

# Chromatin properties of temperature-adapted eukaryotes

David Albir Haba

**UPV Tutor:**

Prof. Jose Gadea Vacas

**NUIG Supervisor:**

Prof. Andrew Flaus

**Autor:**

David Albir

**Valencia,**

June 2017

**UNIVERSITAT POLITÈCNICA DE VALÈNCIA**

ESCOLA TÈCNICA SUPERIOR D'ENGINYERIA AGRONÒMICA I DEL MEDI  
NATURAL



**Grado en Biotecnología**

**NATIONAL UNIVERSITY OF IRELAND GALWAY**



**NUI Galway**  
**OÉ Gaillimh**



BIOMEDICAL SCIENCES BUILDING, NUIG

**Title:** Chromatin properties of temperature-adapted eukaryotes

**Author:** David Albir Haba

**UPV Tutor:** Prof. Jose Gadea Vacas

**NUIG Supervisor:** Prof. Andrew Flaus

**Location and date:** Valencia, June 2017

**Abstract:**

Chromatin packaging is linked with the genome sequence, molecular mechanisms and biological properties of cells. We hypothesize that the histone proteins packaging cold-living eukaryotes (psychrophiles) have enhanced flexibility over mesophiles such as commonly studied animals and fungi, whereas the histones of warm adapted eukaryotes (thermophiles) have enhanced stability compared to mesophiles.

In this project we will use synthetic genes to recombinantly express and purify histones with sequence of the Antarctic fish *Notothenia coriiceps* living below 4°C and the fungus *Chaetomium thermophilum* with optimum living temperature of 60°C.

We attempted to assemble these into histone octamers and nucleosomes to enable comparison of their properties with human equivalents.

**Key words:** Histones, Temperature-adapted, Chromatin, Eukaryotes, Fungi, Bacteria, Thermophiles, Mesophiles.

**Título:** Propiedades de la cromatina en eucariotas adaptados a la temperatura

**Autor:** David Albir Haba

**UPV Tutor:** Prof. Jose Gadea Vacas

**NUIG Supervisor:** Prof. Andrew Flaus

**Lugar y fecha:** Valencia, Junio 2017

**Resumen:**

La compactación de la cromatina está relacionada con la secuencia genómica, los mecanismos moleculares y las propiedades biológicas de las células. Se hipotetiza que el compactamiento de las proteínas histonas de los animales eucariotas adaptados al frío (psicrófilos) han ganado en flexibilidad frente a los mesófilos. En cambio las histonas de los eucariotas adaptados a altas temperaturas (termófilos) como los hongos han ganado en estabilidad en comparación frente a los mesófilos.

En este proyecto genes recombinantes serán usados para expresar y purificar las histonas del pez de la Antártida *Notothenia coriiceps* que vive por debajo de los 4°C y el hongo *Chaetomium thermophilum* con una temperatura óptima de desarrollo de 60°C.

Se ensamblaran estas histonas en octámeros y nucleosomas para permitir la comparación de sus propiedades con sus equivalentes humanos.

**Palabras clave:**

Histonas, Adaptación-temperatura, Cromatina, Eucariotas, Hongos, Bacterias, Termófilos, Psicrófilos, Mesófilos

# ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisor, Prof. Andrew Flaus, for mentoring me doing this project, for teaching me and for being an exceptional reference as science person and human being.

To Martin Browne, for his patience, for giving me the best advices in the hardest circumstances and being always a good company in these long hours doing science.

To Alejandro Puchol, for being a mate during this prepossessing journey, a source of intellectual inspiration, and which is more, a good friend.

And last but not least to my family, especially to my uncle, for still guiding me, like the wise Melquíades, in a world of dreams, mazes, mirrors and white tigers.

# INDEX

1. INTRODUCTION .....	3
1.1. Characteristics and thermal adaptation of eukaryotes.....	3
1.1.1. Psychrophiles.....	4
1.1.2. Thermophiles.....	5
1.2. Functions of chromatin in eukaryotes .....	6
1.2.1. Structure of chromatin.....	7
1.2.2. Epigenetics .....	8
1.3. Histones and nucleosomes.....	8
1.3.1. Histone proteins.....	8
1.3.2. Histone fold dimers and octamers .....	9
1.3.3. Nucleosomes .....	9
1.4. Aims and objectives .....	10
2. MATERIALS AND METHODS .....	11
2.1. Materials .....	11
2.2. Methods .....	13
2.2.1. DNA methods .....	13
2.2.2. Protein methods.....	14
2.2.3. Nucleosome methods .....	17
3. RESULTS.....	18
3.1. Establishing novel histone expression .....	18
3.1.1. De novo sequence optimization.....	18
3.1.2. Plasmid construction.....	18
3.1.3. <i>N. coriiceps</i> H3 mutagenesis .....	20
3.2. Optimization of protein expression .....	21
3.2.1. Expression media trial .....	21
3.2.2. Expression strain trial.....	21
3.2.3. Variation in expression between inoculums .....	23
3.2.4. Optimization of <i>C. thermophilum</i> H4 ribosome binding site .....	23
3.3. Preparation of histones.....	25
3.3.1. Large scale expression.....	25
3.3.2. Purification by chromatography .....	25
3.3.3. Purification by gel filtration chromatography.....	26
3.4. Assembly of octamers refolding from histones .....	27

3.4.1.	<i>N. coriiceps</i> octamers .....	28
3.4.2.	<i>C. thermophilum</i> histones do not refold into octamers.....	29
3.5.	Nucleosome assembly.....	29
3.5.1.	<i>N. coriiceps</i> nucleosomes in vitro by salt dialysis.....	30
4.	DISCUSSION .....	31
4.1.	Histone selection and gene design.....	31
4.2.	Protein expression and purification .....	31
4.3.	Octamers refolding and nucleosome .....	32
5.	CONCLUSIONS .....	33
6.	APPENDINX.....	34
6.1.	Histone Genes sequences .....	34
6.2.	Histone Protein sequences.....	36
6.3.	Histone proteins alignments .....	37

## LIST OF FIGURES AND TABLES

**Figure 1.** Stability curve for a human protein

**Figure 2.** *C. thermophilum* protein Arx1

**Figure 3.** Nucleosome compaction of DNA into

**Figure 4.** Dimers conformation.

**Figure 5.** Nucleosome dialysis

**Figure 6.** Structure of the plasmid with the histone gene CtH2AX

**Figure 7.** *N. coriiceps* H3 mutagenesis sequences.

**Figure 8.** Test expression media optimization.

**Figure 9.** Test expression and media optimization.

**Figure 10.** Expression trial.

**Figure 11.** Plasmid structure marking the region where the insertion of the gap and primers Design for *C. thermophilum* H4 mutagenesis

**Figure 12.** SDS-Page Gel with the mutagenesis of *C. thermophilum* H4.

**Figure 13.** Anion exchange chromatography for NcH3

**Figure 14.** Process column filtration eliminate DNA.

**Figure 15.** *N. coriiceps* histones refolded in octamers

**Figure 16.** *C. thermophilum* histones do not refolded in octamers.

**Figure 17.** *N. coriiceps* Nucleosomes and tetrasomes on 54A54 DNA sequence

**Table 1.** Solutions and Buffers

**Table 2.** PCR conditions for KOD hot start polymerase.

**Table 3.** PCR program using KOD hot start polymerase.

**Table 4.** Description of the vector with the correspondent histone gene

**Table 5.** Summary table of origin, state, properties and production of each histone.

# 1. INTRODUCTION

## 1.1. Characteristics and thermal adaptation of eukaryotes

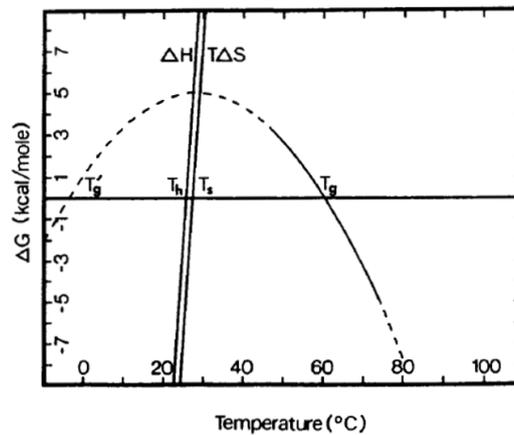
Eukaryote is an ancient Greek word which means 'true nucleus' (sic). The Eukaryote domain of life comprises cells with a compact nucleus and defined compartments separated by membranes, which are thought to be invaginations of the external plasma membrane. The nuclear envelope delimits the nucleus, which is the largest organelle in the cell and contains the genetic code of the cell in the form of chromosomes (Pollard et al., 2007). This specific compaction of DNA is distinct in the Eukaryote domain compared to the other two prokaryotic domains, Bacteria and Archaea.

The Eukaryote domain includes a great variety of organisms apparently very different from each other, that form five supergroups: Amoebozoa, Excavata, Archaeplastida, Stremnopiles, Alveolata and Rhizaria (SAR) and Opisthokonta (Adl et al., 2012). These supergroups in turn, comprise kingdoms including Fungi (*Chaetomium thermophilum*) and Metazoan/Animata (*Homo sapiens*, *Xenopus laevis*, *Caenorhabditis elegans* and *Notothenia coriiceps*) among others.

Eukaryotes and other organisms have struggled against their environment, so under Darwin's theory they adapt, change and evolve. Temperature is one environmental variable for which adaptation has been necessary, and the impact of the temperature in Eukaryotes is well known (Krenek et al., 2012).

At the molecular level, this adaptation depends on the behavior of proteins and other biomolecules. The stability curve of a protein is defined as the plot of the free energy of unfolding as a function of temperature. Human proteins are in their most stable conformation around 25°C (Figure 1).

## PROTEIN STABILITY CURVES



**Figure 1.** Stability curve for a human protein. The curve is obtained from temperature variation of enthalpy and entropy of unfolding (Becktel and Schellman, 1987).

A mesophile organism has an optimal living temperature between 15° and 35° C. Other organisms are specialized to survive at extreme temperatures and their adaptation is interesting. In this project we have investigated how Eukaryotic cells have solved the problem of DNA compactation in variable thermal circumstances.

### 1.1.1. Psychrophiles

The word psychrophiles refers to the organisms able to live at temperatures below 5°C. However, their optimal temperatures for development can be between 12° and 15°C, and their upper temperature tolerance can be between 15° and 20°C.

In a 1928 expedition in Norway, Ditlef Rustad captured a translucent fish in the south shore of the island of Bouvetøya, and gave it the name of "Farvelöst Blood" ('Colorless blood') in his diary (Rustad, 1930). Many years later it was realized that these ice-fish are the only vertebrate animals without myoglobin or hemoglobin in their blood. In these cold temperatures oxygen dissolves to a higher concentration than in mesophile temperatures so can dissolve directly into plasma, which is less viscous because it does not require erythrocytes. The absence of oxygen carrying proteins enabled the fish to develop much larger hearts, denser capillary beds, higher volumes of blood and larger blood vessels where plasma can circulate faster with lower energy expenditure.

Another adaptation of ice fish is the appearance of antifreeze proteins, which derive from an ancestral trypsinogen gene that was duplicated in an example of neofunctionalisation (Beers et al., 2015). One copy maintained the original role as protease, but the other accumulated mutations that eventually provided the antifreeze function. These proteins bind to ice crystals and prevent their growth that results in cell damage. As shown by genomic and proteomic

studies, cold-shock proteins are highly expressed in ice fish and can have crucial roles in protein folding, control of nucleic-acid secondary structure, and transcription and translation (Shin et al., 2014).

Proteins are the main targets of these adaptations as they control the equilibrium between substrates and products, influx of nutrients, outflow of waste products, macromolecular assemblies, nucleic-acid dynamics and appropriate folding. Their adaptation seems to rely on a higher flexibility of key parts of the molecular structure or of whole folded domains through a decreased stability that partly compensates for the freezing effect of low temperatures on the three-dimensional structure. The Heat Shock Response (HSR) is a mechanism of defense against thermal stress from elevated temperatures. Curiously, HSR proteins seems to be constitutively expressed in cold adapted organisms, presumably to mitigate cold denaturation of proteins (Shin et al., 2014) .

Global warming does not bode well for Antarctic fish, and even less for ice-fish which are more sensitive to temperature changes than red-blooded fish. In addition, an increase in temperature leads to an increase in the acidity of the ocean and, as a result, an imbalance in ecosystems and food webs.

Despite their adaptations to the cold environment, genome-wide studies have only recently been performed due to the lack of a sequenced genome (Shin et al., 2014). This revealed that *N. coriiceps* has bigger genome, higher dN/dS ratio comparing with the other species that belong to the same family but not exposed to this environment and multiple post-translational modifications (Shin et al., 2014). The Antarctic bullhead Notothen, *N. coriiceps*, is therefore an endemic teleost ice fish which can be a good model to understand and represent the cold evolution adaptations to the genome.

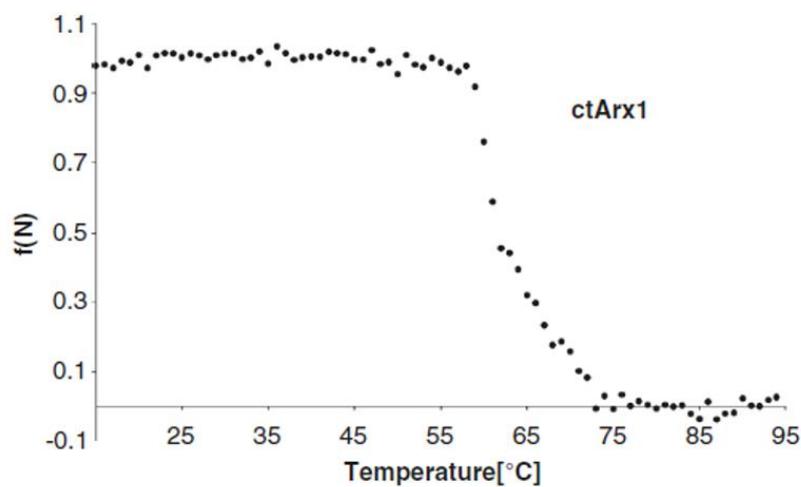
### 1.1.2. Thermophiles

Thermophile is a term applied to the organisms that can grow in extreme conditions of high temperatures above 50°C. Such organisms have been widely used in biotechnology because of the inherent property that their proteins are resistant to high temperatures. *Chaetomium thermophilum* is a thermophilic filamentous fungus whose genome has been published recently (Kellner et al., 2016). It grows on dung or compost and has an optimum temperature of 60°C, which is one of the highest for Eukaryotes. In contrast, *Saccharomyces cerevisiae*, a model yeast, has 73% protein homology to this fungus but has a temperature optimum of only 30°C (van den Brink et al. 2015).

