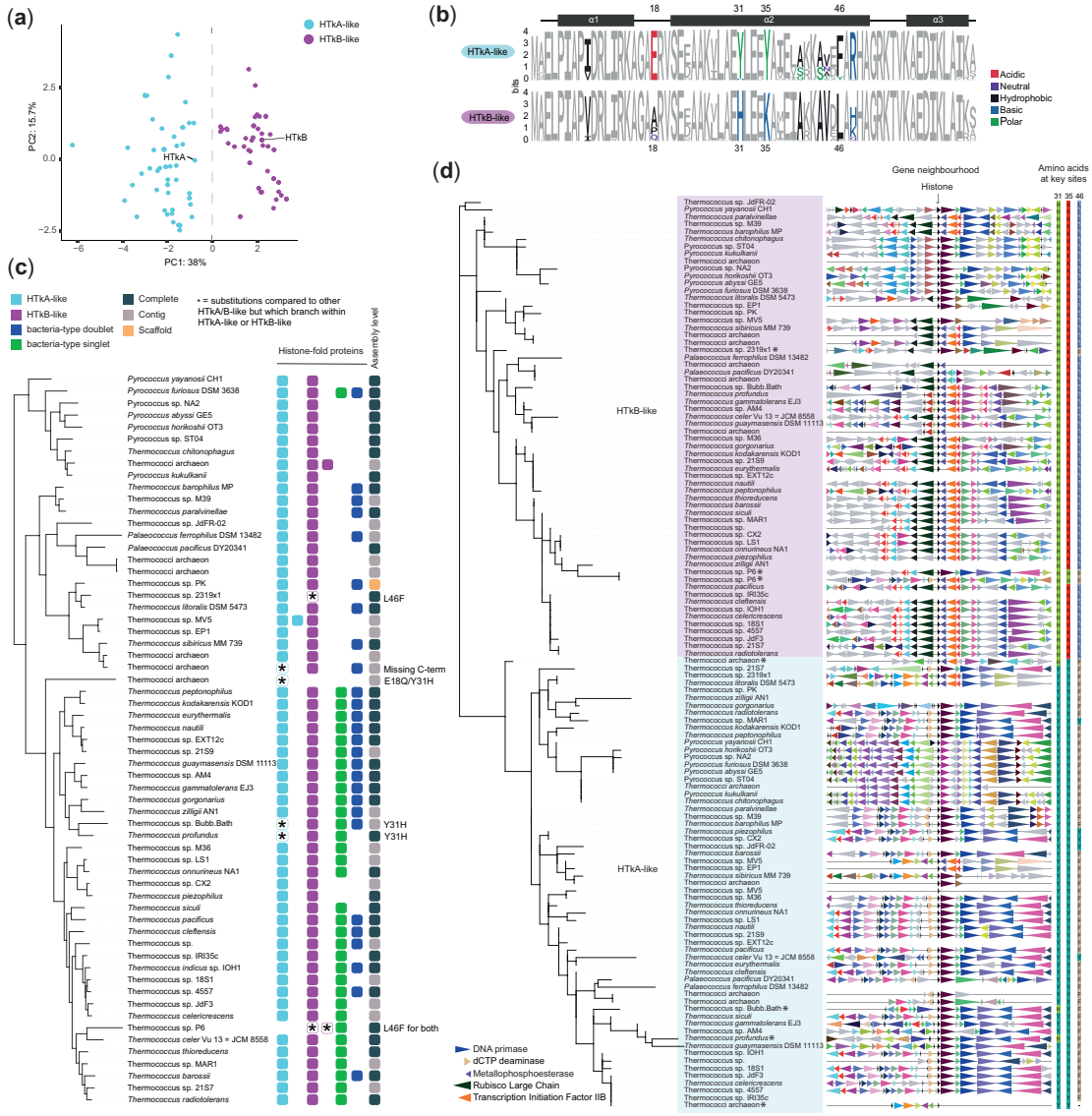


Fig. 1.—(*a*) Principal component analysis of Thermococcales HMf-like histones based on amino acid properties (AAStats, see Materials and Methods). Histones that cluster with either HTkA or HTkB along the first principal component are colored accordingly (HTkA-like histones in blue, HTkB-like in purple). (*b*) Sequence logos showing amino acid composition of HTkA- and HTkB-like histones across 61 Thermococcales. Amino acids that differ substantially between the two groups are colored. Positions are numbered relative to HMfB from *Methanothermus fervidus* to facilitate comparison with prior studies. (*c*) GTDB species tree for all Thermococcales in the data set indicating presence/absence of histones of a particular type in each genome. Each square represents one histone and is colored by histone type. (*d*) Protein-level phylogenetic tree of all HMf-like Thermococcales histones. Genes in the neighborhood are colored to indicate ortholog identity (see Materials and Methods). Note that, while the gene neighborhood is broadly conserved for HTkA- vis-à-vis HTkB-like orthologs, some HTkB-like genes (e.g., in *Thermococcus chitonophagus*) have a 3′ neighborhood normally found for HTkA. This is owing to a large genomic rearrangement event in which HTkB served as the breakpoint (not shown).

HTkA-like and HTkB-like paralogs neatly separate into two groups defined by their position on the tree (fig. 1*d*). This pattern of separation indicates that one HTkA- and one HTkB-like histone were present in the last common ancestor of Thermococcales. We detect only a small number of lineage-specific duplications and deletions and find no

Downloaded from https://academic.oup.com/gbe/article/14/1/evab274/6459647 by Imperial College London Library user on 20 June 2022

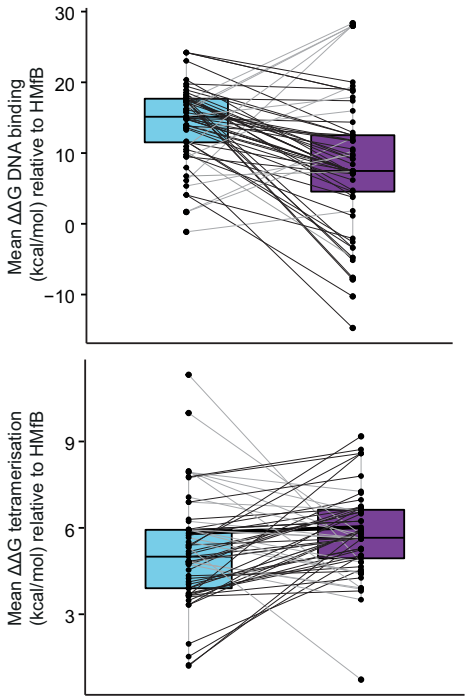


Fig. 2.—Predicted DNA-binding affinity (top) and tetramer stability (bottom) for HTkA/B-like paralogs. Lines connect paralogs from the same genome. Lines are black when DNA-binding affinity is stronger or tetramer interface energy weaker, respectively, for the HTkB-like paralog.

evidence of rampant horizontal gene transfer (fig. 1*c*). The observation that both paralogs have been maintained along divergent Thermococcales lineages strongly suggests that at least some of the amino acid differences between them are functionally important and under selection. Along with our recent report of ancient histone paralogs in the Methanobacteriales (Stevens et al. 2020), this finding provides further evidence that histone variants exist in archaea, evolving in parallel to those in eukaryotes. Note that, at present, we have no convincing evidence that histone paralogs in the Thermococcales and those found in the Methanobacteriales arose from the same ancient duplication event.

Some Thermococcales Encode Histone-Fold Proteins Similar to Those Found in Bacteria

Our survey also revealed that, alongside the HTkA/B-like histones, many Thermococcales genomes encode histone-fold proteins similar to those found in some bacteria (fig. 1*c*, supplementary table S1, Supplementary Material online), which harbor either a single or two (pseudodimeric) HF domains (Alva and Lupas 2019). We will refer to these as bacteria-type singlets and doublets, respectively. Both types are, on average, less well conserved than HTkA/B-like histones (fig. 3*a*) and their distribution across the Thermococcales is noticeably patchier (figs. 1*c* and 3*f* and *g*). Neither type is present in the closest sister clades (Methanofastidiosa, Theionarchaea).

The bacteria-type singlet is confined to a monophyletic group that comprises some (but not all) *Thermococcus* spp. (figs. 1*c* and 3*f*). It is present in all these species in the same highly conserved syntenic context, suggesting a single origin (fig. 3*f*). Following acquisition, the histone has been maintained along almost all lineages. There is only a single loss event (a clean deletion) in the branch leading to *Thermococcus piezophilus* (fig. 3*f*). Conserved synteny is also consistent with a single, earlier origin for the bacteria-type doublet, with multiple subsequent losses (figs. 1*c* and 3*g*). Based on the branching patterns and different syntenic context, the bacteria-type histones in *Pyrococcus furiosus* have likely been acquired secondarily from an extant or ancient *Thermococcus* species.

Bacteria-type doublets differ considerably in sequence from doublet histones previously described in haloarchaea and *Methanopyrus kandleri* (fig. 3*d*). Outside the Thermococcales, we only detect additional bacteria-type doublets in some (hyper)thermophilic Methanocaldococcus and Archaeoglobus species, but never their mesophilic relatives, suggestive of horizontal gene transfer in a high-temperature niche.

Bacteria-type histone-fold proteins have only recently been recognized and await functional characterization. The only functional data we have at present comes from transcriptome/proteome profiling. Consistent with lower sequence-level conservation (fig. 3*a*), the relative expression levels of these genes in *T. kodakarensis* (singlet: TK1040; doublet: TK0750) are lower than those of HTkA/B-like histones at both the transcript and protein levels (fig. 3*b*; Jäger et al. 2014; Sas-Chen et al. 2020). Together, they make up 0.37% of the measured exponential-phase proteome compared with 0.66% for HTkB and 1.1% for HTkA. TK1040 was previously identified in *T. kodakarensis* chromatin fractions (Maruyama et al. 2011), suggesting a (direct or indirect) association with DNA. However, the same study estimated that $<$1% of the amount of chromatin-associated proteins were attributable to TK1040. We therefore consider it unlikely that these histones are global organizers of DNA similar to HTkA/B-like histones, but might modulate chromatin state, either locally or globally, in response to environmental change. In both *P. furiosus* and *T. kodakarensis*, the bacterial-type doublets are under the control of the heat shock regulator Phr (encoded by PF1790 and TK2291, respectively) and upregulated upon Phr deletion, suggesting a potential role in response to heat shock in these archaea (Kanai et al. 2010; Keese et al. 2010). The *T. kodakarensis* doublet is also downregulated at lower temperatures, similar to HTkA/B (Hocher

Downloaded from https://academic.oup.com/gbe/article/14/1/evab274/6459647 by Imperial College London Library user on 20 June 2022