Proteins for thermophilic organisms are not only stable to unfolding at higher temperatures, but also generally more stable than their mesophilic relatives. Alongside this high stability and

resistance of denaturation of proteins, thermophiles such *C. thermophilum* show codon usage bias and reduced genome size with fewer protein coding genes than mesophiles of the same family (*C. thermophilum* 7267, *T. terrestris* 9813; *T. heterothallica* 9110). Shorter introns and intergenic regions and less transposable elements may indicate that transposition is unfavorable at higher temperatures (van den Brink et al. 2015).

On the other hand, thermophile proteins show a high concentration of isoleucine, tryptophan and tyrosine. There is also an overrepresentation of cysteine for catalytic residues, disulfide bridges and metal binding contribute to protein thermal stability, and proline which makes the structure of the proteins more rigid and less unfoldable.



**Figure 2.** *Ct. thermophilum* protein Arx1. This specific protein of *C. thermophilum*, which is homologous to yeast pre-ribosomal export factor Arx1, unfolds around 62°C. (Van Noort et al. 2013)

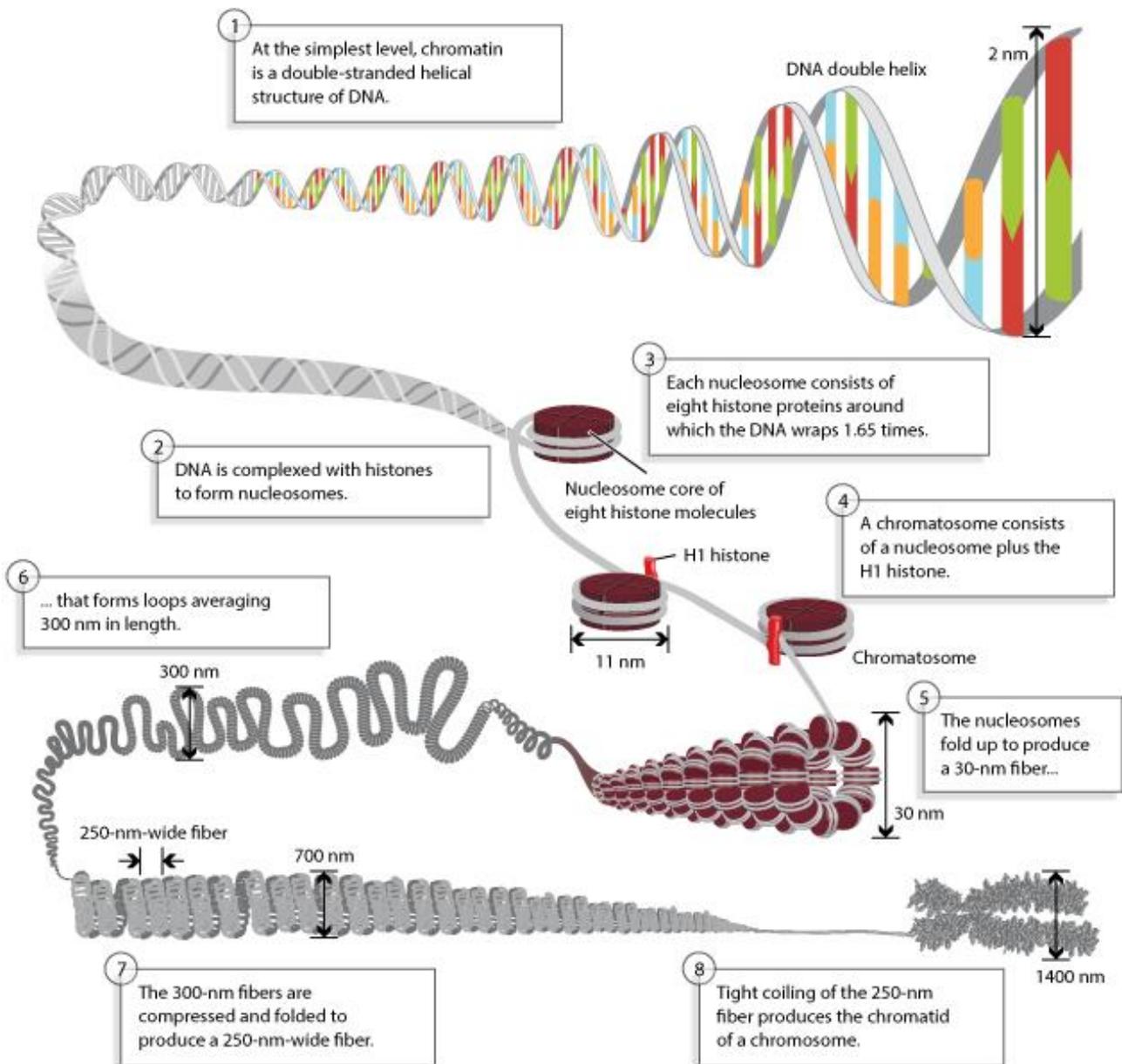
By solving the 3D structure of the *C. thermophilum* protein Arx1 (Figure 2), three types of adaptive mutations were uncovered: 1) the loop rigidity thanks to the increased proline frequency; 2) an increased protein core hydrophobicity, and 3) an increase in electrostatic interactions that stabilize neighboring secondary structure elements (van Noort et al. 2013).

This makes *C. thermophilum* an interesting model for adaptation because its genome and proteome are very well characterized and provide a basis to interpret how histones have evolved in order to perform the nuclear compaction of the DNA.

## 1.2. Functions of chromatin in eukaryotes

The main function of the chromatin is to package the genetic material in the eukaryotic cell. DNA is packaged in a series of repeating nucleosome units in chromosomes, and this structure is predictable and highly conserved. The conformation of the nucleosomes provides a balance

of static and dynamic organization which is essential for chromatin function. The other functions of the chromatin apart from packaging the DNA are to organize and control the expression of the proteins, DNA replication and DNA repair (Figure 3).



**Figure 3.** Nucleosome compaction of DNA into chromosomes (Pierce 2003).

### 1.2.1. Structure of chromatin

To perform special cell functions and respond to a changing environment, the way that the genome is accessible to the transcriptome and regulation complex machinery is dynamic and coordinated. Chromatin serves as platform for all nuclear processes from gene expression to DNA replication and DNA damage repair. This structure changes in part thanks to remodeling complexes that control nucleosome positioning. The profile of nucleosomes varies throughout

the cell-cycle involving histone post-translation modifications and incorporation of histone variants via chromatin remodeling complexes during gene regulation (Bruno et al., 2003).

### 1.2.2. Epigenetics

A variety of post-translational modifications including acetylation, phosphorylation, methylation, ubiquitination and ADP-ribosylation occur on the tail regions of histones. These modifications led to the 'histone code hypothesis' whereby marks in some histones tails encode changes in the chromatin-templated processes. This histone code seems to be, at least in some part, heritable or epigenetic.

Some studies show that the exposure to high temperatures for one generation in *C. elegans* changes the expression of temperature-induced proteins until at least the 14<sup>th</sup> generation. This change is related with the altered trimethylation of histone H3 lysine 9 (H3K9me3) before the onset of zygotic transcription. When the worms were at 20°C, they produced a small quantity of a fluorescent protein. However, when the temperature of the habitat was raised to 25°C, the amount of fluorescent protein produced increased. Afterwards, even after return to the lower the temperature of 20°C, the fluorescent protein production was maintained. Most surprising was that this kind of memory of the warm period was not only maintained in individuals who experimented it but also in the children and grandchildren of these single worms maintained only at 20°C. These later generations continued to show the fluorescence that indicated the experience of their parents and grandparents to heat. The effect lasted up to seven generations and, for a generator of five generations at 25°C, the fluorescence was maintained up to 14 generations.

One of the explanations for this phenomenon is that because the generations of these worms are so short and the environment could change more slowly, that adaptation for long-lived epigenetic inheritance may be useful to them (Klosin et al., 2017).

## 1.3. Histones and nucleosomes

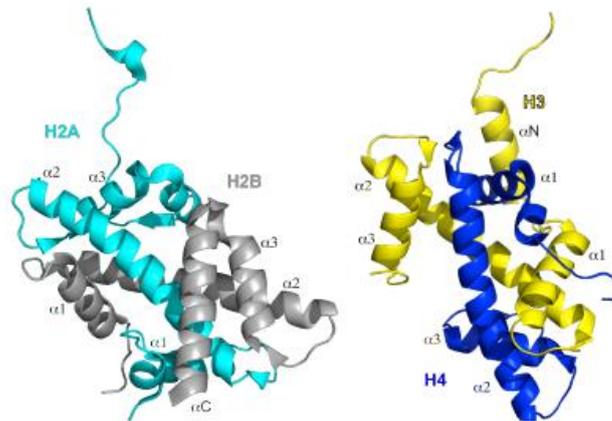
### 1.3.1. Histone proteins

Histones are small proteins of 10-15 kDa that are highly conserved in evolution. There are 5 types of histones, H2A, H2B, H3, H4 and H1. H2A has a variant, H2AX, found in almost all eukaryotes that makes up a significant but variable proportion of bulk chromatin and plays a role in DNA damage and repair. H2A and H2B form a histone fold dimer, and H3 and H4 form another one. Nucleosomes involve 2 each of the 4 core histones H2A, H2B, H3 and H4 and one molecule of H1 for every 200 bp of DNA. The nucleosome core particle (NCP) structures have been solved using histones from *Xenopus* and Human and other organisms before (Patwal, 2014). NCPs are connected to each other by a DNA linker of 20-100 bp, that can be associated

with a molecule of H1. It has been extremely difficult to understand the principles of binding histones to DNA because wrapped DNA behaves as an irregular and cooperative superhelix.

### 1.3.2. Histone fold dimers and octamers

The histone octamer in an NCP is an arrangement of (H3-H4)<sub>2</sub> tetramer at the center and H2A-H2B dimers on either side of the tetramer. These histones are associated as heterodimers by specific hydrophobic interactions in anti-parallel pairs of H2A-H2B and H3-H4 (Figure 4).



**Figure 4.** Histone fold dimers conformation. The histones adopt this conformation this order to perform an octamer: (H2A–H2B:H4–H3:H3–H4:H2B–H2A) (Patwal, 2016)

The octamer conformation provides a characteristic spiral of histone fold dimers in a specific way that makes possible the histone–DNA interactions which is crucial for nucleosome positioning. The histone fold motif comprises a long central  $\alpha 2$  helix flanked by shorter  $\alpha 1$  and  $\alpha 3$  helices. Folding forces based mainly on hydrophobic interactions allow antiparallel pairs H2A–H2B and H3–H4 to interact along the  $\alpha 2$  helix, with the  $\alpha 1$  and  $\alpha 3$  helices folded back across the long  $\alpha 2$  helix (Dechassa and Luger, 2012).

The octamer conformation is the platform enabling DNA contacts. Two types of DNA interaction are generated in the surfaces of  $\alpha 1$  and  $\alpha 3$  helices. The first is formed by the combination of pairs of loops connecting alpha helices known as L1L2 motifs because the pairs involve the loops between  $\alpha 1$  and  $\alpha 2$  (L1), and  $\alpha 2$  and  $\alpha 3$  (L2). The second type of interaction surface is formed by paired N-terminal ends of  $\alpha 1$  helices, and is known as the  $\alpha 1\alpha 1$  motif (Flaus, 2011).

### 1.3.3. Nucleosomes

The nucleosome is defined as the repeating unit of chromatin. The NCP is formed by a structured octamer of the core histones H2A, H2B, H3 and H4 wrapped with 147 bp DNA (Oudet et al., 1975). Nucleosomes are arranged into regularly spaced array, with the linker region between nucleosomes. Nucleosome positioning influences gene expression, and has

implications for chromatin packaging, sequence evolution, and the evolution of gene expression. The three principles of nucleosome dynamics and positioning are (i) DNA sequence accessibility depends on its position within the nucleosome structure, (ii) DNA sequences can influence their location in the nucleosome and (iii) DNA and the histone octamer core together control the potential for structural changes in the nucleosome. The propensity for nucleosomes to resist or undergo structural changes is referred to as nucleosome stability (Flaus, 2011).

#### 1.4. Aims and objectives

Different types of genomes need different properties of chromatin compaction, so variations in histones are expected. However, histones and nucleosomes appear to be highly conserved in Eukaryotic evolution. Significant but small changes in histone sequences are likely to affect the compaction of the DNA and genome, especially in divergent eukaryotes such as those exposed to different thermal adaptations. We set out to compare the histones of the thermophile *C. thermophilum* and the psychrophile *N. coriiceps* with those from more well studied mesophile animals like *H. sapiens* and *X. laevis*.

The aim of this project is uncover the nature of histones variations that enable adaptation to extreme temperatures. To accomplish this, our first objective was to express all four core histones for each organism recombinantly in *E. coli* and then to purify the recombinant histones and assemble them into nucleosomes.