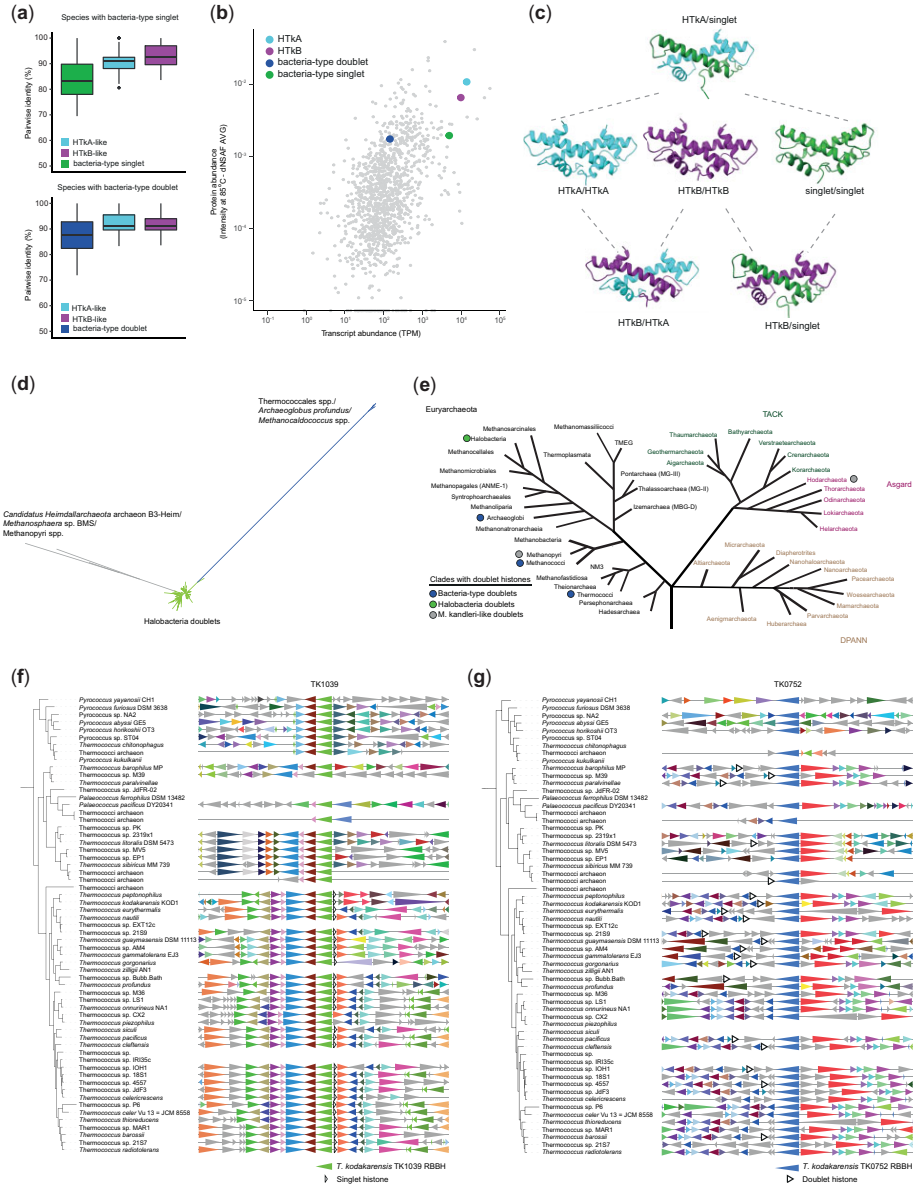


Fig. 3.—(*a*) Pairwise identity for histones in species with HTkA-like, HTkB-like, and bacteria-type singlet (top) or doublet histones (bottom) across all Thermococcales. (b) Protein and transcript abundance (see Materials and Methods) of genes in *T. kodakarensis*. TPM: transcripts per million. (c) AlphaFold-predicted homodimeric structures of TK1040 (green), HTkA (light blue), and HTkB (purple) and heterodimers of HTkA and HTkB (light blue/purple), HTkA and the bacteria-type singlet TK1040 (light blue/green), and HTkB and TK1040 (purple/green). (*d*) Protein-level tree for doublet histones (containing an end-to-end duplication) in archaea. Green branches contain proteins from Halobacteria, blue branches contain bacteria-type doublet histones in Thermococcales, *Methanocaldococcus* spp. and one *Archaeglobus* sp., gray branches show doublet histones including HMk from *Methanopyrus kandleri* (Slesarev et al. 1998; Fahrner et al. 2001). (*e*) Clades with doublet histones highlighted on an tree capturing archaeal diversity, adapted from Borrel et al. (2020) and Stevens et al. (2020) to include the clade for the species *Candidatus Heimdallarchaeota* archaeon B3-Heim, more recently suggested to be a member of the Hodarchaeota (Liu et al. 2021). Circles denote clades containing a doublet histone and are colored by the type of doublet (see *d*). (*f*, *g*) Thermococcales GTDB species tree showing syntenic regions for TK1039 (light green, in *f*) and TK0752 (blue, in *g*) commonly found located close to the bacteria-type singlet or doublet histones. Genes are colored to indicate homology (see fig. 1*d* and Materials and Methods).

Downloaded from https://academic.oup.com/gbe/article/14/1/evab274/6459647 by Imperial College London Library user on 20 June 2022

et al. 2021), further consistent with a role in temperature adaptation.

Can these HF proteins interact with HTkA/B-like histones? We used AlphaFold (Jumper et al. 2021) to predict the structure of combinations of HTkA, HTkB and the bacterial HF singlet from *T. kodakarensis*. Using this approach, all three are predicted to form homodimers and, as expected, HTkA and HTkB form a stable heterodimer. When a combination of either HTkA or HTkB and the bacterial singlet are used, Alphafold also predicts that these will form a heterodimer (fig. 3c). The presence of non-HMf-like HF proteins in Thermococcales genomes adds to the potential functional diversity of histone-based chromatin in these species and may dynamically alter DNA accessibility at different stages of cell growth or in response to environmental challenges. Further experimental investigation is required, however, before meaningful conclusions can be drawn in this regard, including whether they do interact, both structurally and functionally, with the HTkA/B-like histones.

Materials and Methods

Identification of Histones in Thermococcales Genomes

Protein sets, genomes, and GFF files for all available genomes of class Thermococci were downloaded from GenBank (https://www.ncbi.nlm.nih.gov/assembly) using taxid 183968 (accessed on May 27, 2021). Genomes not present in the GTDB tree (Parks et al. 2021) (https://gtdb.ecogenomic.org, accessed August 1, 2021, see below) were removed. Two species that were annotated as Thermococci in NCBI but branched outside the main group on the GTDB tree were removed from the analysis, leaving a final set of 61 genomes, all from the order Thermococcales. Protein sequences were predicted using Prodigal v2.6.3 (Hyatt et al. 2010) where not provided through GenBank. Histone proteins were extracted from the protein sets through HMM searches using HMMER v3.3.1 (hmmsearch–noali (Eddy 2011; Finn et al. 2011) using Pfam models CBFD_NFYB_HMF and DUF1931 (Finn et al. 2014) as well as a Jackhmmer searches using the singlet and doublet histones from bacteria as a seed (Alva and Lupas 2019). Bacteria-type histones hit in the initial search were used to build an HMM model and the Thermococcales protein set was re-searched using HMMER v3.3.1 as above (hmmsearch --noali). Some proteins incorrectly identified as histones at this stage were manually filtered out.

Classification of Thermococcales Histones into HTkA-Like and HTkB-Like Groups

HMf-like histones in Thermococcales downloaded from GenBank (accessed on May 27, 2021) (see above) were aligned using MAFFT (Katoh and Standley 2013) (–localpair --maxiterate 1000). Histones were clustered based on amino acid composition of their peptide sequences using AAStats from the R package Seqinr (Charif and Lobry 2007). Histones that clustered with HTkA and HTkB were assigned HTkA- or HTkB-like status, respectively (see fig. 1a). Twenty maximum likelihood phylogenetic trees were built using Raxml-NG (Kozlov et al. 2019) with the LG+G4 model of evolution as suggested by ModelTest-NG (Darriba et al. 2020). The unrooted best maximum likelihood tree is shown. All trees were plotted using iTOL (Letunic and Bork 2019). Orthologous genes in the genomic neighborhood of each histone were highlighted on the tree using Genespy as best reciprocal hits (Garcia et al. 2019). Orthologs in the histone neighborhood were identified by performing reciprocal best hits for each genome against *T. kodakarensis* using BLAST (Altschul et al. 1990), retaining those that have a similarity score of >40% and are within 20% length of one another (Rocha 2006). Note that the reciprocal best hit approach was only applied to identify putative orthologs in the neighborhood of histone genes. The histones themselves were identified using HMMER searches (see above) and subsequent analyses not restricted to reciprocal best hits of HTkA/B. To generate sequence logos, histones were aligned using MAFFT (Katoh and Standley 2013) (–localpair --maxiterate 1000) and visualized using ggseqlogo in R (Wagih 2017).

Predicted DNA Binding and Tetramerization

Predicted DNA-binding and interaction (tetramerization) strength between dimers for Thermococcales species was computed as in Stevens et al. (2020). In brief, sequences were aligned to HMfB and substitutions were mapped onto a tetrameric model of HMfB (extracted from PDB structure 5t5k) using FoldX (Schymkowitz et al. 2005) to generate models of homotetramers with DNA. Structures were then energy-minimized using AmberTools (Maier et al. 2015) and binding affinity was calculated using an MMPBSA approach with the ff14SB forcefield (Miller et al. 2012). $\Delta\Delta G$ was calculated relative to HMfB. The mean value for five replicates is shown for each model.

Species Tree

The archaeal species tree was downloaded from GTDB (https://gtdb.ecogenomic.org) on August 1, 2021.

Expression Data

Expression data for *T. kodakarensis* were obtained from primary sources and NCBI's Gene Expression Omnibus (GEO) (Barrett et al. 2013). Proteomics data from (Sas-Chen et al. 2020) was processed as in Hocher et al. (2021). Protein abundance at 85°C is shown. Transcript abundance data were obtained from Jäger et al. (2014) and were shown as transcripts per million.

Downloaded from https://academic.oup.com/gbe/article/14/1/evab274/6459647 by Imperial College London Library user on 20 June 2022

Bacteria-Type Singlet HF and HTkA/B Structure Prediction

The AlphaFold v2.0 (Jumper et al. 2021) collab notebook (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb, accessed August 21, 2018) was used to predict the structure of HTkA, HTkB, and the bacterial singlet HF protein (TK1040) as homodimers and all heterodimer combinations. The MSA method used was jackhmmer, and models were ranked by PTMscore. The top model is shown for all homodimers and heterodimers. Images shown were generated using UCSF, ChimeraX developed by the Resource for Biocomputing, Visualization and Informatics at the University of California, San Francisco, with support from National Institutes of Health (R01-GM129325) and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases (Pettersen et al. 2021).

Identification of Doublet Histones in Archaea

The genomes, protein sets, and GFF files for a balanced set of archaea species were downloaded on May 21, 2021 from GenBank (https://www.ncbi.nlm.nih.gov/assembly) using taxid 2157 and processed as above (see Identification of Histones in Thermococcales Genomes). Doublet histones, including doublets from Halobacteria (Dulmage et al. 2015) and HMk in *Methanopyrus kandleri* (Fahrner et al. 2001; Slesarev et al. 1998), were identified by their position on a coding sequence-level tree of archaeal histones, and by length. This larger tree was built from 20 ML trees using Raxml-NG (Kozlov et al. 2019). Trees for the doublet histones were then built as previously described (see Classification of Thermococcales Histones into HTkA-Like and HTkB-Like Groups) and all figures plotted using iTOL (Letunic and Bork 2019). Annotation for the orthologous genes in the genomic neighborhood was generated using GeneSpy (Garcia et al. 2019). Clades with doublet histones are annotated on a tree of archaea adapted from (Borrel et al. 2020) to include Hodarchaeota (Liu et al. 2021).

Percentage Identity of HTkA-/HTkB-Like and Bacteria-Type Histones

HTkA-like, HTkB-like and bacteria-type histones were aligned using MAFFT (Katoh and Standley 2013) (–localpair --maxiterate 1000) separately for species containing either bacteria-type doublet or bacteria-type singlet histones. For each histone type, pairwise sequence identity was calculated using seqidentity from the R package bio3d (Grant et al. 2006).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was funded through UKRI Medical Research Council funding (MC-A658-5TY40) to T.W.

Data Availability

All data underlying were derived from sources in the public domain: GenBank (https://www.ncbi.nlm.nih.gov/assembly), GTDB (https://gtdb.ecogenomic.org) and data from original publications cited throughout the manuscript.

Literature Cited

Alpha-Bazin B, et al. 2021. Lysine-specific acetylated proteome from the archaeon *Thermococcus gammatolerans* reveals the presence of acetylated histones. J Proteomics. 232:104044.
Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.
Alva V, Lupas AN. 2019. Histones predate the split between bacteria and archaea. Bioinformatics 35(14):2349–2353.
Bailey KA, Marc F, Sandman K, Reeve JN. 2002. Both DNA and histone fold sequences contribute to archaeal nucleosome stability. J Biol Chem. 277(11):9293–9301.
Bailey KA, Pereira SL, Widom J, Reeve JN. 2000. Archaeal histone selection of nucleosome positioning sequences and the procaryotic origin of histone-dependent genome evolution. J Mol Biol. 303(1):25–34.
Barrett T, et al. 2013. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 41:991–995.
Bönisch C, Hake SB. 2012. Histone H2A variants in nucleosomes and chromatin: more or less stable? Nucleic Acids Res. 40(21):10719–10741.
Borrel G, Brugère JF, Gribaldo S, Schmitz RA, Moissl-Eichinger C. 2020. The host-associated archaeome. Nat Rev Microbiol. 18(11):622–636.
Bowerman S, Wereszczynski J, Luger K. 2021. Archaeal chromatin 'slinkies' are inherently dynamic complexes with deflected DNA wrapping pathways. Elife 10:e65587.
Charif D, Lobry JR. 2007. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis.In: Bastolla U, Porto, M, Roman, HE, Vendruscolo, M, editors. Structural approaches to sequence evolution: molecules, networks, populations. New York: Springer Verlag. p. 207–232.
Čuboňová L, et al. 2012. An archaeal histone is required for transformation of *Thermococcus kodakarensis*. J Bacteriol. 194(24):6864–6874.
Darriba D, et al. 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol Biol Evol. 37(1):291–294.
Decanniere K, Babu AM, Sandman K, Reeve JN, Heinemann U. 2000. Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon *Methanothermus fervidus*. J Mol Biol. 303(1):35–47.
Dulmage KA, Todor H, Schmid AK. 2015. Growth-phase-specific modulation of cell morphology and gene expression by an archaeal histone protein. MBio 6(5):e00649–15.
Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol. 7(10):e1002195.
Fahrner RL, Cascio D, Lake JA, Slesarev A. 2001. An ancestral nuclear protein assembly: crystal structure of the Methanopyrus kandleri histone. Protein Sci. 10(10):2002–2007.
Finn RD, et al. 2014. Pfam: the protein families database. Nucleic Acids Res. 42:222–230.
Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39:29–37.

Downloaded from https://academic.oup.com/gbe/article/14/1/evab274/6459647 by Imperial College London Library user on 20 June 2022

Garcia PS, Jauffrit F, Grangeasse C, Brochier-Armanet C. 2019. GeneSpy, a user-friendly and flexible genomic context visualizer. Bioinformatics 35(2):329–331.
Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. 2006. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics 22(21):2695–2696.
Henikoff S, Smith MM. 2015. Histone variants and epigenetics. Cold Spring Harb Perspect Biol. 7(1):a019364.
Henneman B. 2019. Histone-DNA assemblies in archaea: shaping the genome on the edge of life. Leiden, Netherlands: Leiden University.
Henneman B, et al. 2021. Mechanical and structural properties of archaeal hypernucleosomes. Nucleic Acids Res. 49(8):4338–4349.
Henneman B, van Emmerik C, van Ingen H, Dame RT. 2018. Structure and function of archaeal histones. PLoS Genet. 14(9):e1007582.
Higashibata H, Fujiwara S, Takagi M, Imanaka T. 1999. Analysis of DNA compaction profile and intracellular contents of archaeal histones from *Pyrococcus kodakaraensis* KOD1. Biochem Biophys Res Commun. 258(2):416–424.
Hocher A, et al. 2021. Growth temperature is the principal driver of chromatinization in archaea. BioRxiv. doi:10.1101/2021.07.08.451601.
Hyatt D, et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.
Jäger D, Förstner KU, Sharma CM, Santangelo TJ, Reeve JN. 2014. Primary transcriptome map of the hyperthermophilic archaeon Thermococcus kodakarensis. BMC Genomics 15:684.
Jang C, Shibata Y, Starmer J, Yee D, Magnuson T. 2015. Histone H3.3 maintains genome integrity during mammalian development. Genes Dev. 29(13):1377–1392.
Jumper J, et al. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596(7873):583–589.
Kanai T, Takedomi S, Fujiwara S, Atomi H, Imanaka T. 2010. Identification of the Phr-dependent heat shock regulon in the hyperthermophilic archaeon, *Thermococcus kodakaraensis*. J Biochem. 147(3):361–370.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.
Keese AM, Schut GJ, Ouhammouch M, Adams MWW, Thomm M. 2010. Genome-wide identification of targets for the archaeal heat shock regulator Phr by cell-free transcription of genomic DNA. J Bacteriol. 192(5):1292–1298.
Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35(21):4453–4455.
Letunic I, Bork P. 2019. Interactive Tree of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 47:256–259.
Li WT, Shriver JW, Reeve JN. 2000. Mutational analysis of differences in thermostability between histones from mesophilic and hyperthermophilic archaea. J Bacteriol. 182(3):812–817.
Liu Y, et al. 2021. Expanded diversity of Asgard archaea and their relationships with eukaryotes. Nature 593(7860):553–557.
Maier JA, et al. 2015. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput. 11(8):3696–3713.
Marc F, Sandman K, Lurz R, Reeve JN. 2002. Archaeal histone tetramerization determines DNA affinity and the direction of DNA supercoiling. J Biol Chem. 277(34):30879–30886.
Martire S, Banaszynski LA. 2020. The roles of histone variants in fine-tuning chromatin organization and function. Nat Rev Mol Cell Biol. 21(9):522–541.
Maruyama H, et al. 2013. An alternative beads-on-a-string chromatin architecture in Thermococcus kodakarensis. EMBO Rep. 14(8):711–717.
Maruyama H, et al. 2011. Histone and TK0471/TrmBL2 form a novel heterogeneous genome architecture in the hyperthermophilic archaeon *Thermococcus kodakarensis*. Mol Biol Cell. 22(3):386–398.
Mattiroli F, et al. 2017. Structure of histone-based chromatin in Archaea. Science. 357(6351):609–612.
Miller BR, et al. 2012. MMPBSA.py: an efficient program for end-state free energy calculations. J Chem Theory Comput. 8(9):3314–3321.
Nalabothula N, et al. 2013. Archaeal nucleosome positioning in vivo and in vitro is directed by primary sequence motifs. BMC Genomics 14:391.
Palmer DK, O'Day K, Trong HL, Charbonneau H, Margolis RL. 1991. Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc Natl Acad Sci USA. 88(9):3734–3738.
Parks DH. 2021. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. Advance Access published Sep 14, 2021, doi:10.1093/nar/gkab776.
Pettersen EF, et al. 2021. UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci. 30(1):70–82.
Rocha EPC. 2006. Inference and analysis of the relative stability of bacterial chromosomes. Mol Biol Evol. 23(3):513–522.
Rojec M, Hocher A, Stevens KM, Merkenschlager M, Warnecke T. 2019. Chromatinization of Escherichia coli with archaeal histones. Elife 8:e49038.
Sanders TJ, et al. 2021. Extended archaeal histone-based chromatin structure regulates global gene expression in *Thermococcus kodakarensis*. Front Microbiol. 12:681150.
Sanders TJ, et al. 2019. TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin. Mol Microbiol. 111(3):784–797.
Sandman K, Grayling RA, Dobrinski B, Lurz R, Reeve JN. 1994. Growth-phase-dependent synthesis of histones in the archaeon Methanothermus fervidus. Proc Natl Acad Sci USA. 91(26):12624–12628.
Sandman K, Krzycki JA, Dobrinski B, Lurz R, Reeve JN. 1990. HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. Proc Natl Acad Sci USA. 87(15):5788–5791.
Sandman K, Soares D, Reeve JN. 2001. Molecular components of the archaeal nucleosome. Biochimie 83(2):277–281.
Sas-Chen A, et al. 2020. Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. Nature 583(7817):638–643.
Schymkowitz J, et al. 2005. The FoldX web server: an online force field. Nucleic Acids Res. 33(Web Server issue):W382–W388.
Sitbon D, Boyarchuk E, Dingli F, Loew D, Almouzni G. 2020. Histone variant H3.3 residue S31 is essential for Xenopus gastrulation regardless of the deposition pathway. Nat Commun. 11(1):1256.
Slesarev AI, Belova GI, Kozyavkin SA, Lake JA. 1998. Evidence for an early prokaryotic origin of histones H2A and H4 prior to the emergence of eukaryotes. Nucleic Acids Res. 26(2):427–430.
Soares DJ, Sandman K, Reeve JN. 2000. Mutational analysis of archaeal histone-DNA interactions. J Mol Biol. 297(1):39–47.
Starich MR, Sandman K, Reeve JN, Summers MF. 1996. NMR structure of HMfB from the hyperthermophile, Methanothermus fervidus, confirms that this archaeal protein is a histone. J Mol Biol. 255(1):187–203.
Stevens KM, et al. 2020. Histone variants in archaea and the evolution of combinatorial chromatin complexity. Proc Natl Acad Sci USA. 117(52):33384–33395.
Talbert PB, Henikoff S. 2010. Histone variants—ancient wrap artists of the epigenome. Nat Rev Mol Cell Biol. 11(4):264–275.
Wagih O. 2017. ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics 33(22):3645–3647.
Wolfe JM, Fournier GP. 2018. Horizontal gene transfer constrains the timing of methanogen evolution. Nat Ecol Evol. 2(5):897–903.

Associate editor: John Archibald

Downloaded from https://academic.oup.com/gbe/article/14/1/evab274/6459647 by Imperial College London Library user on 20 June 2022

ARTICLE IN PRESS

Seminars in Cell and Developmental Biology xxx (xxxx) xxx

Contents lists available at ScienceDirect

Seminars in Cell and Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Histone variants in archaea – An undiscovered country

Kathryn M. Stevens [a,b], Tobias Warnecke [a,b,*]

[a] Medical Research Council London Institute of Medical Sciences, London, United Kingdom
[b] Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London, United Kingdom

ARTICLE INFO

Keywords:
Archaea
Histone variants
Chromatin

ABSTRACT

Exchanging core histones in the nucleosome for paralogous variants can have important functional ramifications. Many of these variants, and their physiological roles, have been characterized in exquisite detail in model eukaryotes, including humans. In comparison, our knowledge of histone biology in archaea remains rudimentary. This is true in particular for our knowledge of histone variants. Many archaea encode several histone genes that differ in sequence, but do these paralogs make distinct, adaptive contributions to genome organization and regulation in a manner comparable to eukaryotes? Below, we review what we know about histone variants in archaea at the level of structure, regulation, and evolution. In all areas, our knowledge pales when compared to the wealth of insight that has been gathered for eukaryotes. Recent findings, however, provide tantalizing glimpses into a rich and largely undiscovered country that is at times familiar and eukaryote-like and at times strange and uniquely archaeal. We sketch a preliminary roadmap for further exploration of this country; an undertaking that may ultimately shed light not only on chromatin biology in archaea but also on the origin of histone-based chromatin in eukaryotes.

1. Introduction

To regulate genome accessibility, different forms of life employ a wide range of tools. In eukaryotes, transcription factors, polymerases, repair enzymes and other DNA-binding proteins have to contend with the fact that most genomic DNA has been assembled into nucleosomes. One way to alter genome accessibility is to exchange bulk histones in these nucleosomes for paralogous variants that – either directly or indirectly – render chromatin more or less accessible. In eukaryotes, there is a large catalogue of these variants, which play specific roles in, for example, responding to DNA damage, establishing centromere identity and altering gene expression in response to temperature change [1–3]. These variants are encoded outside the main clusters of replicative histone genes. Their deposition is not coupled to the cell cycle and occurs at specific locations, regulated by dedicated chaperones and remodellers (reviewed in [4–7]).

Histone-fold proteins are also widespread in archaea [8–10] (Fig. 1f). In some species, they are relatively lowly expressed [10] and appear to behave more like transcription factors. HpyA, for example, the only histone found in the halophilic archaeon *Halobacterium salinarum,* binds to < 60 locations along the genome and regulates the expression of a small set of genes involved in the adaptation to low salt conditions [11]. In other species, histones mediate genome-wide packaging reminiscent of what is happening in eukaryotes. This includes members of the orders Thermococcales and Methanobacteriales where histones are amongst the most abundant proteins in the cell [10,12–14].

Many archaeal genomes encode multiple histone genes [8], which can differ significantly at the amino acid level (Fig. 1b). For example, the most dissimilar histone sequences in *Methanosphaera stadtmanae*, a methanogen found in the human gut, have only 41% identity. Can these homologs be considered histone variants in the sense we know them from eukaryotes? That is, do they play distinct (though perhaps overlapping) functional roles in the cell and make non-redundant contributions to genome function and fitness? Or are the amino acid differences that distinguish these variants largely immaterial - tolerated, rather than adaptive?