## 2. MATERIALS AND METHODS

### 2.1. Materials

<b>Buffers and solutions</b>	<b>Composition</b>	<b>Application</b>
Ampicillin	Stock Solution: 50 mg/ml in ddH <sub>2</sub> O. Used at 50 ug/ml	Selection for antibiotic resistance
Binding Buffer	100mM Tris-HCl pH 7.5, 500 mM NaCl and 0.1 mM EDTA	For binding proteins for column filtration
Chloramphenicol	stock solution: 34 mg/ml in absolute Ethanol. Used at 34 ug/ml	Selection for antibiotic resistance
Coomassie Stain	50% v/v Methanol, 10% v/v Acetic Acid, 40% v/v deionized water, 0,1% 2wv Brilliant Blue G, 0,01% w/v Brilliant Blue R	Protein stain in SDS-PAGE
Denaturing Histrap Buffer A	20 mM Tris-HCl pH 7.5, 100 mM NaCl, 7 M Urea	His Trap Chromatography for Histones
Denaturing Histrap Buffer B	20 mM Tris-HCl pH 7.5, 100 mM NaCl, 7 M Urea, 500 mM Imidazole	His Trap Chromatography for Histones
Denaturing Ion Exchange Buffer A	20 mM Tris-HCl pH 7.5, 0,1 EDTA, DTT, 7 M Urea	Ion exchange Chromatography for Histones
Denaturing Ion Exchange Buffer B	20 mM Tris-HCl pH 7.5, 0.1 EDTA, DTT, 7 M Urea, 2 M NaCl	Ion exchange Chromatography for Histones
DNA Purification Buffer A	20 mM Tris-HCl pH 7.5, 0.1 EDTA	Protein purification of DNA
DNA Purification Buffer B	20 mM Tris-HCl pH 7.5, 0.1 EDTA, 2 M NaCl	Protein purification of DNA
Histone Octamer Refolding Buffer	20 mM Tris-HCl pH 7.5, 0.1 EDTA, 2 M NaCl	Histone Octamer Preparation for Gel filtration and CGC/FPLC
Histone Unfolding Buffer	20 mM Tris-HCl pH 7.5, 10 mM DTT, 7 M Guanidine HCl	Histone Octamer Preparation for Gel filtration and CGC/FPLC
Histone Wash Buffer	50 mM Tris-HCl pH 7.5, 10 mM DTT, 100 mM NaCl, 1 mM EDTA, 1 mM Benzamidine	Protein Extraction and Inclusion Bodies Preparation
Histone Wash Buffer with Triton X-100 1%	51 mM Tris-HCl pH 7.5, 10 mM DTT, 100 mM NaCl, 1 mM EDTA, 1 mM Benzamidine, 1% v/v Triton X-100	Inclusion Body Preparation

Kanamycin	Stock Solution: 30 mg/ml in ddH <sub>2</sub> O. Used at 30 µg/ml	Selection Antibiotic Resistance
Nucleosome Dialysis Buffer A	0.8 M NaCl, 10mM Tris-HCl pH 7.5, 1mM DTT	Nucleosome Stepwise Dialysis
Nucleosome Dialysis Buffer B	0.6 M NaCl, 10mM Tris-HCl pH 7.5, 1mM DTT	Nucleosome Stepwise Dialysis
Nucleosome Dialysis Buffer C	0.5 M NaCl, 10mM Tris-HCl pH 7.5, 1mM DTT	Nucleosome Stepwise Dialysis
Nucleosome Dialysis Buffer D	0.1 M NaCl, 10mM Tris-HCl pH 7.5, 1mM DTT	Nucleosome Stepwise Dialysis
SDS Loading Buffer 6x	125 mM Tris-HCl pH 6.8, 2% v/v of 20% SDS, 60% v/v of 100% Glycerol, DTT, 0,8 mg/ml Bromophenol Blue	SDS-PAGE
Soluble His Trap buffer A	20 mM Tris-HCl pH 7.5, 100 mM NaCl	His Trap Chromatography for Histones
Soluble His Trap buffer B	20 mM Tris-HCl pH 7.5, 100 mM NaCl, 500 Imidazole	His Trap Chromatography for Histones
Soluble Ion Exchange Buffer A	20 mM Tris-HCl pH 7.5, 50 mM NaCl	Histone column filtration
Soluble Ion Exchange Buffer B	20 mM Tris-HCl pH 7.5, 1 M NaCl	Histone column filtration
Sucrose Loading Buffer 6x	30 % v/v sucrose in deionized H <sub>2</sub> O	Loading Sample in Native PAGE
Terrific Broth (TB)	12 g/l Tryptone, 24 g/l Yeast extract, 9.4 g/l KH <sub>2</sub> PO <sub>4</sub> , 4ml/l Glycerol	Media for Histone Expression
Wash Buffer	50 mM Tris-HCl pH 7.5, 100 mM NaCl	Preparation for Inclusion Bodies
2YT Broth	16 g/l Tryptone, 10 g/l Yeast Extract, 5 g/l NaCl	Media for Histone Expression

**Table 1.** Solutions and Buffers

## 2.2. Methods

### 2.2.1. DNA methods

#### 2.2.1.1. Transformation

50  $\mu$ l of E.coli competent cells (could be Rosetta2, Star pRIL, BL21 PLysS) were studded on ice with 1  $\mu$ l of DNA on ice for 30 min. Then a heat shock at 42°C for 60 sec. and placed back to ice for another 10 min. 700  $\mu$ l of LB were added to the tube and then was incubated at 37°C, 180 rpm, 45 min. 150  $\mu$ l of this culture were plated in agar plates with resistance antibiotics selection (ampicillin, kanamycin, and chloramphenicol) and incubated over night at 37°C.

#### 2.2.1.2. PCR Mutagenesis

The complementary primers with the single base mutation were designed with the flanking region on each side with an extension of 20 bp with a Tm of 50°C. The plasmid was amplified using the primers with the conditions and reagents of the table 2 and 3. The same reaction with no primers was set as negative control. The template and the newly synthesized strands in the PCR product can be distinguished by their mutilation status basing on that, 1,5 $\mu$ l of DpnI was directly added to each PCR reaction and incubated for 16h at 37°C. Once digested both samples were transformed in Rossetta2 competent cells (as described in 2.2.1 before) and plated in different agar plates. Depending on the number of colonies of the PCR with primers and without primers would validate sadistically if the mutation were inserted or no. if the number on the PCR with primers is significantly bigger than the other, a single colony of that plate will be chose to do a miniprep (Thermo Scientific) and sending it for sequencing to double check if the single mutation is correctly inserted in the plasmid.

Description	Final concentration
DNA	10 ng
10 KOD buffer	1 x
2mM dNTPs	0.2 mM
25mM Mg <sup>2+</sup>	2 mM
10 $\mu$ M Primer 1	0.3 $\mu$ M
10 $\mu$ M Primer 2	0.3 $\mu$ M
KOD polymerase	1U
water	to adjust volume
Total volume	50 $\mu$ L

**Table 2.** PCR conditions for KOD hot start polymerase.

Step	Temperature (°C)	Time (min:sec)	Cycles
Initial denaturing	95	2:00	1
Denaturing	95	0:30	
Annealing	56	0:20	16
Extension	68	0:30/kb	
Final Extension	72	5:00	1
Hold	4	-	-

**Table 3.** PCR program using KOD hot start polymerase.

### 2.2.1.3. Plasmids

Histone genes in were in the pD54-SR or in the pET3 plasmid vectors, as shown in the table 4.

Vector	Gene insert	Gap between RBS and Gene sequence	De novo synthesis	Antibiotic Resistance
pD451-SR	Ct H2AX	No		
pD451-SR	Ct H2B	No		
pD451-SR	Ct H3	No	Yes	Kanamycin
pD451-SR	Ct H4	No		
pD451-SR	Nc H2A	No		
pD451-SR	Nc H2B	No		
pEt3a	Nc H3	Yes	No	Ampicillin
pEt3a	Hs/Nc H4	Yes		

**Table 4.** Description of the vector with the correspondent histone gene

## 2.2.2. Protein methods

### 2.2.2.1. Histone test expression

Transformed competent cells with the plasmids containing the histones sequences were plated on antibiotic selection plates of agar overnight. Next day a swipe of the colonies was inoculated into 20 ml of fresh media at 37°C, 180 rpm, approx. 3-4h until the OD was between 0.6-0.8. Then IPTG was added to a final concentration of 0.4 mM. Samples can be taken at 30 min, 2h, 4h, and overnight since that point to run a SDS-PAGE with them.

### 2.2.2.2. Histone large expression

A swipe of colonies were inoculated in a 2L flask containing 1L of fresh media. The culture was incubated at 37°C, 180 rpm, approx. 3-4h when the OD was between 0.6-0.8, 1 ml of the sample was taken to run a SDS-PAGE, IPTG was added to a final concentration of 0.4 mM, let the culture incubate for 2-3h more and collect another 1ml of the sample for the SDS-PAGE gel to check if the protein induction expression worked. The media was collected and centrifuged to obtain the pellet that was stored at 20 °C until further use. The samples collected to check

the protein expression, were as well spun down to a pellet and resuspended in protein gel loading buffer (PGLB) and heated at 95°C for 5 minutes. SDS-PAGE was run with the samples and standard in Coomassie blue stain for 2 hours. And destained in Coomassie Destain after.

*Equipment used in histone expression:*

- Centrifuge Avanti J-26 XP Beckman Coulter
- Rotor JA17 and JLA 10.5 Beckman Coulter
- Shaker incubator Innova 44 New Brunswick Scientific
- Sonicador. ULTRATURRAX Branson sonifier

### 2.2.2.3. Histone purification

#### 2.2.2.3.1. *Inclusion body preparation*

The first step for histone purification is inclusion body preparation. Cell pellet containing histone inclusion bodies were resuspended in Wash Buffer and sonicated on ice bath for 2 minutes at 40% amplitude using Branson sonifier. Cell suspension was then transferred to a 50mL NALGENE tube and centrifuged for 15 minutes at 4°C with 23 000 g. The pellet was retained and washed again with Wash Buffer and centrifuged.

Wash steps were repeated twice with Triton Wash Buffer and finally twice again with Wash Buffer. Inclusion bodies were then solubilized by incubating in 0.5mL DMSO at room temperature on a roller for half an hour followed by 1 h incubation in 5mL Unfolding Buffer.

When pellet was almost dissolved the suspension was centrifuged for 20 minutes with 35 000 G. The supernatant containing protein was collected in a new 50mL tube and diluted to 1 in 20 with histone binding buffer to purify with ion exchange chromatography.

#### 2.2.2.3.2. *Cation exchange chromatography*

The histone solution, after inclusion body preparation, was centrifuged to remove E. coli DNA contamination or any undissolved matter and filtered through 0.2µm PVDF filter. Protein solution was loaded on Fast flow liquid chromatograph (FPLC) with histone binding buffer using HiTrap SP FF 5mL column with flow 1mL/min-1 and eluted in Elution Buffer.

#### 2.2.2.3.3. *Dialysis and Lyophilization*

Histones were dialyzed against water to eliminate all the urea containing in the solution. 3h 3 times and then over night. Dialysis was performed at 4°C against four changes of refolding buffer. For 1 mg of each histone, dialysis was carried out first in 350mL of refolding buffer twice for 1h each followed by 400mL for 2 h. The result of the dialysis was frozen at -80°C over night and then lyophilized for one or two day until was dry.

#### 2.2.2.4. Concentration determination

Concentration of purified protein was measured with UV spectrophotometer. Purified protein solution was either transferred to a cuvette or measured on a nanodrop pedestal. Absorbance at 276nm ( $A_{276}$ ) measured and purity of the protein was determined by 260:280 ratio in a ThermoScientific spectrophotometer. The absorbance values were noted to calculate accurate concentration using the formula:

$$C \text{ (mgmL}^{-1}\text{)} = [A_{276} / \epsilon] * MW$$

$$C \text{ (}\mu\text{M)} = [A_{276} / \epsilon] * 10^6$$

*\*Concentration (C) determination of pure protein. Where molecular weight (MW) is in Dalton and molar extinction coefficient ( $\epsilon$ ) is in  $M^{-1}/cm^{-1}$ .*

#### 2.2.2.5. Histone complex refolding

Histones were refolded into dimer, tetramer or octamers using salt-dialysis method. The lyophilized histone aliquots were dissolved in unfolding buffer to a final concentration of 2  $mg/mL^{-1}$ . Unfolding was carried out in unfolding buffer at room temperature for 30 minutes. Concentration of each histone was determined by UV spectrophotometer. The four histone proteins were mixed in an equimolar ratios and volume was adjusted to a final protein concentration of 1  $mg/mL^{-1}$  using unfolding buffer. Histone mixture was then transferred into dialysis bags with 8,000 MW cut off. Dialysis was performed at 4°C against four changes of refolding buffer. For 1 mg of each histone, dialysis was carried out first in 350 mL of refolding buffer twice for 1 h each followed by 400 mL for 2 h. The final dialysis step was carried out overnight in 1 L refolding buffer. Precipitated protein was removed by centrifugation and concentrated to a final volume of 1 mL. This concentrated sample was then purified by gel filtration chromatography. Refolded histone complexes were always kept at 4°C. Histone proteins were refolded into octamer using refolding buffer containing 2 M NaCl and histones dimer or tetramer were assembled in refolding buffer with 1 M NaCl.

#### 2.2.2.6. Gel filtration chromatography

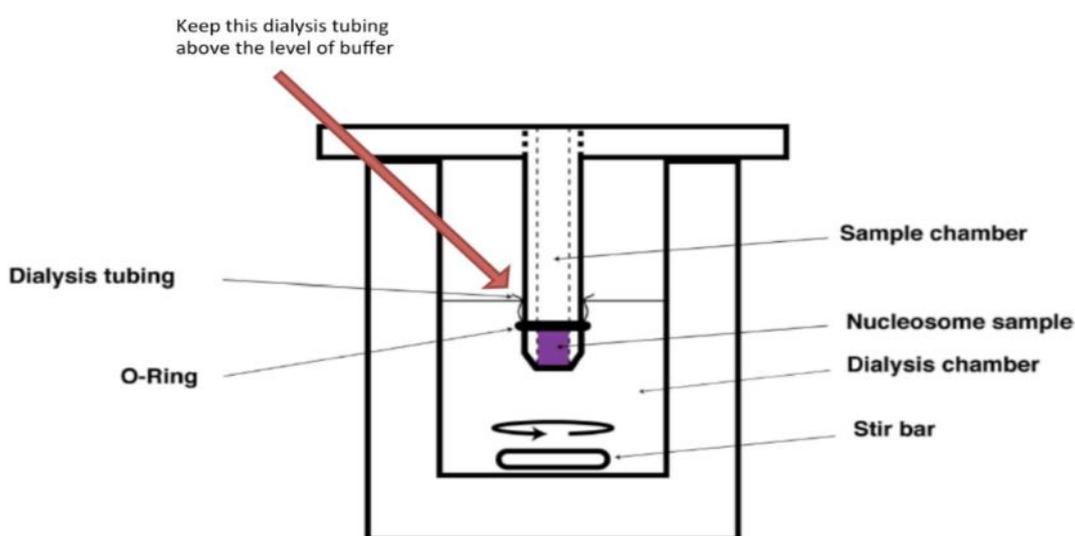
Purification of histone complexes were carried out on a Superdex-200 gel filtration column with 0.5 mL flow rate. Column was pre-equilibrated with filtered and degassed refolding buffer. Concentrated histone octamer sample of volume 0.5 mL were loaded into the FPLC system using 1 mL syringe. Purified octamer, tetramer and dimer peaks were collected in 96 well plates with a set fraction size of 0.5 mL. High molecular weight aggregates elutes first at about 7 mL followed by histone octamer at 12 mL to 13 mL and histone dimer at 15 mL for *N*.

*Coriiceps*. The purity and stoichiometry of fractions were checked on 15% SDS PAGE. Fractions of equimolar quantity of the histone proteins were pooled and stored at 4°C.

### 2.2.3. Nucleosome methods

#### 2.2.3.1. Salt dialysis

Purified octamer was mixed with equal molar amounts of DNA cy-labelled in a solution containing final concentration of 2 M NaCl and 10 mM Tris-Cl, pH 7.5 buffer. The reaction mix was transferred into mini dialysis blocks (Figure) after 30 minutes incubation on ice followed by exchange of buffers for 2 h each in Nucleosome Dialysis Buffer A 0.8 M NaCl, Nucleosome Dialysis Buffer B 0.6 M NaCl, Nucleosome Dialysis Buffer C 0.5 M NaCl and Nucleosome Dialysis Buffer D 0.1 M NaCl at 4°C (Figure 2). Nucleosomes were stored at 4°C.



**Figure 5.** Nucleosome dialysis (Doran, 2013). The DNA and histone octamer are placed into the sample chamber in 25  $\mu$ l volumes and this one is then placed into the pre-chilled dialysis chamber which contains 8 ml of 4°C dialysis buffer. The assembled unit is then placed on a magnetic stirrer and left to dialyze in a 4°C cold room.

#### 2.2.3.2. Native gel electrophoresis

A 6% native polyacrylamide gel was prepared and was left to set for 1 h before pre-equilibrating for 3 h at 300 V in 0.2xTBE. From a 100 pmol nucleosome reaction, 2  $\mu$ L was mixed with 8  $\mu$ L of 5% sucrose in a 1.5 mL tube on ice. This mixture was loaded on the pre-equilibrated native gel and was further run for 3 h at 300 V. The gel was scanned using fluorescent imager for Cy-labelled samples.

## 3. RESULTS

### 3.1. Establishing novel histone expression

The first objective that is required for study the nucleosome of the divergent thermal organisms is to express a significant amount of protein for each histone. To achieve this purpose, proteins were expressed in *E. coli*, where they are expected to be free of post-translational modifications with high expression levels in inclusion bodies that should limit the proteolytic degradation.

#### 3.1.1. De novo sequence optimization

The genes encoding each histone were designed. As no genomic material from the divergent organisms was available, the sequences were synthesized de novo using algorithms optimized for maximal production in *E. coli*. Each gene was optimized for codon usage and to minimize 5' mRNA secondary structure. It is important to achieve this optimization of gene sequence to suit the codon usage of the heterologous host because certain codons that are rare in *E. coli* affect recombinant protein expression (Kane, 1995). Codon optimization also plays an important role to avoid tRNA depletion by eliminating repetitive codons in the genes (Henaut and Danchin, 1996).

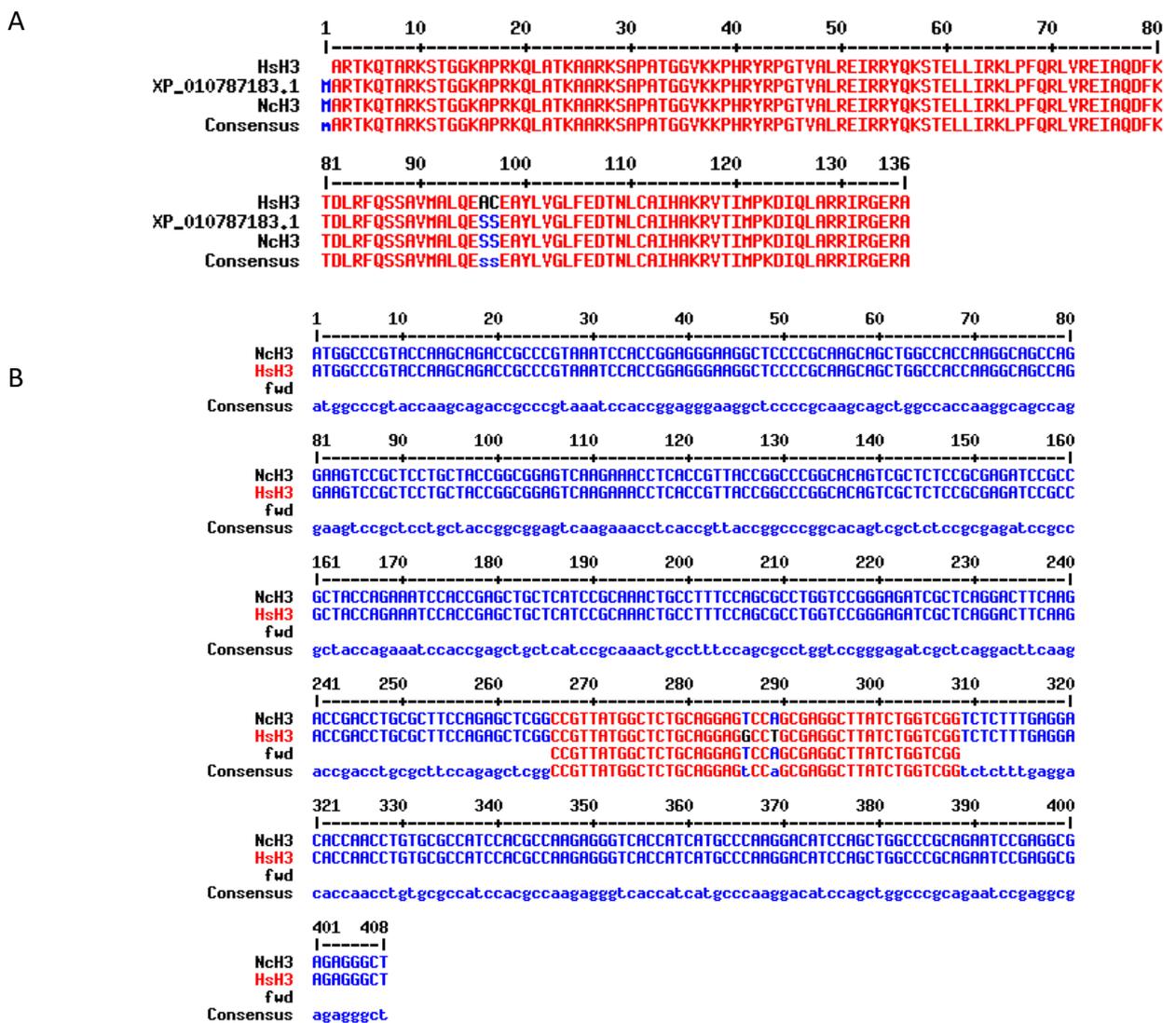
#### 3.1.2. Plasmid construction

The chosen vector was pD451-SR (Figure 6) from the company DNA 2.0. This plasmid includes a gene coding for kanamycin resistance (NPT II/Neo) to enable selection. Secondly it contains, the lacI regulatory region of the lac operon that codes for the repressor that binds very tightly to a short DNA sequence just downstream of the promoter near the beginning of lacZ called the lac operator. The repressor binding to the operator interferes with binding of RNAP to the promoter, and therefore transcription occurs only at very low levels. Thirdly, the LacO regulatory region of the lac operon is included. If lactose is missing from the growth medium, the repressor binds very tightly to a short DNA sequence just downstream of the promoter near the beginning of lacZ called the lac operator. The repressor binding to the operator interferes with binding of RNAP to the promoter, and therefore transcription occurs only at very low levels. When cells are grown in the presence of lactose, a lactose metabolite called allolactose, which is a combination of glucose and galactose, binds to the repressor, causing a change in its shape. Thus altered, the repressor is unable to bind to the operator, allowing RNAP to transcribe and thereby leading to higher levels of the encoded proteins. Silencing of the promoter prior to IPTG induction is achieved using symmetrical lac operators (Sadler et al., 1983) spaced around the promoter to maximize cooperativity (Oehler et al., 1994). This operator pair ensures significantly tighter repression than regular lac operators (Lanzer and



### 3.1.3. *N. coriiceps* H3 mutagenesis

*N. coriiceps* H3 is highly homologous to human histones so this gene was not synthesized de novo, but instead the H3 gene was generated by mutagenesis from human sequences already available in the plasmid pET3a. Mutagenesis primers for the two amino acid changes were designed from the *H. sapiens* H3 sequence to generate the *N. coriiceps* H3 encoding gene (Figure 2). The mutagenesis was successful as the plate with the negative control had 2 colonies and the positive one had 15. The mutagenesis was checked by sequencing the plasmid.



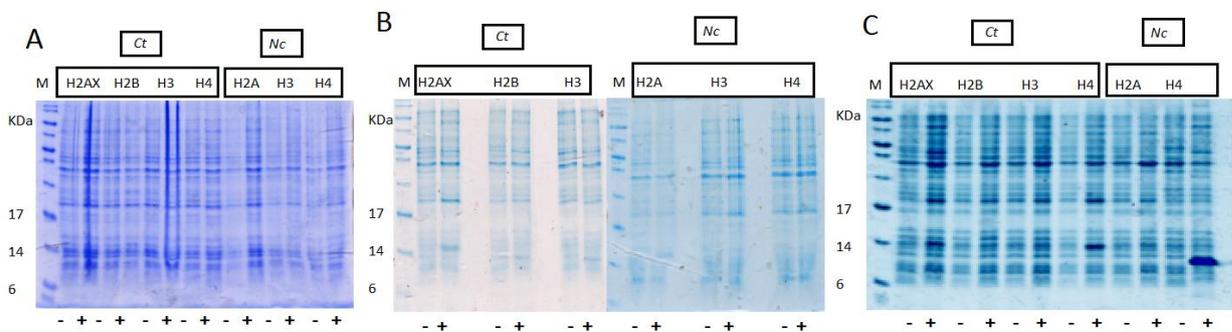
**Figure 7.** *N. coriiceps* H3 mutagenesis sequences. (A) Alignment of protein sequences *H. sapiens* H3 against *N. coriiceps* cH3. (B) Alignment DNA sequence of HsH3 against NcH3. In Black is the two base mutation, in red primer desing.

## 3.2. Optimization of protein expression

All eight plasmids encoding the four histones from the two organisms were transformed into Rosetta2(DE3) pLysS cells. This *E. coli* strain express T7 lysozyme that inhibits the basal activity of T7 RNA polymerase and contributes for lower gene toxicity (Studier, 1991). Growth conditions were at 37°C with 180 rpm agitation. The *E. coli* cells were induced to express the proteins when the cell density reached OD<sub>600</sub> 0.6-0.8.

### 3.2.1. Expression media trial

Three different growth media were tested with Rosetta2 to see if a change in media components could enhance the expression. Histones were expressed using 2YT, Terrific Broth and Miller media (Figure8).



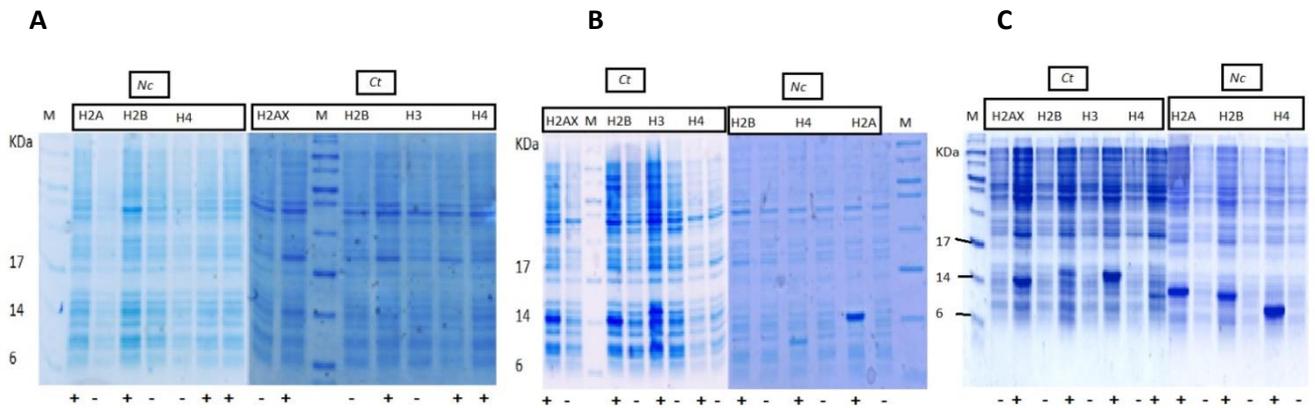
**Figure 8.** Expression media optimization. SDS-Gel showing test expression using *E. coli* strains Rosetta2 in different media: (A) Miller (B) Terrific broth (C) 2YT media showing induction with IPTG (OD 0.6-0.8). Lanes showing uninduced (-) and induced (+). However, expression levels for human H4 (HsH4) were comparatively high.