Below, we will review the evidence that histones encoded in the same archaeal genome do different things. We consider three lines of evidence: structural properties, regulation, and evolution. We will argue that histone variants with distinct, adaptive functions exist in archaea as they do in eukaryotes, but that we have barely scratched the surface when it comes to elucidating their physiological roles, an effort that will ultimately help us not only to understand chromatin biology in archaea but also the evolution of histone-based chromatin in eukaryotes.

* Corresponding author at: Medical Research Council London Institute of Medical Sciences, London, United Kingdom.
E-mail address: tobias.warnecke@lms.mrc.ac.uk (T. Warnecke).

https://doi.org/10.1016/j.semcdb.2022.02.016
Received 12 November 2021; Received in revised form 20 February 2022; Accepted 20 February 2022
1084-9521/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Please cite this article as: Kathryn M. Stevens, Tobias Warnecke, *Seminars in Cell and Developmental Biology*, https://doi.org/10.1016/j.semcdb.2022.02.016

2. Structure

2.1. Similar building blocks, different assembly

Regarding both secondary and tertiary structure, the core histone fold in archaea and eukaryotes is remarkably similar. Monomers assemble into dimers in the familiar handshake configuration (Fig. 1c, d), mediated through hydrophobic contacts between the anti-parallel α2 helices of the two monomers and through hydrogen bonding between neighbouring loop regions [15,16]. Contacts with DNA are also made in a similar manner, principally via residues in the loop regions [17] (Fig. 1c,d).

In eukaryotes, side chain residues at the dimer interface form complementary shapes. As a consequence, eukaryotic histones form obligate heterodimers: H3-H4 and H2A-H2B [18]. H4 does not form dimers with another H4 molecule, H3 does not pair with H3, and so forth. Dimerization of (at least some) archaeal histones is comparatively more flexible as it allows formation of homo- as well as heterodimers. Direct experimental confirmation of this *in vivo* comes from *Methanothermus fervidus*, which has served as the model system for much of the early

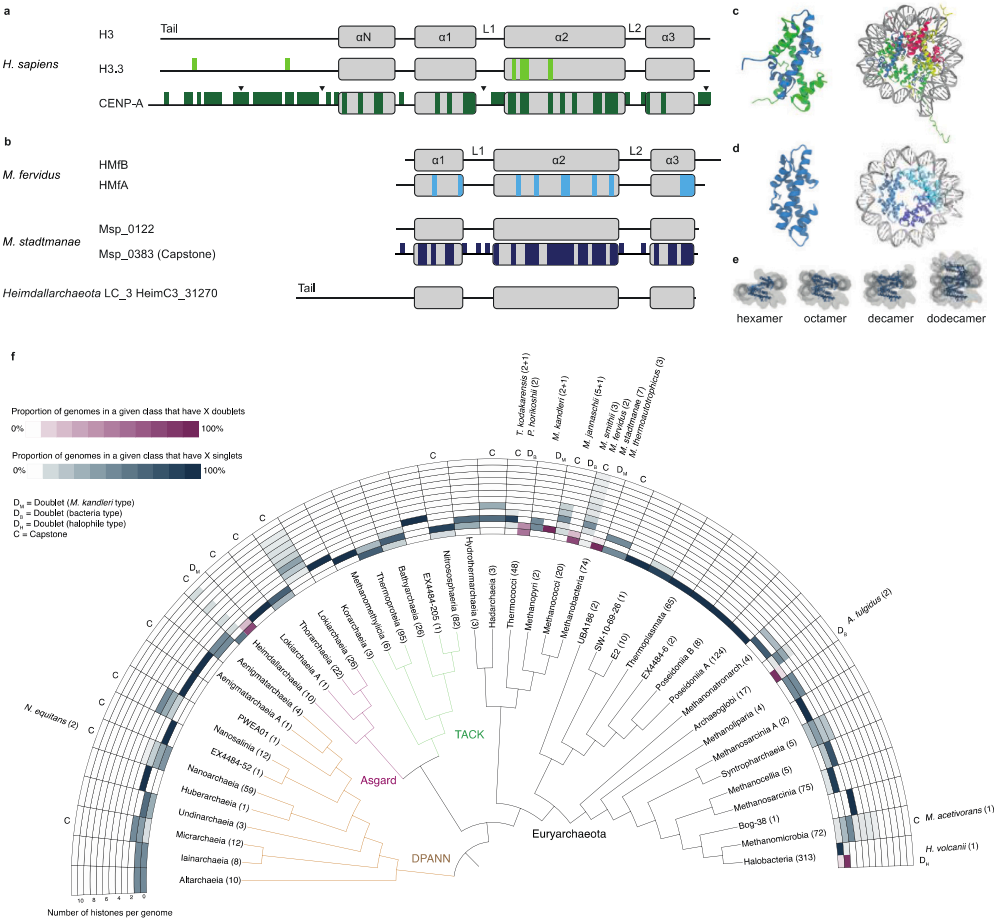


Fig. 1. Overview of archaeal and eukaryotic histone variants. a. Comparison of human histone variants H3.3 and CENP-A (cenH3) to replicative H3. Sequences were downloaded from HistoneDB 2.0 [68]. Sites which are not identical to replicative H3 are highlighted in coloured blocks on the secondary structure for each variant. Insertions are marked with an arrow. Alpha helices (αN-α3) and loop regions (L1/L2) are annotated. b. Comparison of histones from three archaea that each encode more than one histone. Top: for *Methanothermus fervidus*, residues which differ in HMfA relative to HMfB are shown on the secondary structure. Centre: in *Methanosphaera stadtmanae*, residues which differ in Msp_0383 relative to Msp_0122 are highlighted. Only two of the seven histones in this species are shown. Bottom: an example of a tailed histone from *Candidatus Heimdallarchaeota* archaeon LC_3. c. Structure of the H3-H4 dimer and the eukaryotic nucleosome (PDB: 1AOI [18]; red = H2B, yellow = H2A, green = H3, blue = H4). d. Structure of the HMfB dimer and hexamer (PDB: 5T5K [21]). e. Model of higher-order archaeal histone oligomers, showing sequential addition of HMfB histone dimers from hexamer to dodecamer. Structures in c. and d. were rendered using VMD [69]. f. Cladogram of archaea illustrating the distribution of histone proteins across different archaeal classes. The number of genomes available for each class is given in parentheses following the class name. The number of histones is shown separately for proteins containing either a single (singlet, blue) or end-to-end-duplicated (doublet, red) histone-fold domain. Clades containing putative capstone histones (see text and [66]) are annotated "C". Doublet histones are annotated by type (see [66]); bacteria-type doublet (D_B), *M. kandleri*-type doublet (D_M), Halobacteria-type doublet (D_H). Species mentioned in the main text are annotated with the number of singlet (+ doublet) histones that are present in their genome. The cladogram is based on GTDB (https://gtdb.ecogenomic.org), with the tree rooted at the split between DPANN and Euryarchaeota/Asgard/TACK. Information on histone presence/absence and number was obtained from [10]. Only information for species present in GTDB is shown, supplemented with additional Asgard species from Genbank (taxid 1935183) and [76]. Seven archaeal classes present in GTDB were not included in the histone survey dataset [10] and are therefore not included in the tree. For reference, Lokiarchaeia_A contains Odinarchaeota species. Some counts of species with no histones may be as a result of missing information/incomplete genome assemblies rather than indicating a definitive lack of histones.

biochemical work on archaeal histones by John Reeve and colleagues [19]. Homodimerization has also been extensively demonstrated *in vitro* using recombinant histones from *M. fervidus* [16,19–21], *Thermococcus kodakarensis* [22], *Pyrococcus horikoshii* OT3 [23] and *Methanobacterium formicicum* [24,25]. In addition, *in vitro* unfolding transitions indicate that homodimers are also formed in Pyrococcus GB-3a [25], *Archaeoglobus fulgidus* and *Methanocaldococcus jannaschii* [26]. Finally, there are several archaea that encode only a single histone gene, making homodimerization a necessary requirement for dimer formation [27]. Taken together, this suggests that the capacity for promiscuous homo- and heterodimerization is, if not a universal property of archaeal histones, then certainly widespread.

2.2. Variants that restrict promiscuous dimerization?

Some archaeal histone variants might break this promiscuity. Most members of the (confusingly named) Halobacteria as well as some distantly related species such as *Methanopyrus kandleri* [28,29] carry histone doublets, where two histone-fold domains are encoded back-to-back, connected through a flexible linker peptide. This physical tether should strongly favour dimerization between the fused histone-fold domains, which have begun to diverge in sequence and are thought to act – effectively – as obligate heterodimers [although structural modelling suggests that, if the tether were to be cut, both N- and C-terminal halves could still form homodimers (K. Stevens, unpublished results)]. Heterodimerization here is enforced by physical linkage, not structural incompatibility as observed for eukaryotic histones.

Outside of the forced marriage of histone doublets, we are not aware of any examples of obligate heterodimerisation in archaea. There are, however, at least two documented instances where heterodimerization appears to be required for DNA binding. One of the three histones of *Methanothermobacter thermoautotrophicus* OSU (HMtB) has lost its ability to bind DNA as a homodimer but can do so when assembled into a heterodimer with one of its homologs (HMtA2) [30]. Similarly, one of the two histones of *Nanoarchaeum equitans* can only bind DNA as a heterodimer [31]. These histones can still *form* homodimers but these dimers do not appear to interact with DNA, at least under the conditions tested *in vitro*. In both instances, the requirement for a non-identical histone to generate dimers capable of DNA binding has evolved through small changes to the protein sequence; via a single amino acid substitution in *M. thermoautotrophicus* [30] and a four-amino acid insertion in the L1 region in *N. equitans*, reminiscent of the extended L1 region in eukaryotic H3 compared to H4 [31]. The physiological role of these variants, how they are deployed and interact with other histones to affect chromatin properties *in vivo*, remains to be elucidated. Studying them further might help us understand the evolution of obligate heterodimerization, which – thus far – appears unique to eukaryotes and some viruses [32].

2.3. Archaeal histone variants modify the histone-DNA complexes that contain them

Model archaeal histone dimers assemble into tetramers very similar in structure to eukaryotic $(H3\text{-}H4)_2$ tetramers [21] (Fig. 1c,d). These tetramers can have different properties depending on their constituent histone variants, including altered stability, DNA binding affinity and propensity to assemble into larger complexes (see below). Most of our experimental knowledge in this regard again comes from *in vitro* work on the two histones of *M. fervidus*, HMfA and HMfB, which are 85% identical at the amino acid level. HMfA binds DNA at a lower concentration than HMfB but is associated with less extensive compaction [19,33]. The two histones of another hyperthermophile, *T. kodakarensis*, also differ in their binding properties: HTkB[1] has greater affinity for DNA and compacts DNA more extensively than HTkA [22]. Similarly, one of the histones in *M. jannaschii* (HMjA4) has lower affinity to DNA and is less effective at inhibiting transcription *in vitro* [34]. Interestingly, this homolog is encoded on a plasmid that is not found in closely related strains, suggesting that HMjA4 might be a recent acquisition, yet to be fully integrated into the workings of the new host genome.

Assembling combinations of histone paralogs with different properties into heterologous complexes can result in structures with intermediate properties. Tetramers built from fusion dimers of HMfA and HMfB, for example, have DNA binding affinities that are stronger than $(HMfA)_4$ and (slightly) weaker than $(HMfB)_4$ [35]. Is this true for archaeal histone variants in general? Making defined, homogeneous populations of heteromeric complexes is difficult (although feasible by way of making fusion proteins). As a result, this question has not been addressed experimentally at scale. However, recent systematic *in silico* modelling of all possible tetrameric complexes that can be built from the seven histones of *Methanosphaera stadtmanae* suggests that combinations of paralogs often yield tetramers with intermediate DNA binding affinities or tetramer stabilities and that the properties of the complex can change in a graded, dosage-dependent fashion depending on the number of copies of a given variant that are incorporated into the complex [36].

2.4. Archaeal histone variants modify the oligomerisation potential of higher-order complexes

Variant incorporation might also be important for larger, super-tetrameric complexes. In *M. fervidus* [13,14], *T. kodakarensis* [21,37,38], and likely other species [8], tetramers can extend into longer oligomers of variable length, where dimers are added in a stepwise fashion to wrap increasingly more DNA (Fig. 1e). Using a histone such as HMfA/B or HTkA/B to build these structures, previously referred to as hypernucleosomes [8] or archaeosomes [39], this process can – theoretically – go on forever. Footprints of these complexes as large as 450 bp have been found *in vivo* using micrococcal nuclease digestion, consistent with complexes composed of 30 histones [21,38,40], although complexes are typically smaller (4–10 histones). This variability in size represents a sharp departure from the fixed octameric architecture of eukaryotic nucleosomes.