Most of the histones of the two divergent thermal organisms demonstrated almost the same moderate expression in *E. coli* after our media optimization strategy, whereas *C. thermophilum* H3 and H4 expressed poorly in all cases. This suggests that media choice does not make a large difference to histone production in any of our cases.

### 3.2.2. Expression strain trial

Different *E. coli* strains were tested in order to improve the histone expression levels. Although our sequences were already codon optimized, Rosetta 2(DE3) pLysS contains genetically incorporated tRNAs genes for 7 rare codons whereas the Star pRIL strain codes for only 3. The Star pRIL strain also contains a mutation in the gene encoding RNaseE to enhance mRNA stability (Lopez et al., 1999). The pLysS plasmid carries the gene encoding T7 lysozyme and does not provide codon supplementation.

We compared expressions of each histone in Rosetta2 pLysS, BL21(DE3) pLysS and Star pRIL competent cells using 2YT media. The expression levels in Star pRIL and BL21(DE3) didn't improve significantly for any histone compared to Rosetta2 (Figure 9).



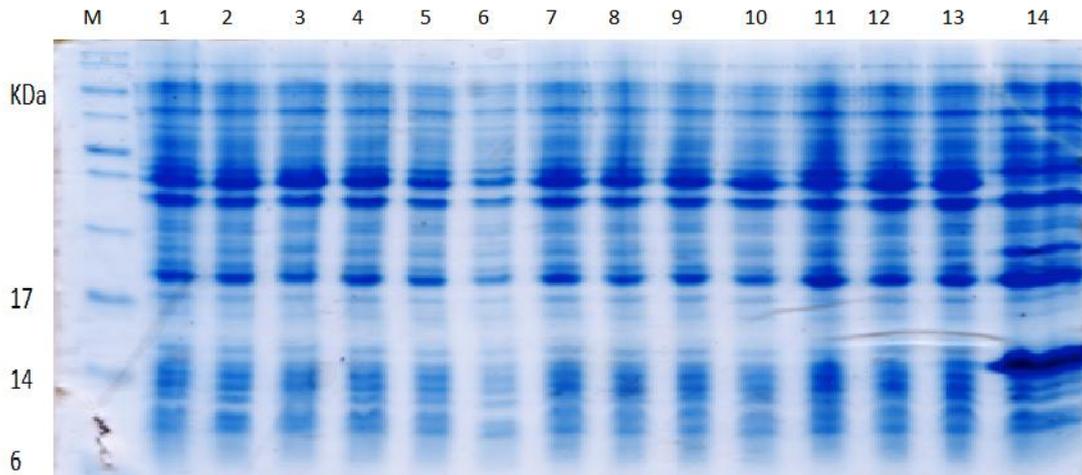
**Figure 9.** Test expression and media optimization. SDS-Gel showing test expression using *E. coli* strains (A) BL21 pLysS (B) Star pRIL (C) Rosetta2 in 2YT media showing 4 h induction. Lanes showing uninduced (-) and induced (+) cultures with IPTG for H2A, H2B, H3 and H4.

Organism	Histone	Gene sequence precedence	Protein lengths	Molecular weight (MW)	Extinction coefficient ( $\epsilon$ )	Test Expression	Production Protein
<i>N. coriiceps</i>	H2A	De novo Synthesized	127	13663,98	4470	+	+
	H2B	De novo Synthesized	125	13760,07	7450	+	+
	H3	Mut. from <i>Hs</i> H3	136	15404,02	4470	+	++
	H4	Same as <i>Hs</i> H4	103	11381,37	5960	+	++
<i>C. thermophilum</i>	H2AX	De novo Synthesized	133	14140,37	4470	+	+
	H2B	De novo Synthesized	140	15089,31	7450	+	+
	H3	De novo Synthesized	136	15391,85	4470	+	-
	H4	De novo Synthesized	103	11386,34	5960	+	-

**Table 5.** Summary table of origin, state, properties and production of each histone.

### 3.2.3. Variation in expression between inoculums

Expression levels can sometimes vary from one colony to another. An experiment was designed to compare inoculums derived from 13 separate single colonies of *C. thermophilum* H3 grown independently for expression. All 13 single colony showed equivalent expression (Figure 10).



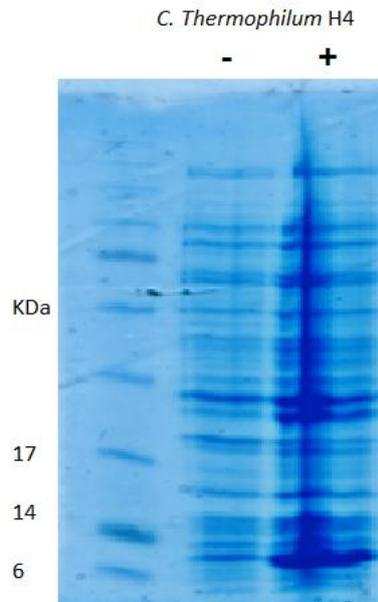
**Figure 10.** Expression trial. 13 single colonies picked from the same plate of *C. thermophilum* H3. In the 14<sup>th</sup> line is the + control with the *N. coriiceps* H3.

### 3.2.4. Optimization of *C. thermophilum* H4 ribosome binding site

We hypothesized that the reason for some cases of poor expression could be in the structure of the plasmid because noticed that there was no space between the RBS and the sequence of the protein (Figure 11). Other histone expression plasmids with much higher expression for proteins of the same weight and complexity have spaces of at least 6 bases from the RBS to the gene sequence. By reducing mRNA folding free energy near the RBS the rates of translation initiation can dominate expression levels of the histone genes (Kudla et al., 2009; de Smit and Van Duin, 1990). Modifying the stability and folding of mRNA secondary structure near the ribosomal binding site by introducing a sequence with reduced GC content favors translation and quantitatively affects expression levels of recombinant proteins (Kudla et al., 2009; Wu et al., 2004).

Mutagenesis PCR was designed for the plasmid of *C. thermophilum* H4 for insert a 7 base pairs (ATATACAT) spacer between the RBS and the ATG codon.





**Figure 12.** SDS PAGE gel comparing expression of the *C. thermophilum* H4. The result of the PCR mutagenesis of *C. thermophilum* H4 in lane 2 (+). In lane 1, negative control (-) of *C. thermophilum* H4 before the mutagenesis. We can appreciate in same growth conditions, media and *E. coli* strains, how the mutagenesis make an improvement in the expression of the protein.

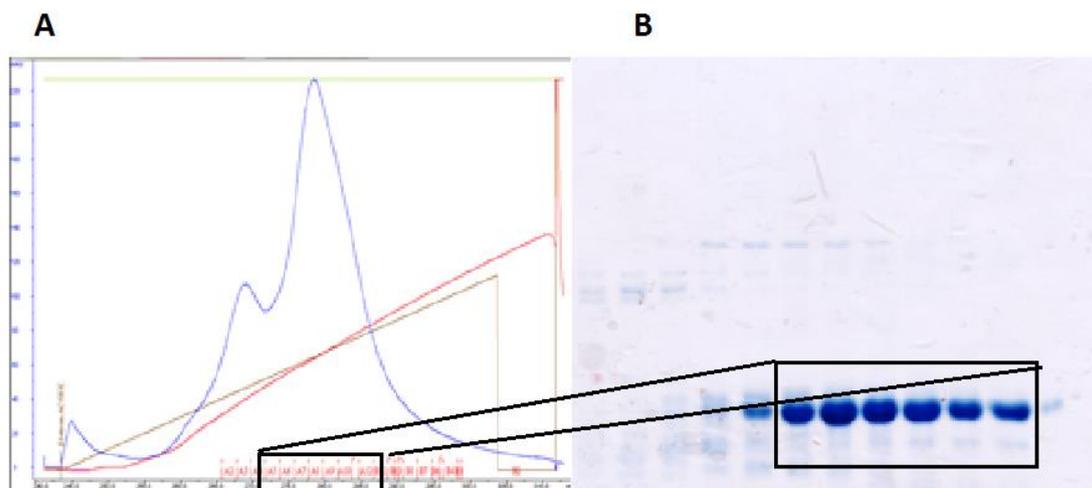
### 3.3. Preparation of histones

#### 3.3.1. Large scale expression

Recombinant histones proteins were expressed in 2 L flask with 1 L media and selection antibiotics chloramphenicol (34 µg/ml) for pLysS together with kanamycin (30 µg/ml) for the expression plasmid pD451-SR or ampicillin (100 µg/ml) for pET3a. After growth and induction by IPTG this media was centrifuged in order to obtain the cell pellet with protein in inclusion bodies due to hydrophobic regions that promote the aggregation of the expressed histones at high concentrations (Kane and Hartley, 1988).

#### 3.3.2. Purification by chromatography

The inclusion bodies were lysed by sonication, and washed by centrifugation then solubilized using guanidine hydrochloride due to the positive net charge of histones. Histones were then purified by cation exchange chromatography using a NaCl gradient for elution. All histones showed a similar behavior, with the most ideal case overall being *N. coriiceps* H3 (Figure 13). SDS PAGE showed that all fractions correspond with the same purity as the chromatogram peak (Figure 6).

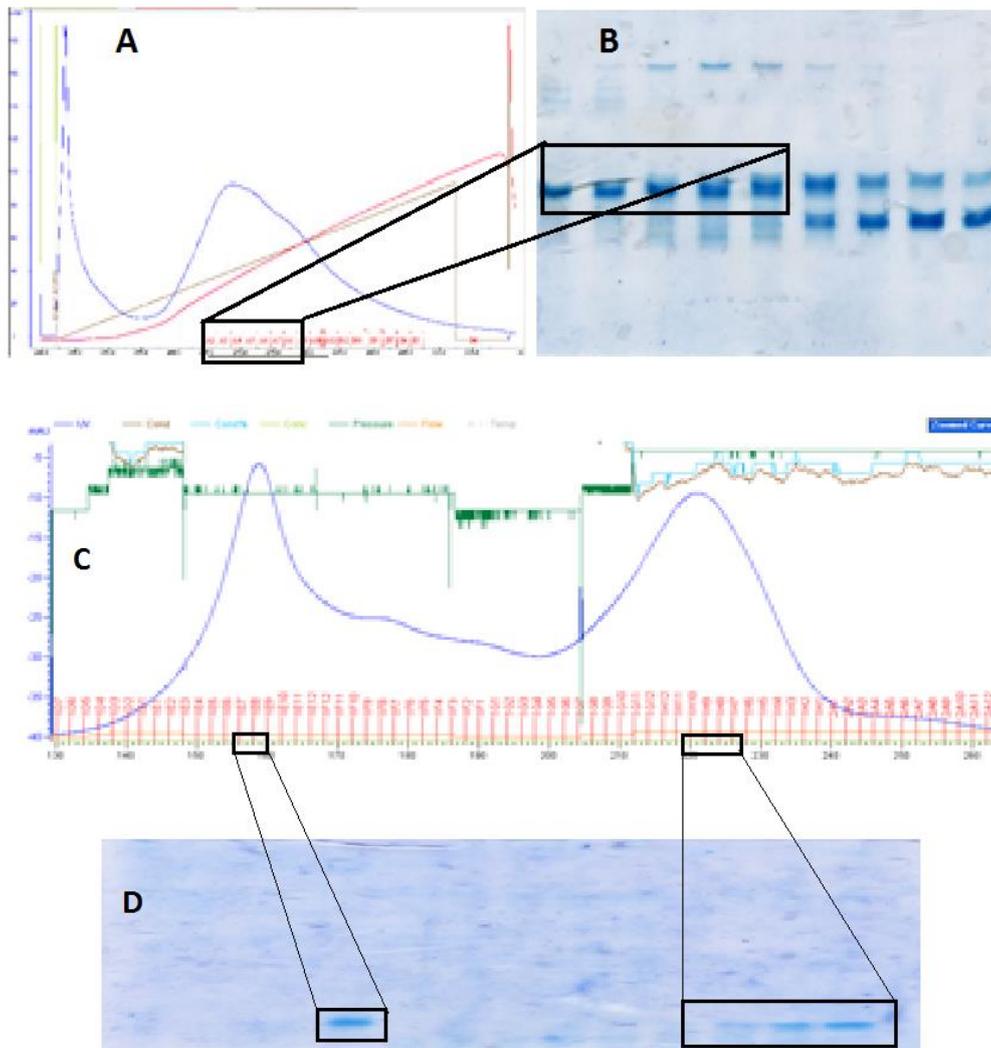


**Figure 13.** Anion exchange chromatography for *N. coriiceps* H3. (A) Chromatogram of the Cation exchange HiTrap SP FF 5mL GE Healthcare column showing in blue the 280Abs referring to the protein concentration in the fractions. (B) Validation SDS PAGE with the fractions of the pure protein indicated in the chromatogram.

Proteins were analyzed by UV spectrophotometry to calculate the 260:280 nm ratio of absorbance, and by SDS-PAGE. The 260:280 nm ratio was used to check for DNA and RNA contamination. An ideal 260:280 ratio value is in the range 0.5 to 0.8 for proteins with little or no DNA contamination. Above this ratio there is excess nucleic acid contamination which leads to subsequent complications because histones are poorly UV absorbing DNA binding proteins. Can interfere in the refolding of the octamers.

### 3.3.3. Purification by gel filtration chromatography

The expression levels of *C. thermophilum* H4 after purification were less than 2mg for a 1 L culture. Purification from inclusion bodies was never reproducible to give sufficient quantities for experiments and led to an excessive 260:280 ratio above 1.5 indicating DNA contamination. To remove DNA contamination, an additional purification over a Superdex 200 10/300 gel filtration column was used in a reducing buffer containing 7 M urea (Figure 14). After this step, the purity of *C. thermophilum* H4 showed an acceptable 260:280 ratio of 0.8.



**Figure 14.** Process column filtration eliminate DNA. (A) Cation exchange chromatography of *C. thermophilum* H4, (B) SDS-page gel validating the protein in the fractions with high rate of DNA contamination (C) DNA contamination which was removed by gel filtration chromatography using urea buffer after cation exchange chromatography purification Superdex 200 10/300 GL column. (D) SDS-page gel validating the protein with free-DNA removal filtration column. Quality controls measures included, apart SDS-PAGE gel, the UV spectrophotometry.

After purification, histones were dialyzed at 4°C to eliminate all urea and salt. The dialysate was frozen at -80°C over night, proteins were then lyophilized in individuals tubes for long term storage at room temperature.

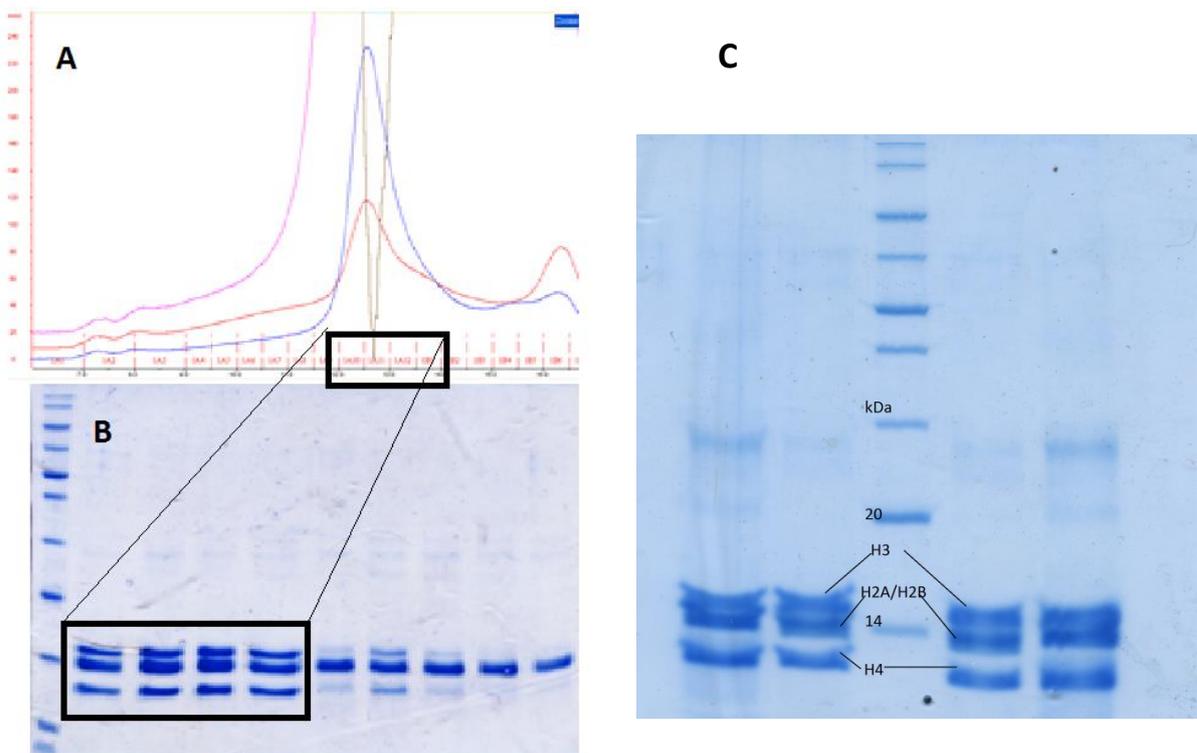
### 3.4. Assembly of octamers refolded from purified histones

The octamer refolding property of histones was compared for histones from two thermal organisms by mixing equal molar amounts of H2A, H2B, H3 and H4 after they were completely unfolded in 7M Guanidium. This mix was then dialyzed against 3 changes of 2M NaCl buffer to

fold histones into octamers (Luger et al., 1997b). It is in 2M NaCl where the histone octamer complex is in the most stable form because it is dominated by the hydrophobic interactions between the histones and by potential cationic repulsions of the highly basic proteins (Arents et al., 1991; Luger et al., 1997a). Octamer complexes of histones were separated from aggregates and other partial complexes using a Superdex 200 10/300 column. A typical chromatogram of octamers shows elution high molecular weight aggregates at 7 mL followed by octamers complexes at 12 mL to 13 mL, and H2A-H2B dimer peak at 15 mL (Figure 7A).

### 3.4.1. *N. coriiceps* octamers

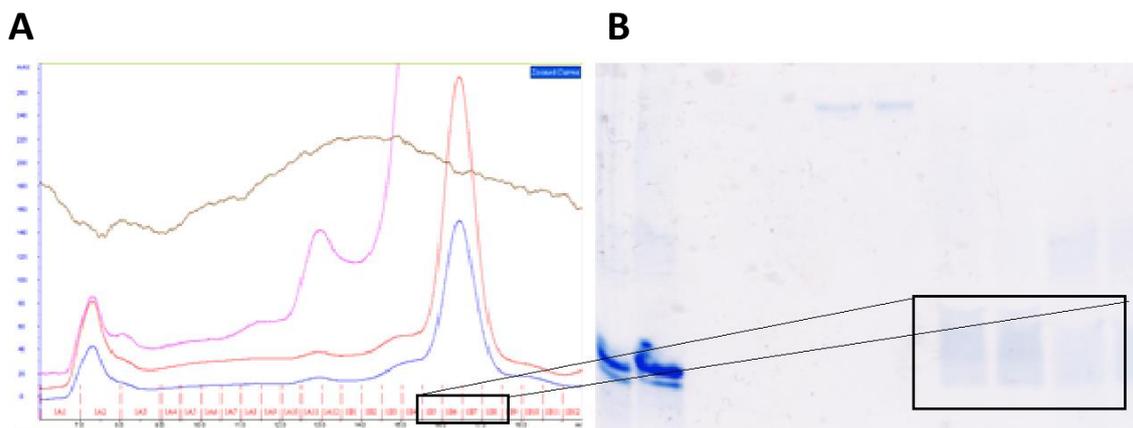
*N. coriiceps* octamer assembly was very successful by mixing in equimolar quantities all four histones and using the gel filtration (Figure 15A) to obtain pure octamer fractions (Figure 15B). The fractions were checked by SDS PAGE (Fig 15C) for the nucleosome assembly. Octamers could be stored at 4°C for several weeks.



**Figure 15.** *N. coriiceps* histones refolded into octamers. (A) Chromatogram with the *N. coriiceps* octamers by column filtration Superdex 200 10/300 GL column. (B) SDS PAGE gel with GFC fractions confirming the presence of individual histones and their stoichiometry corresponding to the octamer peak at 12mL elution volume. The composition of the four histones in this octamer complex is qualitatively proportional as expected resulting in perfect yield of histone octamer. (C) Nucleosomes of *N. coriiceps* and *X. laevis*.

### 3.4.2. *C. thermophilum* histones do not refold into octamers

The peak elution fraction resulting from octamer gel filtration is 11-12 ml, but this zone did not show any octamers by SDS PAGE for *C. thermophilum* refolding (Figure 8A). This suggests that the small peak at 11 mL corresponds to a complex different than octamer, possibly due to limited aggregation of H3, H4 or H2A, H2B histones or from DNA (Figure 16). A reason for the problem of *C. thermophilum* not forming octamers could be the high contamination of DNA in the histones production making it very difficult to perform exact quantification by UV absorbance of the amount protein and secondly, difficult the histone interaction and assembly into octamers in all our attempts. We were unable to refold *C. thermophilum* octamers. Consequently, *C. thermophilum* nucleosomes could not be formed. The inability of *C. thermophilum* histones to form octamers could maybe show that the histone octamer assembly must be mediated by DNA sequence as *C. thermophilum* histones are unable to refold into stable tetramers and octamers in absence of DNA by standard salt dialysis method.



**Figure 16.** *C. thermophilum* histones do not refold in octamers. (A) Chromatogram with the *N. coriiceps* octamers by column filtration Superdex 200 10/300 GL column. The GFC analysis shows a very small peak at 11mL that corresponds to a structure larger than the octamer, suggesting an inability of *C. thermophilum* to form stable octamers. Large aggregates elute first at 7mL followed by dimers and tetramers form 16mL. (B) *C. thermophilum* histone refolding almost entirely resulted in precipitation; we can appreciate two bands in Lanes 1 and 2 from the raw sample before load to the GFC. The rest of the SDS Page could not validate the other peaks that were registered in the chromatogram. The occurrence of these aggregates and absence of any octamer peak were consistent in all trials despite many refolding attempts.

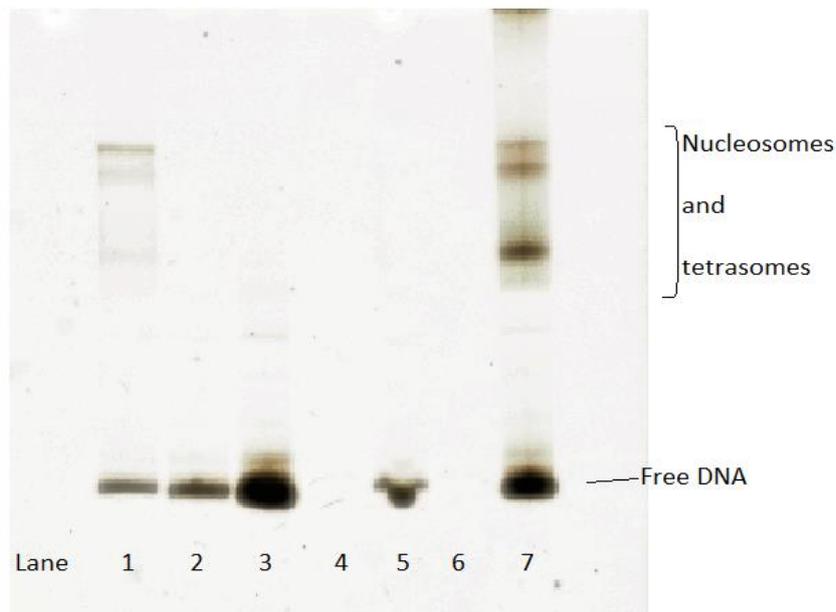
### 3.5. Nucleosome assembly

Nucleosomes can be assembled *in vitro* using histone octamer complexes along with a DNA sequence of 147 bp or greater by salt gradient dialysis. We have used the very well characterized nucleosome positioning DNA sequence 54A54 (255 bp). Purified octamers were mixed with DNA at 2 M NaCl and then sequentially dialyzed in reducing salt concentrations to physiologically relevant level 0.1 M NaCl. When the salt concentration is reduced to 1 M, the

octamer dissociates into  $(H3-H4)_2$  tetramer and two  $H2A-H2B$  dimers, and then at 0.85 M NaCl DNA binds to the tetramer, and at 0.65 M NaCl,  $H2A-H2B$  dimers bind on either side of the tetramer to form the nucleosome (Yager, 1989; Germond et al., 1976). The reason to choose this method is that it is highly efficient to avoid non-specific histone–DNA aggregation when mixing pure octamer with DNA.

### 3.5.1. *N. coriiceps* nucleosomes in vitro by salt dialysis

The ability of *N. coriiceps* octamers to form nucleosomes on 147 bp of the 54A54 sequence of DNA was tested. Purified octamer fractions from gel filtration chromatography of *N. coriiceps* and control *X. laevis* were mixed with DNA at 1:1 ratio and dialyzed by reducing the salt concentration. Nucleosome samples were then validated by native PAGE (Figure 17). Native gel interpretations of nucleosome positioning are not always obvious because the flexibility of DNA at entry/exit angles is variable (Pennings et al., 1991).



**Figure 17.** *N. coriiceps* Nucleosomes and tetrasomes on 54A54 DNA sequence. Native PAGE showing migration of tetrasomes and nucleosomes of *N. coriiceps* and *X. laevis* on 54A54 fluorescence labeled by salt dialysis. First and second band are for nucleosomes and di-tetrasomes respectively and third band is free-DNA. *X. laevis* was loaded in Lane 1, *N. coriiceps* was loaded in Lane 7 and Free-DNA was loaded in Lane 3 as negative control.

The native PAGE of histone assembly with free DNA showed the capability to form nucleosomes with in both cases, although with an excess of unbound DNA. The mobility of the nucleosome bands of *N. coriiceps* equivalent to *X. laevis*.

## 4. DISCUSSION

### 4.1. Histone selection and gene design

There is a general assumption that at the chromatin level all eukaryotes package and process their genomes in an equivalent fashion. We wanted to include two divergent thermal organisms that would test eukaryotic adaptation of chromatin to temperature to allow us to investigate the genome packaging mechanism. To compare relevant nucleosome properties, it was important that the genomes of divergent species be completely sequenced to ensure representative canonical histones were used, avoiding divergent isoforms or variants.

The classification of histones is mainly dependent on their amino acid sequence identity with mammals. Alignment of histone sequences from divergent thermal eukaryotes with *H. sapiens* shows that *C. thermophilum* histones are somewhat divergent, with only 65% identity in H2B. There is very high similarity of H3 and H4 between *N. coriiceps* and *H. sapiens*. H4 is widely considered to be the most conserved histone and only differs by a few amino acids animals, plants and most fungi.