Not all archaeal histones support oligomerization equally well. Tethered particle motion experiments have revealed weaker stacking and lower stability of oligomeric complexes that contain HMfA compared to HMfB, consistent with differences in DNA compaction between HMfA and HMfB [33]. In other species, variants might interfere with oligomer extension more radically: We recently identified histone variants in several Methanobacteriales archaea that we called 'capstone' variants. These capstones are characterized by unstable tetramer interfaces and are predicted to prevent further stable extension when incorporated at the end of a growing oligomer (Fig. 2) [36]. Structural modelling of histones from *M. stadtmanae* also suggests that this capstone property is easy to evolve: a single amino acid change can significantly destabilize the tetramer interface [36]. Using a single substitution (loss of histidine at the end of α2, see Fig. 2) as a guide for capstone identity, we discovered histones with predicted capstone function in several independent archaeal lineages (Fig. 1f), suggesting parallel evolution of capstone functionality. Previously, Henneman and colleagues had also highlighted residues likely to disrupt stacking interactions between dimers in the context of longer oligomeric complexes [8]. Histones with substitutions at those residues might act in a similar manner to that proposed for capstones and destabilize or outright prevent multimerization.

It remains to be established if, when, and why these histone variants are incorporated in archaeal chromatin *in vivo*. The same is true for heteromeric histone complexes in general. With the exception of HMfA/B heterodimers (see above), we have no experimental confirmation that specific heteromeric complexes exist *in vivo*. Yet we strongly suspect that they do given what we know about how these complexes assemble: in

[1] Note that the A/B labels here simply reflect naming conventions rather than ancient paralogy.

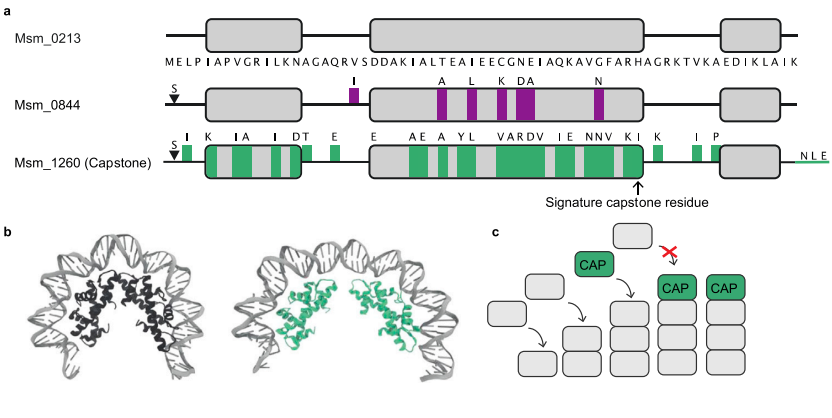


Fig. 2. Capstone histone variants. a. Histone secondary structure showing the positions of substitutions in the *Methanobrevibacter smithii* histones Msm_0844 and Msm_1260 (the putative capstone histone) relative to Msm_0123. Sites which are not conserved are highlighted in coloured blocks on the secondary structure for each histone. Insertions are marked with an arrow. The residue responsible for the capstone property (position 49 in HMfB coordinates) is highlighted. b. Representative frames from simulations of the most highly expressed histone (Msp_0122, left, black) and the capstone histone from *Methanosphaera stadtmanae* (Msp_0383, right, green). Images were rendered using VMD [69]. c. Model illustrating the potential capping function of capstone variants when integrated into oligomeric histone complexes.

contrast to eukaryotes, there are no known variant-specific (or even general) histone chaperones or remodellers in archaea. Instead, archaeal histones self-assemble onto DNA. This is not only true *in vitro* but also when histones are heterologously expressed in *E. coli* where – in the absence of any co-evolved deposition machinery – they associate with DNA and form complexes that protect DNA from micrococcal nuclease (MNase) digestion [14]. This self-assembly process is likely very dynamic. Recent work using cryo-EM and molecular dynamics simulations showed that HTkA oligomers can exhibit spontaneous breathing motions, where the oligomer opens up like a clam, providing both intermittent access to normally concealed DNA and perhaps a mechanism for exchange of histones paralogs without the need for dedicated remodelling [39]. We therefore suspect that variants will, perhaps somewhat haphazardly, be incorporated into and shape the properties of a heterogeneous population of complexes, which will affect DNA transactions throughout the nucleoid (Box 1).

2.5. Archaeal histone complexes and the nature of variants

The dynamic size of histone-DNA complexes and the capacity for homo- as well as heterodimerization have important implications for how we think about histone variants in archaea. Archaeal histone variants do not have positional identity in the same way that eukaryotic variants do. They are not swapped into a defined position in a fixed larger structure, like H2A.Z replacing replicative H2A in the octameric nucleosome. Archaeal histone complexes have rather more degrees of freedom in how different histones can be assembled relative to each other. In this regard, archaeal histones behave more like bacterial nucleoid-associated proteins, where homologous variants present in the same species (e.g. HupA and HupB in *E. coli*) can also often homo- and heterodimerise [41,42] and assemble into longer oligomeric complexes on DNA. When thinking about the nature and potential roles of histones variants in archaea, we must therefore avoid having our thoughts too narrowly guided by prior knowledge of (and pre-formed notions about) eukaryotic variants.

Box 1
Global versus local action

For lack of evidence to the contrary, we have assumed that assembly of tetramers and higher order structures is random, and that composition and assembly are consequently dictated primarily by mass action. Under such a model, chromatin architecture is altered on a global scale rather than locally. Changes in the expression of a particular histone variant tune chromatin properties such as genome accessibility genome-wide. This might be useful for a concerted response to environmental change (e.g. in response to heat stress) or in preparation for altered nutrient availability (e.g. during stationary phase). This is an obvious departure from the world of eukaryotes, where variants are often deposited in a precisely targeted fashion. Instead, the situation appears more like that found in bacteria where nucleoid composition and resultant architecture are constantly remodelled, as different nucleoid-associated proteins assume dominant roles in the chromatin landscape, depending on physiological need. For example, Dps is a major constituent of the *E. coli* nucleoid in stationary phase, where it mediates a compact, crystalline chromatin state, but barely detectable in exponential phase [70]. Archaeal histone variants might mediate similar global transitions in chromatin state, just in a more LEGO-like modular fashion, where one histone variant seamlessly slots in for another. Locally specific changes, on the other hand, require either variant-specific interactions with factors that confer specificity or that the variants themselves bind preferentially to different DNA sequences. To our knowledge, neither has been demonstrated for any archaeal histone *in vivo*. When expressed in *E. coli*, the binding footprints of HMfA and HMfB from *M. fervidus* are highly correlated, with only minor differences[14]. Similarly, in *T. kodakarensis*, Δ*htkA* and Δ*htkB* strains show similar MNase profiles [40]. If sequence preferences exist, they are likely to be subtle. SELEX experiments comparing the rank order and absolute differences in sequence preferences between paralogs would be informative, as would variant-specific pulldowns to establish topography of binding along the genome *in vivo*. Successful attempts at the latter have yet to be reported. In part, this might be because the lack of tails has made it harder (though likely not impossible) to introduce tags without disrupting native binding properties difficult (K. Sandman, personal communication). Similarly, obtaining antibodies with sufficient specificity to discern between paralogs for experiments such as ChIP-seq has been difficult, but would likely succeed eventually. As the capacity for locally specific incorporation is key to the compartmentalized nature of chromatin in eukaryotes, whether archaeal histones are deposited in a global or local manner remains one of the outstanding questions of archaeal histone biology.

3. Regulation

3.1. Archaeal histone variants exhibit dynamic differences in expression

Although we have a basic understanding of which variant complexes can form in principle, when and why those complexes are formed, how long they persist – and what their physiological relevance is – remains unknown. In the absence of known chaperones or remodellers, it seems reasonable to assume that complex assembly and deposition is more or less random and that specific variants are incorporated into dimers and higher-order complexes indiscriminately. As a result, chromatin properties would be altered globally rather than locally, a key difference to eukaryotes (Box 1). Under this model of chromatin assembly, controlled by mass action, the relative abundance of histone paralogs assumes a key role for chromatin structural change. How different variants are expressed, throughout the growth cycle and in the face of altered environmental conditions, might therefore give clues to their function.

What, then, do we know about differential expression of histone paralogs in archaea? To begin with, different histone paralogs, concurrently expressed in the same cell, often differ substantially in their relative expression level. Examples here include histone paralogs in *Methanococcus voltae* [43], *T. kodakarensis* [44,45], *M. fervidus* [19] and particularly in *M. stadtmanae,* where the abundance of different histone paralogs varies over a 27-fold range [36]. More importantly, these differences are often dynamic. Relative paralog dosage in *M. fervidus* [19], *M. voltae* [43] and *M. stadtmanae* [36], for example, changes at different stages of growth. In *M. fervidus,* HMfA is the most abundant histone in exponential phase, yet HMfB levels reach parity as cells enter stationary phase [19]. As recombinant HMfB compacts DNA more extensively than HMfA, this might suggest that stationary phase cells have more compact chromatin than cells at earlier growth stages [19]. Conversely, combining growth-phase specific expression data with structural modelling, we previously suggested that histone variants that increase in stationary phase in *M. stadtmanae* should generate *less* stable histone-DNA complexes on average than are found during exponential phase [36]. Both these predictions should be taken with a pinch of salt because they consider histones in isolation, rather than as part of a more complex chromatin ecosystem that, in archaea, often involves other highly expressed DNA-binding proteins [10,46]. How histones interact with these proteins, and whether histone variants display differential interactions, is very poorly understood and should be a key area of future research into archaeal chromatin [see [47] for a study breaking important ground in this regard].

3.2. Variants that change expression in response to environmental change

Regarding the role of histone variants in responding to environmental change, our knowledge remains similarly embryonic. The only information we have is that, at least on a few occasions, variants respond differentially to a change in external conditions. Expression of HTkA and HTkB in *T. kodakarensis,* for instance, varies with growth temperature. When shifted from their optimal growth temperature of 85 °C to 65 °C, the relative abundance of both paralogs drops, but more so for HTkA than HTkB (Table 1) [45]. Similarly, two of the six histones of *M. jannaschii* (MJ0168, MJ1258) are upregulated upon heat shock (98 °C vs 88 °C), one more strongly than the other (Table 1) [48]. Upon cold shock (65 °C), the same two histones, plus another (MJ0932), are downregulated [49]. MJ1258 is also more highly expressed following a pressure change from 7.8 atm to 500 atm alongside another, plasmid-encoded histone (MJECL17) [48]. Finally, only one of its six histone paralogs, MJ0932, is significantly upregulated in high H_2 environments and downregulated when *M. jannaschii* is grown alongside *Thermococcus paralvinellae* rather than in isolation [50]. Thus, differential responses of histone variants to a variety of stimuli are clearly evident. The physiological significance of these changes, however, remains unknown.

Meaningful progress here can be made at several levels. Even in species that remain genetically inaccessible, changes to chromatin composition and interaction partners under different conditions can be monitored using cross-linking mass-spectrometry, complemented by MNase-Seq and RNA-seq to monitor size and distribution of histone complexes and how they relate to gene expression. Ultimately, to assess changes in chromatin composition in different conditions, variant-specific labels or antibodies will need to be developed (see Box 1). Also desirable would be wider application of a classic approach to understanding gene function – deleting the gene of interest. In principle, deletion of histones with different structural properties may result in distinct changes to growth and gene expression. Where deletion studies in archaea with more than one histone gene have been carried out, variant-specific effects have been found: Deletion of *hstA* (but not *hstB*) in *M. voltae* [43] and *htkB* (but not *htkA*) in *T. kodakarensis* [51] reduces growth rate. Further, *T. kodakarensis* cells deficient in HTkA can no longer be transformed while deletion of *htkB* causes no such issue [51]. On a systems level, Δ*hstA* and Δ*hstB* strains of *M. voltae* exhibit distinct changes in gene expression [43], as do Δ*htkA* and Δ*htkB* strains in *T. kodakarensis* [51]. No obvious link was found between the positioning of tetrameric complexes and gene expression changes for each mutant, and MNase profiles from each strain were highly similar [40]. It is therefore tempting to speculate that structural properties of the variants are at fault for divergent effects on transcription. However, non-concordant deletion effects might also be caused by dosage effects, as deleting one histone might reduce overall histone titres more so than deleting another, an issue that has received insufficient attention in the current literature. Position effects might also play a role: deletions can disturb gene expression locally, which might have locus- and therefore variant-specific, system-wide knock-on effects. It is interesting to note in this context, that in *T. kodakarensis* expression of HTkA from a plasmid is toxic for wild-type but also Δ*htkA* cells, suggesting that ectopic expression cannot replace genomic expression [51]. Paralog swap experiments, where one histone variant is replaced in its native genomic location by a second variant (and vice versa), would be an interesting approach to try and disentangle these effects.

Table 1
Differential expression of archaeal histone variants in response to environmental change.

Organism	Conditions compared	Change in expression (log2-fold)	Reference
Methanocaldococcus jannaschii	Heat shock (98 °C/88 °C) at low pressure (7.8 atm)	MJ0168 = 1.1MJ1258 = 1.5	[48]
	Heat shock (98 °C/88 °C) at high pressure 500 atm	MJ0168 = 1.1MJ1258 = 1.8	[48]
	1.5 h after pressure shock (500 atm) at 88 °C	MJ1258 = 1.7MJECL17 = 1.3	[48]
	Gas substrate limited (600 ml/200 ml) at 7.8 atm	MJ0168 = 1.3MJ1258 = 1.3	[55]
	Gas substrate limited (600 ml at 500 atm/200 ml at 7.8 atm)	MJ0168 = 1.6MJ1258 = 1.2	[55]
	High H_2 v low H_2	MJ0932 = 1.4	[50]
	Co-culture with *Thermococcus paralvinellae* v monoculture	MJ0932 = −1.2	[50]
	Temperature change (65 °C/85 °C)	MJ0168 = −1.7MJ1258 = −1.5MJ0932 = −1.1	[56]
Thermococcus onnurineus	Nutrient change (modified minimal-CO/YPS)	TON_0185 = 1.6TON_1235 = 0.18	[57]
	Nutrient change (modified minimal-formate/YPS)	TON_0185 = 0.48TON_1235 = 0.56	[57]
Methanocella conradii	Co-culture v monoculture with *Pelotomaculum thermopropionicum*	Mtc_0089 = 0.17Mtc_0839 = −1.1	[58]
Thermococcus kodakarensis	Temperature change (65 °C/85 °C)	HTkA = −1.1HTkB = −0.51	[45]

Finally, a low-hanging fruit: Earlier investigations suggested that promoters of histone genes in several archaea harbour binding sites for heat shock regulators [52]. Pertinently, in species with two paralogs, one binding site was predicted to be stronger than the other [52]. Not all of these histones turned out to exhibit the predicted increase in expression following heat shock *in vivo* [53,54], suggesting that predictions here are not trivial. Nonetheless, with an increasingly diverse set of genomes and transcriptomes at our disposal, and an improved understanding of archaeal promoters and transcription factors, this area is ripe for renewed investigation and might take us closer to understanding the regulation of histone variants at the transcriptional level.

3.3. Post-transcriptional regulation

We know next to nothing about post-transcriptional regulation of histones in archaea, including how such regulation might differ between variants. This includes degradation of histones, about which we know nothing at all, but also an aspect of histone biology that takes up a prominent position in thinking about histone functional diversity in eukaryotes: post-translational modifications (PTMs).

In eukaryotes, histones are dynamically modified, particularly on their N-terminal tails. These PTMs, which include methylation, acetylation, and a myriad of other chemical modifications, can impact nucleosome dynamics and promote or restrict interactions with other proteins. Although some modifications are shared between replicative histones and their variants, others are variant-specific, often because the modified residue is unique to that variant [59]. For example, in response to DNA double strand breaks, H2A.X becomes phosphorylated on serine 139 in the C-terminal SQE/Dϕ motif (where ϕ is a hydrophobic amino acid), which is absent from replication-dependent H2A [60]. This phosphorylation event allows DNA repair through further interactions with a cascade of DNA repair proteins [2].

Almost all archaeal histones lack sizeable tails [8] and with them the major hotspots for modification in eukaryotes. The histone fold itself, however, can also be modified [61]. The finding that a number of archaeal genomes encode homologs of histone-modifying enzymes seemed to strengthen the case for histone PTMs in archaea, especially the discovery that, in some halophilic archaea, a putative histone deacetylase is located in an operon with the doublet histone [62]. Yet initial mass spectrometry analyses found no PTMs in histones from either *M. jannaschii* or *Methanosarcina acetivorans* [63] – what looked like a fire produced no smoke.

A more recent investigation, however, did find evidence for histone PTMs - specifically acetylation - in two members of the Thermococcales: *T. kodakarensis* and *Thermococcus gammatolerans* [64]. Both species encode two histone paralogs, each of which was found to be acetylated on several lysine residues. Interestingly, not all lysines are shared between the two paralogs, providing an opportunity for differential regulation of each variant via PTMs [64]. A few of the acetylated lysines, located towards the end of the core fold (L2; α3, Fig. 3), are found in near-orthologous positions to those in eukaryotes. PTMs of these residues might therefore have similar structural repercussions. Yet other lysines are found in parts of the protein, in particular the α2 helix, where eukaryotic histones rarely harbour lysines (Fig. 3), suggesting that any structural effects will have no eukaryotic equivalent. It is worth noting in this regard that four of the lysines in *T. kodakarensis* HtkB, including one in α2, have been predicted to play a role in hypernucleosome stability [8]. Of those, three showed evidence of acetylation (one is acetylated in both paralogs, the other two alternately acetylated in either HTkB or HTkA) suggesting a potential mechanism to control oligomer stability post-translationally in a variant-specific manner [64].

The discovery of histone PTMs is exciting, but comes with a major caveat: as of now, we lack direct information on modification stoichiometry, i.e. the fraction of histones that are modified. Establishing stoichiometry should be a priority for follow-up work; for those interested in physiological relevance, it matters whether 100% of histones are modified or 0.1%. The current data sketch out tantalizing possibilities for parallels with eukaryotes but fall short of demonstrating physiological significance. If modifications were indeed found to be common, future research should address how exactly those PTMs impact chromatin structure, how their deposition and removal is regulated, and how this process differs from variant to variant.

4. Evolution

The physiological role of histone variants is often hard to unravel experimentally. Many archaea remain genetically inaccessible or have eluded attempts to culture them. In the absence of experimental data, evolution can provide a powerful pointer to whether histone variants carry out distinct functions. From an evolutionary perspective, if sequence differences between paralogs have been maintained over long periods of time, this likely indicates that the paralogs play non-redundant physiological roles. In eukaryotes, H2A.Z from distantly related taxa cluster together on a phylogenetic tree, to the exclusion of other H2A variants, suggesting long-term maintenance of this particular variant [65]. Other variants, such as H2A.X, are polyphyletic, i.e. they have sequence or functional properties that have arisen multiple times independently in evolution [4,5].

We recently used this phylogenetic line of reasoning to search for evidence of long-term maintenance or parallel evolution of distinct histone types in archaea. We found evidence for both. In the Thermococcales, almost all species encode two paralogs, one with properties similar to *T. kodakarensis* HTkA and one more similar to HTkB (Fig. 4) [66]. Tracing the evolutionary history of these paralogs, we inferred that both were present in the last common ancestor of Thermococcales [66], approximately 750Mya [67]. Similarly, we uncovered several instances of long-term paralog maintenance in a distantly related order of archaea, the Methanobacteriales, which emerged around 1.6Gya [67]. Even if absolute divergence times are difficult to estimate and verify, a conservative interpretation of our results still indicates that a number of archaeal histone paralogs have been maintained as distinct entities for hundreds of millions of years [36,66], strongly suggesting that each plays a unique and important role in the cell. To note, these cases of long-term maintenance and initial diversification have happened in parallel to eukaryotes. So far, we have no evidence for shared ancestral paralogy (where histone variants in eukaryotes originated prior to eukaryogenesis and have also survived in extant representatives of some archaeal lineages).

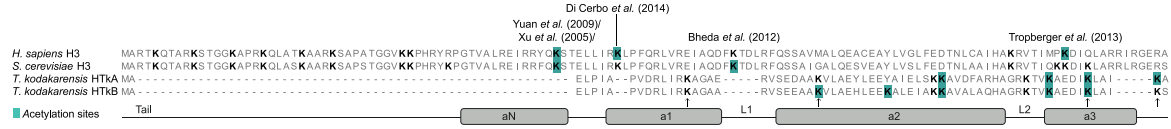


Fig. 3. Post-translational modifications of the core histone fold. Alignment of example eukaryotic histones with the two histones of *T. kodakarensis*. Lysines subject to post-translational modification in the globular domain of eukaryotic histone H3 (based on [61]) and those found acetylated in *T. kodakarensis* HTkA and HTkB [64] are highlighted with a blue shaded box. Lysines at position 56 [71], 64 [72] and 122 [73] have been found acetylated in human H3, as have lysines at positions 56 [74] and 79 [75] in H3 from *Saccharomyces cerevisiae*. The secondary structure for histone H3 is shown at the bottom. Arrows indicate lysines involved in potential stacking interactions for hypernucleosome formation in HTkB according to [8].

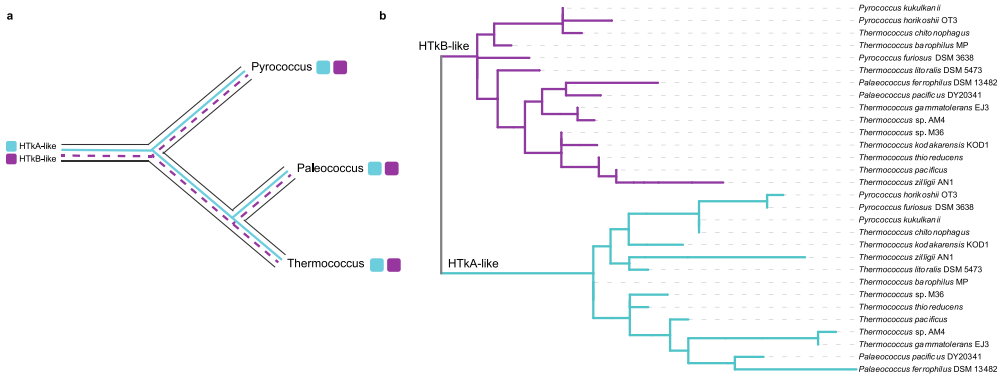


Fig. 4. Long-term maintenance of histone paralogs in the Thermococcales. a. Schematic Thermococcales tree illustrating long-term maintenance of histone paralogs with HTkA- and HTkB-like properties in Pyrococcus, Paleococcus and Thermococcus species, where both paralogs have been retained through multiple speciation events. b. Abridged protein-level tree for HTkA- and HTkB-like histones in the Thermococcales (see [66] for details), illustrating phylogenetic clustering of HTkA- and HTkB-like paralogs, which supports ancestral paralogy.

As highlighted above, we also find evidence for repeated independent emergence of particular histone properties, as histones to which we attribute capstone properties have evolved in several distantly related species. More cases like this will likely be revealed through systematic analysis of different clades of archaea, particularly as new genomes from diverse groups become available. Analysis here is complicated by the fact that archaeal histones are short and therefore typically yield poor phylogenetic signal. Combining evolutionary history with structural and functional data might therefore be critical to pin down conserved histone variants.

5. Conclusion

Our understanding of histone variants in archaea pales in comparison to what decades of diligent work have brought to light in eukaryotes. But this should not be taken as a lament. Instead, it should entice further, faster exploration of a country whose strange fruit are ripe for the picking. Prior work on eukaryotes has given us the tools and strategies to progress this exploration at speed, with a roadmap judiciously adjusted for the unique properties of archaeal histone complexes. Above, we have highlighted several areas where we are missing crucial information: Which other proteins do archaeal histones interact with, and are there interaction partners that are specific to certain variants? How is expression of archaeal histones controlled and how does control differ from one variant to another? Do different variants preferentially bind different sequences, and are differences in affinity large enough to matter *in vivo*. Is histone incorporation into larger oligomers really random? How does the composition of histone complexes change over time and are histones turned over quickly or slowly? What is the significance of post-translational modifications, how are they controlled, and are they widespread across archaea or limited to a few clades? Answers to all of these questions are within our grasp.