However, there are many differences in the amino acid positions of *C. thermophilum* in H3 and H4 histones that can be potentially critical for formation of histone tetramer interfaces and hence for octamer stability. Histone sequence alignments of divergent eukaryotes to higher eukaryotes showed significant differences in the regions that may be critical for histone-histone interaction in octamer and histone-DNA interactions in nucleosomes.

### 4.2. Protein expression and purification

The expression of histones is sensitive sequence and to growth conditions in *E. coli*. We were able to create histone genes for our two divergent organisms by using gene optimization calculations and gene synthesis technology. Sufficient amounts of histones were produced for the biochemical assays by recombinant protein expression in *E. coli*, to produce typically 4 mg from a 1,5 L culture. Testing expression in different strains of *E. coli* and using different media did not produce any obvious changes in yield. However, we found that expression of *C. thermophilum* H4 was improved by optimizing the plasmid to suit *E. coli* to ensure that the RBS sequence is correctly placed. Contamination of DNA was always a problem and we always took care, to avoid excess DNA in the protein sample. Purification using gel filtration was necessary to remove DNA in one case.

### 4.3. Octamers refolding and nucleosome

*C. thermophilum* histones did not form octamers and mostly precipitated, although some histone dimers remained in solution. Inspection of the amino acid sequence of the histones reveals four changes in amino acid sequence in the H4, at A49, S61, S65 and S84 that differ from *N. coriiceps*, *H. sapiens* and *X. laevis*. The mutation of these residues could affect in the interaction of H4 with H3, making the conformation less stable and this disfavoring octamer stability. This could be the reason *C. thermophilum* does not form octamer despite the conservation of the rest of the amino acid sequence.

Native PAGE of *N. coriiceps* histones assembled on 54A54 DNA shows they have the capability to form mono-nucleosomes. There were differences in the mobility of these nucleosome, with *N. coriiceps* bands being more diffuse suggesting they are more dynamic. The diffuse nucleosome bands could also suggest that the DNA is not tightly wrapped around octamer, with DNA less tightly bound at the entry/exit site of the nucleosome affecting hydrodynamic behavior (Flaus et al., 1996).

The inability of *C. thermophilum* histones to form octamers prevented formation of nucleosomes by the normal salt dialysis technique. Instead, a salt-urea dialysis approach to assemble nucleosomes from individual histones could be used in future experiments.

## 5. CONCLUSIONS

Histone sequences, even though are highly conserved, have significant level of diversity across the Eukaryotic kingdom. This shows that histone evolution may be more complicated than previously assumed and the biochemical properties of divergent histones may be explicable based on structural interactions. We used an efficient method to produce recombinant histones in *E. coli* in large quantities. Heterologous protein expression of histones can be improved by optimizing gene sequences and growth conditions. For most of the histones, genes synthesis with codon adaptation to *E. coli* and altering the 5' end of mRNA structure lead sufficiently to high level expression. For poorly expressing histones, optimization of growth conditions by testing *E. coli* strains or growth media didn't show any positive effect on expression levels. It is important for the correct expression of some proteins that a space exists between RBS and the sequence of that protein. A common scale up and purification strategy that could be applied to all the histones to achieve sufficient amounts of protein for biochemical assays. The results from octamer refolding at high ionic strength show differences in complex formation for the two divergent thermal organisms. *N. coriiceps* histones show similar octamer refolding behavior to *X. laevis*. *C. thermophilum* histone largely precipitated and did not form octamers by salt dialysis. Results from nucleosome assembly show that histone octamers from *N. coriiceps* and *X. laevis* were able to assemble into nucleosomes by standard methods of salt dialysis. There was difference in the mobility nucleosomes from *N. coriiceps* and *X. laevis*, demonstrating distinctive hydrodynamic properties of nucleosomes. This suggests some variation in binding of DNA around the octamer either in a static or dynamic mode.

## 6. APPENDINX

### 6.1. Histone Genes sequences

#### *C. thermophilum*

##### **CtH2AX:**

ATGACCGGTGGCAAAGCGGTGGCAAAGCATCGGGTACAAAATCAGCGCAGTCTCGTTCCTCAAAGCTGGCCTG  
GCCTTCCCGGTGGTTCGTGTTACCGTTTGGTGCCTAAAGGTAATTATGCACAGCGCGTTGGTGCAGGCGGCCA  
GTATACCTGGCGGCGGTTCTGGAATATCTGGCTGCCGAAATCCTGGAAGTGGCTGGTAATGCGGCCCGTGACAAT  
AAGAAAACCCGTATCATTCCGCGCCACCTGCAGCTAGCTATTCGCAACGATGAGGAGCTGAATAAACTGCTCGGC  
CATGTGACCATCGCACAGGGCGGTGTACTGCCGAATATCCATCAGAACCTGCTGCCGAAAAAGACTGCTAAGACG  
GGAAAAAACTTAAGCCAAGAACTTTAATGGTTGAGGT

##### **CtH2B:**

ATGGCACCGAAAGCAGACGCGCAGAAAAAACCGGCATCCAAGGCACCAGCATCTACCGCTAGCAAGGCTCCGTCT  
GAAAAAAAAGATGCGGGTAAAAAACTGCCGCTAGCGGTGAAAAAAAAGCGTACCAAAGCGCGTAAAGAAA  
CCTACAGCTCGTATATTTATAAAGTTCTTAAGCAAGTACATCCGGACACTGGCATCTCTAATCGTGCTATGTCTATC  
CTGAATTCATTGTTAATGATATTTTCGAGCGCGTGGCCACGGAGGCGTCTAAACTGGCGGCCTACAATAAAAAAA  
GCACCATCTCTCCCGTAAATCCAGACCGCTGTTGCTTAATCTGCCTGGTGAAGTCCGGAACACGCGGTATC  
CGAAGGCACCAAGGCTGTCACTAAATATAGTTCTAGCACGAAATAAT

##### **CtH3:**

ATGGCTCGTACTAAACAGACTGCGCGTAAATCCACCGGTGGCAAAGCTCCTCGCAAACAGCTGGCGTCTAAAGCA  
GCCCCGAAATCTGCACCGAGCACTGGTGGCGTGAAAAAACACACCGCTATAAACCGGGCACGGTAGCGCTGCGT  
GAAATTCGCCGCTACCAGAAGAGTACCGAGCTGTTGATCCGTAAGTTACCGTTTCAACGTTCTCGTTCGTGAGATCG  
CCCAGGATTTCAAAGCGACCTGCGTTTCCAATCTCAGCGATCGGTGCACTGCAGGAAAGCGTTGAATCCTATCT  
GGTCTCGTGTTCGAAGATACCAACCTTTGTGCAATCCATGCCAAACGTGTGACCATTAGTCTAAGGACATCCAG  
CTGGCGCGTCTGAGAGGAGAACGTAATTAAT

##### **CtH4:**

ATGACTGGTCTGGTAAAGGTGGTAAGGGTCTGGGTAAAGGAGGCGCAAACGTACCGTAAAATCCTTCGCGA  
CAATATCCAAGGCATCAAAAACCGGCTATCCGCCGCTGGCCCGCCGCGGGCGGTGTTAAACGTATTTACGCGATG  
ATCTATGAGGAAACCGTGGCGTCTGAAATCCTTTCTGAAAAGCGTTATTCGCGACGCCGTTACCTACACCGAAC  
ATGCGAAACGTAAGACTGTGACTTCTGATGTAGTGTACGCATTAACGTCAGGGCCGTACCCTCTATGGCTT  
CGGGGGTAAAT

#### *N. coriiceps*

##### **NcH2A:**

ATGTCTGGTCTGGCAAACCGGTGGTAAAGCGCGTGCAAAGCAAACCCGTTTCGAGCCGTGCAGGTTTGCA  
ATTCCCGGTGGTCTGTGACCCGCATCTGCGCAAAGGCAACTACGCTACCGCGTAGGCGCGGGGGCACCAGT  
CTATCTGGCCCGGTTCTCGAATACCTGACTGCGGAAATCCTGGAGCTGGCTGGCAATGCTGCTCGAGATAATAAA  
AAAACCTCGGATCATTCTCGTACCTGCAGCTGGCGTTGTAATGACGAAGAACTCAACAAGCTGCTGGGTGGC  
GTCACAATCGCGCAGGGTGGCGTTCTTCCGAATATCCAGGCGGTGCTGCTGCCGAAGAAGACGGAAAAAGCTGC  
CAAAAAATAAT

##### **NcH2B:**

ATGCCAGAAGCAGCATCTGTAAGACCTAAAAAGGCTCTAAAAAGCGGTTACCAAACCTCCGTCCAAAACC  
GGTAAGAAACGCCGTAAGTCCCGTAAAGAGTCGTATGCGATCTACGTCTATAAAGTATGAAGCAGGTGCACCCG  
GATACGGGCATTAGCAGCAAGGCGATGGGTATCATGAATTGTTTCGTATCCGACATCTTCGAACGTATCGCTGGT  
GAAGCCTCTCGCTTAGCGCACTACAACAAACGTAGTACCATTACCAGCAGGGAGATCCAGACTGCTGTTGCTCTCC  
TGCTGCCGGGCGAACTTGCTAAACATGCCGTTTCTGAAGGTACGAAAGCAGTCACTAAATATACCTCTTCCAAATA  
AT

##### **NcH3:**

ATGGCCCGTACCAAGCAGACCGCCCGTAAATCCACCGGAGGGAAGGCTCCCCGCAAGCAGCTGGCCACCAAGGC  
AGCCAGGAAGTCCGCTCCTGCTACCGGCGGAGTCAAGAAACCTACCGTTACCGGCCCGCACAGTGCCTCTCCG  
CGAGATCCGCCGCTACCAGAAATCCACCGAGCTGCTCATCCGCAAACCTGCTTTCCAGCGCCTGGTCCGGGAGATC  
GCTCAGGACTTCAAGACCGACCTGCGCTTCCAGAGCTCGGCCGTTATGGCTCTGCAGGAGTCCaGCGAGGCTTATC  
TGGTCTGGTCTTTGAGGACACCAACCTGTGCGCATCCACGCCAAGAGGGTACCATCATGCCAAGGACATCCA  
GCTGGCCCGCAGAATCCGAGGCGAGAGGGCT

**NcH4/HeH4:**

ATGTCTGGTCTGGTAAAGGTGGCAAAGGTCTGGGTAAAGGTGGTGGCAAACGTCATCGTAAATCCTGCGCGA  
TAACATTACAGGCATTACCAAACAGCCATCCGTCGTTGGCTCGCCGCGGCGGCGTTAAGCGTATCTCGGGCCTG  
ATCTATGAAGAGACTCGCGGCGTACTGAAGGTGTTCTTAAAAATGTAATCCGTGACGCGGTTACCTATACTGAAC  
ACGCGAAACGAAAACCGTTACGGCAATGGACGTCGTGTACGCTCTCAAACGTCAGGGCCGTACACTGTACGGGT  
TCGGTGGTTAAT

**H. sapiens**

**HsH2AX:**

ATGTCAGGTCGCGGTAACACTGGCGGGAAAGCCCGTGCAGAAAGCGAAATCGCGCAGTTCCTGCGGGTCTGCA  
GTTCCGGTAGGTCGTGTGCATCGCCTTCTGCGCAAAGGCCACTATGCGGAACGTGTGGGTGCTGGTGCTCCGGT  
CTACTTAGCAGCCGTGTTGGAGTACCTGACAGCCGAAATCCTGGAACGTCAGGTAATGCAGCACGTGACAACAA  
GAAAACCCGCATCATTCCACGCCATCTGCAGTTAGCCATTGCGAACGATGAGGAACTGAACAACTCCTGGGAGG  
CGTACCATTTGCGCAAGGCGGCGTTTTGCCAATATCCAGGCCGTTCTTCTCCGAAGAAAACGAGCGCTACGGTA  
GGACCGAAAGCGCCTTCTGGTGGCAAGAAAGCGACCCAAGCGAGC

**HsH2A:**

ATGTCAGGAAGAGGCAAACAAGGCGGTAAAGCTCGCGTAAGGCCAAGACTCGCTCATCTCGGGCTGGGCTACA  
GTTCCCTGTTGGCCGTGTTACCGGCTGTTAAGGAAAGGCAATTATTCCGAGCGGGTGGGAGCTGGAGCTCCAGT  
CTATCTGGCTGCAGTGTGGAGTATCTGACCGCTGAGATTTTGAATTGGCCGGGAATGCGGCCGTGATAACAA  
GAAGACTCGCATTATCCCAGACACCTGCAGCTCGCTATCCGCAACGATGAGGAACTGAACAACTGCTCGGAAG  
AGTACTATCGCTCAGGGCGGGGTCCTGCCAACATCCAGGCTGTGCTGCTGCCAAGAAAACCGAGAGTACCA  
CAAAGCCAAGGGTAAG

**HsH2B:**

ATGCCAGAACCGCCAAGTCCGCTCCAGCCCCGAAGAAAGGCTCCAAGAAAGCGGTGACCAAGGCTCAGAAGAA  
AGACGGGAAAAAACGCAAACGTTCCAGGAAGGAGAGTTATTCCGTTTACGTGTACAAGGTGCTGAAGCAGGTGC  
ACCCCGATACCGGCATCTCGTCCAAGGCCATGGGCATCATGAACTCCTTTGTCAACGATATCTTTGAGCGCATCGC  
AGGGGAAGCCTCCCGCCTGGCTCATTACAACAAGCGCTCCACCATCACCTCCCGGGAGATCCAGACCGCGGTCCG  
ACTGCTGCTGCTGGGAGTTGGCCAAACACGCCGTGCCGAGGGCACCAAGGCTGTACCAAGTACACCAGCGC  
GAAG