Conflict of interest

The authors declare that they are not subject to a conflict of interest.

Acknowledgements

The authors thank Antoine Hocher for comments on the manuscript. Work on histone variants in the Warnecke lab is funded by the UK Medical Research Council.

References

[1] S.V. Kumar, P.A. Wigge, H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis, Cell 140 (2010) 136–147, https://doi.org/10.1016/j.cell.2009.11.006.
[2] R. Scully, A. Xie, Double strand break repair functions of histone H2AX, Mutat. Res. Mol. Mech. Mutagen 750 (2013) 5–14, https://doi.org/10.1016/j.mrfmmm.2013.07.007.
[3] D.K. Palmer, K. O'Day, H.L.E. Trong, H. Charbonneau, R.L. Margolis, Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone, Proc. Natl. Acad. Sci. USA 88 (1991) 3734–3738, https://doi.org/10.1073/pnas.88.9.3734.
[4] S. Henikoff, M.M. Smith, Histone variants and epigenetics, Cold Spring Harb. Perspect. Biol. 7 (2015), a019364, https://doi.org/10.1101/cshperspect.a019364.
[5] H.S. Malik, S. Henikoff, Phylogenomics of the nucleosome, Nat. Struct. Biol. 10 (2003) 882–891, https://doi.org/10.1038/nsb996.
[6] S. Henikoff, K. Ahmad, Assembly of variant histones into chromatin, Annu. Rev. Cell Dev. Biol. 21 (2005) 133–153, https://doi.org/10.1146/annurev.cellbio.21.012704.133518.
[7] F. Mattiroli, S. D'Arcy, K. Luger, The right place at the right time: chaperoning core histone variants, EMBO Rep. 16 (2015) 1454–1466, https://doi.org/10.15252/embr.201540840.
[8] B. Henneman, C. van Emmerik, H. van Ingen, R.T. Dame, Structure and function of archaeal histones, PLoS Genet. 14 (2018), e1007582, https://doi.org/10.1371/journal.pgen.1007582.
[9] S.P. Laursen, S. Bowerman, K. Luger, Archaea: the final frontier of chromatin, J. Mol. Biol. 433 (2021), 166791, https://doi.org/10.1016/j.jmb.2020.166791.
[10] A. Hocher, G. Borrel, K. Fadhlaoui, J.F. Brugère, S. Gribaldo, T. Warnecke, Growth temperature is the principal driver of chromatinization in archaea, BioRxiv, 2021. doi: 10.1101/2021.07.08.451601.
[11] S. Sakrikar, A. Schmid, An archaeal histone-like protein regulates gene expression in response to salt stress, Nucl. Acid. Res. 49 (2021) 12732–12743, https://doi.org/10.1093/nar/gkab1175.
[12] T.J. Sanders, M. Lammers, C.J. Marshall, J.E. Walker, E.R. Lynch, T.J. Santangelo, TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin, Mol. Microbiol. 111 (2019) 784–797, https://doi.org/10.1111/mmi.14191.
[13] S.L. Pereira, R. a Grayling, R. Lurz, J.N. Reeve, Archaeal nucleosomes, Proc. Natl. Acad. Sci. USA 94 (1997) 12633–12637, https://doi.org/10.1073/pnas.94.23.12633.
[14] M. Rojec, A. Hocher, K.M. Stevens, M. Merkenschlager, T. Warnecke, Chromatinization of Escherichia coli with archaeal histones, Elife 8 (2019), e49038, https://doi.org/10.7554/eLife.49038.
[15] G. Arents, E.N. Moudrianakis, The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization, Proc. Natl. Acad. Sci. USA 92 (1995) 11170–11174, https://doi.org/10.1073/pnas.92.24.11170.
[16] K. Decanniere, A.M. Babu, K. Sandman, J.N. Reeve, U. Heinemann, Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon Methanothermus fervidus, J. Mol. Biol. 303 (2000) 35–47, https://doi.org/10.1006/jmbi.2000.4104.
[17] D.J. Soares, K. Sandman, J.N. Reeve, Mutational analysis of archaeal histone-DNA interactions, J. Mol. Biol. 297 (2000) 39–47, https://doi.org/10.1006/jmbi.2000.3546.
[18] K. Luger, A.W. Mäder, R.K. Richmond, D.F. Sargent, T.J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution, Nature 389 (1997) 251–260, https://doi.org/10.1038/38444.
[19] K. Sandman, R.A. Grayling, B. Dobrinski, R. Lurz, J.N. Reeve, Growth-phase-dependent synthesis of histones in the archaeon Methanothermus fervidus, Proc.

Natl. Acad. Sci. USA 91 (1994) 12624–12628, https://doi.org/10.1073/pnas.91.26.12624.
[20] M.R. Starich, K. Sandman, J.N. Reeve, M.F. Summers, NMR structure of HMfB from the hyperthermophile, Methanothermus fervidus, confirms that this archaeal protein is a histone, J. Mol. Biol. 255 (1996) 187–203, https://doi.org/10.1006/jmbi.1996.0016.
[21] F. Mattiroli, S. Bhattacharyya, P.N. Dyer, A.E. White, K. Sandman, B.W. Burkhart, K.R. Byrne, T. Lee, N.G. Ahn, T.J. Santangelo, J.N. Reeve, K. Luger, Structure of histone-based chromatin in Archaea, Science 357 (2017) 609–612, https://doi.org/10.1126/science.aaj1849.
[22] H. Higashibata, S. Fujiwara, M. Takagi, T. Imanaka, Analysis of DNA compaction profile and intracellular contents of archaeal histones from Pyrococcus kodakaraensis KOD1, Biochem. Biophys. Res. Commun. 258 (1999) 416–424, https://doi.org/10.1006/bbrc.1999.0533.
[23] L. Weng, Y. Feng, X. Ji, S. Cao, Y. Kosugi, I. Matsui, Recombinant expression and characterization of an extremely hyperthermophilic archaeal histone from Pyrococcus horikoshii OT3, Protein Expr. Purif. 33 (2004) 145–152, https://doi.org/10.1016/j.pep.2003.09.004.
[24] W. Zhu, K. Sandman, G.E. Lee, J.N. Reeve, M.F. Summers, NMR structure and comparison of the archaeal histone HFoB from the mesophile Methanobacterium formicicum with HMfB from the hyperthermophile Methanothermus fervidus, Biochemistry 37 (1998) 10573–10580, https://doi.org/10.1021/bi973007a.
[25] W.T. Li, R.A. Grayling, K. Sandman, S. Edmondson, J.W. Shriver, J.N. Reeve, Thermodynamic stability of archaeal histones, Biochemistry 37 (1998) 10563–10572, https://doi.org/10.1021/bi973006i.
[26] W.T. Li, K. Sandman, S.L. Pereira, J.N. Reeve, MJ1647, an open reading frame in the genome of the hyperthermophile methanococcus jannaschii, encodes a very thermostable archaeal histone with a C-terminal extension, Extremophiles 4 (2000) 43–51, https://doi.org/10.1007/s007920050006.
[27] K. Weidenbach, J. Glöer, C. Ehlers, K. Sandman, J.N. Reeve, R.A. Schmitz, Deletion of the archaeal histone in Methanosarcina mazei Gö1 results in reduced growth and genomic transcription, Mol. Microbiol. 67 (2008) 662–671, https://doi.org/10.1111/j.1365-2958.2007.06076.x.
[28] R.L. Fahrner, D. Cascio, J.A. Lake, A. Slesarev, An ancestral nuclear protein assembly: Crystal structure of the Methanopyrus kandleri histone, Protein Sci. 10 (2001) 2002–2007, https://doi.org/10.1110/ps.10901.
[29] A.I. Slesarev, G.I. Belova, S.A. Kozyavkin, J.A. Lake, Evidence for an early prokaryotic origin of histones H2A and H4 prior to the emergence of eukaryotes, Nucleic Acids Res. 26 (1998) 427–430, https://doi.org/10.1093/nar/26.2.427.
[30] K. Sandman, H. Louvel, R.Y. Samson, S.L. Pereira, J.N. Reeve, Archaeal chromatin proteins histone HMtB and Alba have lost DNA-binding ability in laboratory strains of Methanothermobacter thermautotrophicus, Extremophiles 12 (2008) 811–817, https://doi.org/10.1007/s00792-008-0185-3.
[31] U. Friedrich-Jahn, J. Aigner, G. Längst, J.N. Reeve, H. Huber, Nanoarchaeal origin of histone H3? J. Bacteriol. 191 (2009) 1092–1096, https://doi.org/10.1128/JB.01431-08.
[32] G. Yoshikawa, R. Blanc-Mathieu, C. Song, Y. Kayama, T. Mochizuki, K. Murata, H. Ogata, M. Takemura, Medusavirus, a novel large DNA virus discovered from hot spring water, e02130-18, J. Virol. 93 (2019), https://doi.org/10.1128/JVI.02130-18.
[33] B. Henneman, T.B. Brouwer, A.M. Erkelens, G.J. Kuijntjes, C. Van Emmerik, R. A. Van Der Valk, M. Timmer, N.C.S. Kirolos, H. Van Ingen, J. Van Noort, R. T. Dame, Mechanical and structural properties of archaeal hypernucleosomes, Nucleic Acids Res. 49 (2021) 4338–4349, https://doi.org/10.1093/nar/gkaa1196.
[34] S.P. Wilkinson, M. Ouhammouch, E.P. Geiduschek, Transcriptional activation in the context of repression mediated by archaeal histones, Proc. Natl. Acad. Sci. USA 107 (2010) 6777–6781, https://doi.org/10.1073/pnas.1002360107.
[35] F. Marc, K. Sandman, R. Lurz, J.N. Reeve, Archaeal histone tetramerization determines DNA affinity and the direction of DNA supercoiling, J. Biol. Chem. 277 (2002) 30879–30886, https://doi.org/10.1074/jbc.M203674200.
[36] K.M. Stevens, J.B. Swadling, A. Hocher, C. Bang, S. Gribaldo, R.A. Schmitz, T. Warnecke, Histone variants in archaea and the evolution of combinatorial chromatin complexity, Proc. Natl. Acad. Sci. USA 117 (2020) 33384–33395, https://doi.org/10.1073/PNAS.2007056117.
[37] H. Maruyama, J.C. Harwood, K.M. Moore, K. Paszkiewicz, S.C. Durley, H. Fukushima, H. Atomi, K. Takeyasu, N.A. Kent, An alternative beads-on-a-string chromatin architecture in Thermococcus kodakarensis, EMBO Rep. 14 (2013) 711–717, https://doi.org/10.1038/embor.2013.94.
[38] T.J. Sanders, F. Ullah, A.M. Gehring, B.W. Burkhart, R.L. Vickerman, S. Fernando, A.F. Gardner, A. Ben-Hur, T.J. Santangelo, Extended archaeal histone-based chromatin structure regulates global gene expression in Thermococcus kodakarensis, Front. Microbiol. 12 (2021) 1071, https://doi.org/10.3389/fmicb.2021.681150.
[39] S. Bowerman, J. Wereszczynski, K. Luger, Archaeal chromatin 'slinkies' are inherently dynamic complexes with deflected DNA wrapping pathways, Elife 10 (2021), e65587, https://doi.org/10.7554/eLife.65587.
[40] N. Nalabothula, L. Xi, S. Bhattacharyya, J. Widom, J.P. Wang, J.N. Reeve, T. J. Santangelo, Y.N. Fondufe-Mittendorf, Archaeal nucleosome positioning in vivo and in vitro is directed by primary sequence motifs, BMC Genom. 14 (2013) 391, https://doi.org/10.1186/1471-2164-14-391.
[41] J. Rouvière-Yaniv, N.O. Kjeldgaard, Native Escherichia coli HU protein is a heterotypic dimer, FEBS Lett. 106 (1979) 297–300, https://doi.org/10.1016/0014-5793(79)80518-9.
[42] L. Claret, J. Rouviere-Yaniv, Variation in HU composition during growth of Escherichia coli: the heterodimer is required for long term survival, J. Mol. Biol. 273 (1997) 93–104, https://doi.org/10.1006/jmbi.1997.1310.
[43] I. Heinicke, J. Müller, M. Pittelkow, A. Klein, Mutational analysis of genes encoding chromatin proteins in the archaeon Methanococcus voltae indicates their involvement in the regulation of gene expression, Mol. Genet. Genom. 272 (2004) 76–87, https://doi.org/10.1007/s00438-004-1033-5.
[44] D. Jäger, K.U. Förstner, C.M. Sharma, T.J. Santangelo, J.N. Reeve, Primary transcriptome map of the hyperthermophilic archaeon Thermococcus kodakarensis, BMC Genom. 15 (2014) 684, https://doi.org/10.1186/1471-2164-15-684.
[45] A. Sas-Chen, J.M. Thomas, D. Matzov, M. Taoka, K.D. Nance, R. Nir, K.M. Bryson, R. Shachar, G.L.S. Liman, B.W. Burkhart, S.T. Gamage, Y. Nobe, C.A. Briney, M. J. Levy, R.T. Fuchs, G.B. Robb, J. Hartmann, S. Sharma, Q. Lin, L. Florens, M. P. Washburn, T. Isobe, T.J. Santangelo, M. Shalev-Benami, J.L. Meier, S. Schwartz, Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping, Nature 583 (2020) 638–643, https://doi.org/10.1038/s41586-020-2418-2.
[46] Z.F. Zhang, L. Guo, L. Huang, Archaeal chromatin proteins, Sci. China Life Sci. 55 (2012) 377–385, https://doi.org/10.1007/s11427-012-4322-y.
[47] H. Maruyama, E.I. Prieto, T. Nambu, C. Mashimo, K. Kashiwagi, T. Okinaga, H. Atomi, K. Takeyasu, Different proteins mediate step-wise chromosome architectures in Thermoplasma acidophilum and Pyrobaculum calidifontis, Front. Microbiol. 11 (2020) 1247, https://doi.org/10.3389/fmicb.2020.01247.
[48] B.B. Boonyaratanakornkit, L.Y. Miao, D.S. Clark, Transcriptional responses of the deep-sea hyperthermophile Methanocaldococcus jannaschii under shifting extremes of temperature and pressure, Extremophiles 11 (2007) 495–503, https://doi.org/10.1007/s00792-007-0063-4.
[49] B.B. Boonyaratanakornkit, A.J. Simpson, T.A. Whitehead, C.M. Fraser, N.M.A. El-Sayed, D.S. Clark, Transcriptional profiling of the hyperthermophilic methanarchaeon Methanococcus jannaschii in response to lethal heat and non-lethal cold shock, Environ. Microbiol. 7 (2005) 789–797, https://doi.org/10.1111/j.1462-2920.2005.00751.x.
[50] B.D. Topçuoğlu, C. Meydan, T.B. Nguyen, S.Q. Lang, J.F. Holden, Growth kinetics, carbon isotope fractionation, and gene expression in the hyperthermophile Methanocaldococcus jannaschii during hydrogen-limited growth and interspecies hydrogen transfer, e00180-19, Appl. Environ. Microbiol. 85 (2019), https://doi.org/10.1128/AEM.00180-19.
[51] L. Čuboňová, M. Katano, T. Kanai, H. Atomi, J.N. Reeve, T.J. Santangelo, An archaeal histone is required for transformation of Thermococcus kodakarensis, J. Bacteriol. 194 (2012) 6864–6874, https://doi.org/10.1128/JB.01523-12.
[52] M.S. Gelfand, E.V. Koonin, A.A. Mironov, Prediction of transcription regulatory sites in Archaea by a comparative genomic approach, Nucleic Acids Res. 28 (2000) 695–705, https://doi.org/10.1093/nar/28.3.695.
[53] L. Rohlin, J.D. Trent, K. Salmon, U. Kim, R.P. Gunsalus, J.C. Liao, Heat shock response of Archaeoglobus fulgidus, J. Bacteriol. 187 (2005) 6046–6057, https://doi.org/10.1128/JB.187.17.6046-6057.2005.
[54] K.R. Shockley, D.E. Ward, S.R. Chhabra, S.B. Conners, C.I. Montero, R.M. Kelly, Heat shock response by the hyperthermophilic archaeon Pyrococcus furiosus, Appl. Environ. Microbiol. 69 (2003) 2365–2371, https://doi.org/10.1128/AEM.69.4.2365-2371.2003.
[55] B. Boonyaratanakornkit, J. Córdova, C.B. Park, D.S. Clark, Pressure affects transcription profiles of Methanocaldococcus jannaschii despite the absence of barophilic growth under gas-transfer limitation, Environ. Microbiol. 8 (2006) 2031–2035, https://doi.org/10.1111/j.1462-2920.2006.01083.x.
[56] B.B. Boonyaratanakornkit, A.J. Simpson, T.A. Whitehead, C.M. Fraser, N.M.A. El-Sayed, D.S. Clark, Transcriptional profiling of the hyperthermophilic methanarchaeon Methanococcus jannaschii in response to lethal heat and non-lethal cold shock, Environ. Microbiol. 7 (2005) 789–797, https://doi.org/10.1111/j.1462-2920.2005.00751.x.
[57] S. Cho, M.-S. Kim, Y. Jeong, B.-R. Lee, J.-H. Lee, S.G. Kang, B.-K. Cho, Genome-wide primary transcriptome analysis of H2-producing archaeon Thermococcus onnurineus NA1, Sci. Rep. 7 (2017) 43044, https://doi.org/10.1038/srep43044.
[58] P. Liu, Y. Lu, Concerted metabolic shifts give new insights into the syntrophic mechanism between propionate-fermenting Pelotomaculum thermopropionicum and hydrogenotrophic Methanocella conradii, Front. Microbiol. 9 (2018) 1551, https://doi.org/10.3389/fmicb.2018.01551.
[59] D. Corujo, M. Buschbeck, Post-translational modifications of H2A histone variants and their role in cancer, Cancers 10 (2018) 59, https://doi.org/10.3390/cancers10030059.
[60] C. Redon, D. Pilch, E. Rogakou, O. Sedelnikova, K. Newrock, W. Bonner, Histone H2A variants H2AX and H2AZ, Curr. Opin. Genet. Dev. 12 (2002) 162–169, https://doi.org/10.1016/S0959-437X(02)00282-4.
[61] P. Tessarz, T. Kouzarides, Histone core modifications regulating nucleosome structure and dynamics, Nat. Rev. Mol. Cell Biol. 15 (2014) 703–708, https://doi.org/10.1038/nrm3890.
[62] N. Altman-Price, M. Mevarech, Genetic evidence for the importance of protein acetylation and protein deacetylation in the halophilic archaeon Haloferax volcanii, J. Bacteriol. 191 (2009) 1610–1617, https://doi.org/10.1128/JB.01252-08.
[63] A.J. Forbes, S.M. Patrie, G.K. Taylor, Y. Bin Kim, L. Jiang, N.L. Kelleher, Targeted analysis and discovery of posttranslational modifications in proteins from methanogenic archaea by top-down MS, Proc. Natl. Acad. Sci. USA 101 (2004) 2678–2683, https://doi.org/10.1073/pnas.0306575101.
[64] B. Alpha-Bazin, A. Gorlas, A. Lagorce, D. Joulié, J.-B. Boyer, M. Dutertre, J.-C. Gaillard, A. Lopes, Y. Zivanovic, A. Dedieu, F. Confalonieri, J. Armengaud, Lysine-specific acetylated proteome from the archaeon Thermococcus gammatolerans reveals the presence of acetylated histones, J. Proteom. 232 (2021), 104044, https://doi.org/10.1016/j.jprot.2020.104044.

[65] P.B. Talbert, K. Ahmad, G. Almouzni, J. Ausió, F. Berger, P.L. Bhalla, W.M. Bonner, W.Z. Cande, B.P. Chadwick, S.W.L. Chan, G.A.M. Cross, L. Cui, S.I. Dimitrov, D. Doenecke, J.M. Eirin-López, M.A. Gorovsky, S.B. Hake, B.A. Hamkalo, S. Holec, S.E. Jacobsen, K. Kamieniarz, S. Khochbin, A.G. Ladurner, D. Landsman, J. A. Latham, B. Loppin, H.S. Malik, W.F. Marzluff, J.R. Pehrson, J. Postberg, R. Schneider, M.B. Singh, M.M. Smith, E. Thompson, M.-E. Torres-Padilla, D. J. Tremethick, B.M. Turner, J.H. Waterborg, H. Wollmann, R. Yelagandula, B. Zhu, S. Henikoff, A unified phylogeny-based nomenclature for histone variants, Epigenetics Chromatin 5 (2012) 7, https://doi.org/10.1186/1756-8935-5-7.

[66] K.M. Stevens, A. Hocher, T. Warnecke, Deep conservation of histone variants in Thermococcales archaea, Genom. Biol. Evol. 14 (2022) evab274, https://doi.org/10.1093/gbe/evab274.

[67] J.M. Wolfe, G.P. Fournier, Horizontal gene transfer constrains the timing of methanogen evolution, Nat. Ecol. Evol. 2 (2018) 897–903, https://doi.org/10.1038/s41559-018-0513-7.

[68] E.J. Draizen, A.K. Shaytan, L. Mariño-Ramírez, P.B. Talbert, D. Landsman, A. R. Panchenko, HistoneDB 2.0: a histone database with variants—an integrated resource to explore histones and their variants, Database 2016 (2016) baw014, https://doi.org/10.1093/database/baw014.

[69] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38, https://doi.org/10.1016/0263-7855(96)00018-5.

[70] T. Ali Azam, A. Iwata, A. Nishimura, S. Ueda, A. Ishihama, Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid, J. Bacteriol. 181 (1999) 6361–6370, https://doi.org/10.1128/JB.181.20.6361-6370.1999.

[71] J. Yuan, M. Pu, Z. Zhang, Z. Lou, Histone H3-K56 acetylation is important for genomic stability in mammals, Cell Cycle 8 (2009) 1747–1753, https://doi.org/10.4161/cc.8.11.8620.

[72] V. Di Cerbo, F. Mohn, D.P. Ryan, E. Montellier, S. Kacem, P. Tropberger, E. Kallis, M. Holzner, L. Hoerner, A. Feldmann, F.M. Richter, A.J. Bannister, G. Mittler, J. Michaelis, S. Khochbin, R. Feil, D. Schuebeler, T. Owen-Hughes, S. Daujat, R. Schneider, Acetylation of histone H3 at lysine 64 regulates nucleosome dynamics and facilitates transcription, Elife 3 (2014), e01632, https://doi.org/10.7554/eLife.01632.

[73] P. Tropberger, S. Pott, C. Keller, K. Kamieniarz-Gdula, M. Caron, F. Richter, G. Li, G. Mittler, E.T. Liu, M. Bühler, R. Margueron, R. Schneider, Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer, Cell 152 (2013) 859–872, https://doi.org/10.1016/j.cell.2013.01.032.

[74] F. Xu, K. Zhang, M. Grunstein, Acetylation in histone H3 globular domain regulates gene expression in yeast, Cell 121 (2005) 375–385, https://doi.org/10.1016/j.cell.2005.03.011.

[75] P. Bheda, S. Swatkoski, K.L. Fiedler, J.D. Boeke, R.J. Cotter, C. Wolberger, Biotinylation of lysine method identifies acetylated histone H3 lysine 79 in Saccharomyces cerevisiae as a substrate for Sir2, Proc. Natl. Acad. Sci. USA 109 (2012) E916–E925, https://doi.org/10.1073/pnas.1121471109.

[76] Y. Liu, K.S. Makarova, W.-C. Huang, Y.I. Wolf, A.N. Nikolskaya, X. Zhang, M. Cai, C.-J. Zhang, W. Xu, Z. Luo, L. Cheng, E.V. Koonin, M. Li, Expanded diversity of Asgard archaea and their relationships with eukaryotes, Nature 593 (2021) 553–557, https://doi.org/10.1038/s41586-021-03494-3.