**HsH3:**

ATGGCCCGTACCAAGCAGACCGCCCGTAAATCCACCGGAGGGAAAGGCTCCCCGCAAGCAGCTGGCCACCAAGGC  
AGCCAGGAAGTCCGCTCCTGCTACCGGCGGAGTCAAGAAACCTCACCGTTACCGGCCCGGCACAGTCTGCTCCG  
CGAGATCCGCGCTACCAGAAATCCACCGAGTGTCTATCCGCAAACGCTTTCCAGCGCTGGTCCGGGAGATC  
GCTCAGGACTTCAAGACCGACCTGCGCTTCCAGAGCTCGGCCGTTATGGCTCTGCAGGAGGCTGCGAGGCTTAT  
CTGGTGGTCTCTTTGAGGACACCAACCTGTGCGCCATCCACGCCAAGAGGGTACCATCATGCCAAGGACATCC  
AGCTGGCCCGCAGAATCCGAGGCGAGAGGGCT

**HsH4:**

ATGTCTGGTCTGGTAAAGGTGGTAAAGGTCTGGGTAAAGGTGGTCTAAACGTCACCGTAAAGTTCTGCGTGAC  
AACATCCAGGGTATACCAAGCCGGCTATCCGTCGTCTGGCTCGTCTGGTGGTAAACGTATCTCCGGTCTGA  
TCTACGAAGAAACCCGCGGTGTTCTGAAAGTTTTCTGGAAAACGTTATCCGTGACGCTGTTACCTACCCGAACA  
CGCTAAACGTAACCGTTACCGCTATGGACGTTGTTACGCTCTGAAACGTCAGGGTCTACCTGTACGGTTTC  
GGTGGT

**X. laevis**

**XIH2A:**ATGTCAGGAAGAGGCAAACAAGGCGGTAAAACCCGCGCTAAGGCCAAGACTCGCTCATCTCG  
GGCTGGGCTACAGTTCCTGTTGGCCGTGTTACCCGGCTGTTAAGGAAAGGCAATTATGCAGAGCGGG  
TGGGAGCTGGAGCTCCAGTCTATCTGGCTGCAGTGTGGAGTATCTGACCGCTGAGATTTTGAATTG  
GCCGGGAATGCGGCCGTGATAACAAGAAGACTCGCATTATCCCAGACACCTGCAGCTCGCTGTGCG  
CAACGATGAGGAACTGAACAACTGCTCGGAAGAGTCACTATCGCTCAGGGCGGGTCTGCCCAACA  
TCCAGTCCGTGCTGCTGCCAAGAAAACCGAGAGTTCCTAAGTCCGCAAGAGCAAG

**XIH2B:**

ATGGCCAAGTCCGCTCCAGCCCCGAAGAAAGGCTCCAAGAAAGCGGTGACCAAGACTCAGAAGAAAGACGGGAA  
AAAGCGCAGGAAGACAAGGAAGGAGAGTTATGCCATTTACGTGTACAAGGTGCTGAAGCAGGTGCACCCCGATA  
CCGGCATCTCGTCCAAGGCCATGAGCATCATGAACTCCTTTGTCAACGATGTGTTTGAAGCGCATCGCAGGGGAAG

CCTCCCGCCTGGCTCATTACAACAAGCGCTCCACCATCACCTCCCGGGAGATCCAGACCGCGGTCCGACTGCTGCT  
GCCTGGGGAGTTGGCCAAACACGCCGTGTCCGAGGGCACCAAGGCTGTACCAAGTACACCAGCGCCAAG

**XIH3:**

ATGGCCCGTACCAAGCAGACCGCCCGTAAATCCACCGGAGGGAAGGCTCCCCGCAAGCAGCTGGCCACCAAGGC  
AGCCAGGAAGTCCGCTCCTGCTACCGGCGGAGTCAAGAAACCTCACCGTTACCGGCCCGGCACAGTCGCTCTCCG  
CGAGATCCGCCGCTACCAGAAATCCACCGAGCTGCTCATCCGCAAACCTGCCTTCCAGCGCCTGGTCCGGGAGATC  
GCTCAGGACTTCAAGACCGACCTGCGCTTCCAGAGCTCGGCCGTTATGGCTCTGCAGGAGGCCAGCGAGGCTTAT  
CTGGTCGCTCTCTTTGAGGACACCAACCTGTGCGCCATCCAGCCAAGAGGGTCACCATCATGCCAAAGGACATCC  
AGCTGGCCCGCAGAATCCGAGGCGAGAGGGCT

**XIH4:**

ATGTCTGGTCTGGTAAAGGTGGTAAAGGTCTGGGTAAAGGTGGTCTAAACGTCACCGTAAAGTTCTGCGTGAC  
AACATCCAGGGTATACCAAGCCGGCTATCCGTCGCTGGCTCGTCTGGTGGTAAACGTTATCTCCGGTCTGA  
TCTACGAAGAAACCCCGGTTCTGAAAGTTTTCTGGAAAACGTTATCCGTGACGCTGTACCTACCCGAACA  
CGCTAAACGTAACCCGTTACCGCTATGGACGTTGTTTACGCTCTGAAACGTCAGGGTCTACCCTGTACGTTTC  
GGTGGT

## 6.2. Histone Protein sequences

### *C. thermophilum*

**CtH2AX:**

MTGGKSGGKASGTSKSAQSRSSKAGLAFVGRVHRLLRKGNYAQRVAGAPVYLAADVLEYLAAEILELAGNAARDNKKT  
RIIPRHLQLAIRNDEELNKLGHVTIAQGGVLPNIHQNLPPKKTAKTGKNSQEL

**CtH2B:**

MAPKADAQKKPASKAPASTASKAPSEKKDAGKKAASGEKKRRTKARKETYSSYIKVLKQVHPDGTISNRAMSILNSFV  
NDIFERVATEASKLAAYNKSTISSREIQTAVRLLPGELAKHAVSEGTKAVTKYSSSTK

**CtH3:**

MARTKQTARKSTGGKAPRKQLASKAARKSAPSTGGVKKPHRYKPGTVALREIRRYQKSTELLIRKLPFQRLVREIAQDFK  
SDLRFQSSAIGALQESVESYLVSLFEDTNLCAIHAKRVTIQSKDIQLARRLRGERN

**CtH4:**

MTGRGKGGKGLGKGGAKRHRKILRDNIQGITKPAIRRLARRGGVKRISAMIYEETRGVLKSFLESVIRDAVITYTEHAKRK  
TVTSLDVVYALKRQGRTLYGFGG

### *N. Coriiceps*

**NcH2A:**

MSGRGKTGGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYAHRVAGAPVYLAADVLEYLTAEILELAGNAARDNKKT  
RIIPRHLQLAVRNDDEELNKLGGVTIAQGGVLPNIQAVLLPKKTEKAAKK

**NcH2B:**

MPEAASVKAPKKGSKKAVTKTPSKTGKRRRSRKESYAIYVYKVMKQVHPDGTISSKAMGIMNCFVSDIFERIAGEASRL  
AHYNKRSTISSREIQTAVRLLLPGELAKHAVSEGTKAVTKYTSSK

**NcH3:**

MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTELLIRKLPFQRLVREIAQDFK  
TDLRFQSSAVMALQESSEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA

**NcH4:**

MSGRGKGGKGLGKGGAKRHRKILRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLKVFLENVIRDAVITYTEHAKRKT  
VTAMDVVYALKRQGRTLYGFGG

### *H. sapiens*

**HsH2AX:**

MSGRGKTGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNHYAERVAGAPVYLAADVLEYLTAEILELAGNAARDNKKTRI  
IPRHLQLAIRNDEELNKLGGVTIAQGGVLPNIQAVLLPKKTSATVGPAPSGGKATQAS

**HsH2A:**

MSGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYSERVAGAPVYLAADVLEYLTAEILELAGNAARDNKKTR  
IIPRHLQLAIRNDEELNKLGRVTIAQGGVLPNIQAVLLPKKTESHHKAKGK

**HsH2B:**

MPEPAKSAPAPKKGSKKAVTKAQKKDGKKRKRSRKESYSVYVYKVLKQVHPDTGISSKAMGIMNSFVNDIFERIAGEAS  
 RLAHYNKRSTITSREIQTAVRLLLPGLAKHAVSEGTKAVTKYTSAK

**HsH3:**

MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTELLIRKLPFQRLVREIAQDFK  
 TDLRFQSSAVMALQEACEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIGERA

**HsH4:**

MSGRGKGGKGLGKGGAKRHRKVLDRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGLVKVFLENVIRDAVYTEHAKRK  
 TVTAMDVVYALKRQGRPLYGFGG

**X. laevis**

**XIH2A:**

MSGRGKQGGKTRAKAKTRSSRAGLQFPVGRVHLLRKGNYAERVGAGAPVYLAAVLEYLTAEILELAGNAARDNKKTR  
 IIPRHLQLAVRNDEELNKLLGRVTIAQGGVLPNIQSVLLPKKTESKSAKSK

**XIH2B:**

MAKSAPAPKKGSKKAVTKTKQKDGKKRRKTRKESYAIYVYKVLKQVHPDTGISSKAMSIMNSFVNDVFERIAGEASRLA  
 HYNKRSTITSREIQTAVRLLLPGLAKHAVSEGTKAVTKYTSAK

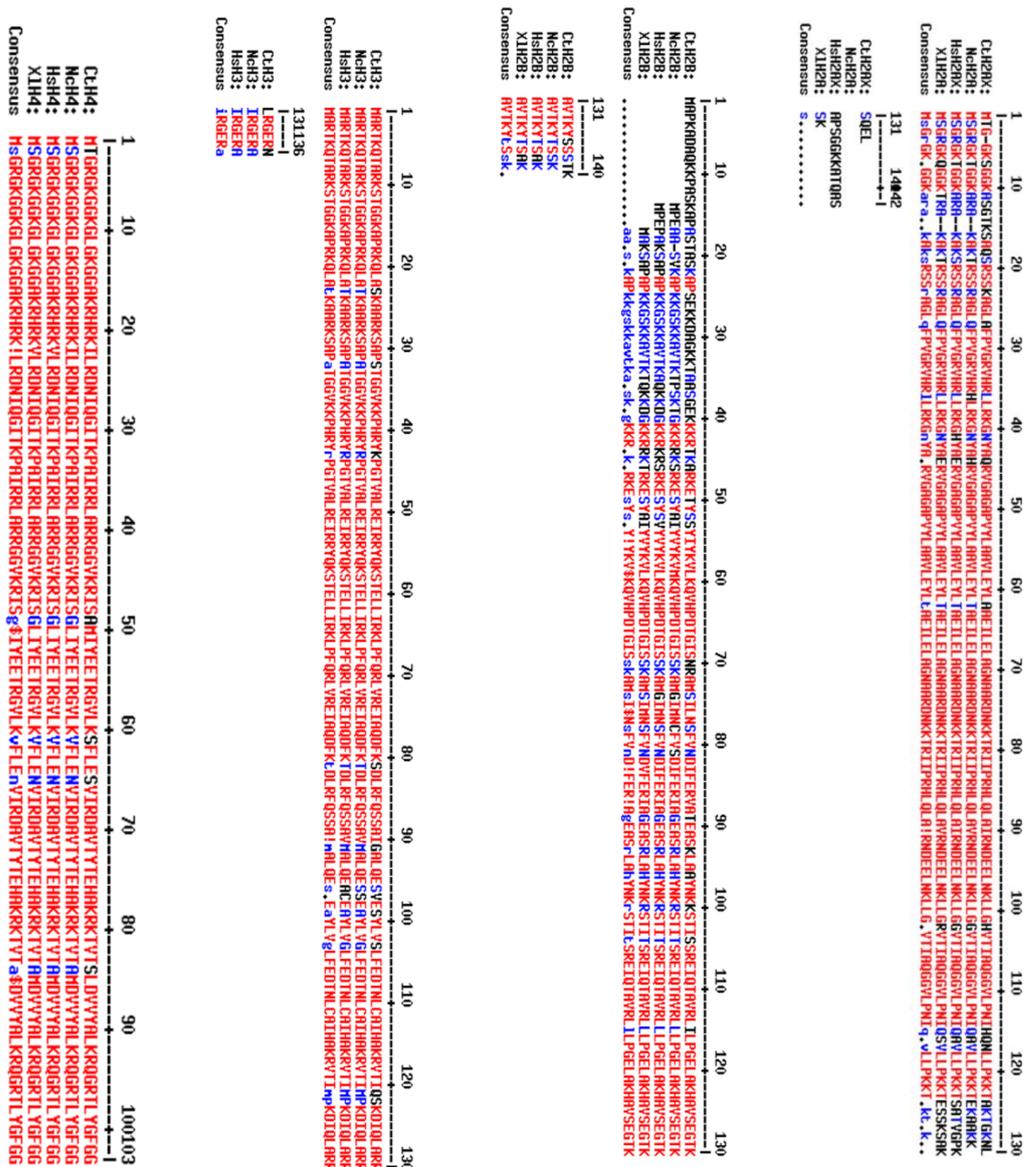
**XIH3:**

MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTELLIRKLPFQRLVREIAQDFK  
 TDLRFQSSAVMALQEASEAYLVLFEDTNLCAIHAKRVTIMPKDIQLARRIGERA

**XIH4:**

MSGRGKGGKGLGKGGAKRHRKVLDRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGLVKVFLENVIRDAVYTEHAKRK  
 TVTAMDVVYALKRQGRPLYGFGG

**6.3. Histone proteins alignments**



## 7. REFERENCES

- Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, et al. 2012. The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology*. 59:429–514.
- Arents, G., Burlingame, R. W., Wang, B. C., Love, W. E., & Moudrianakis, E. N. (1991). The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix. *Proceedings of the National Academy of Sciences*, 88(22), 10148-10152.
- Becktel, W. J., & Schellman, J. A. (1987). Protein stability curves. *Biopolymers*, 26(11), 1859-1877.
- Beers, J. M., & Jayasundara, N. (2015). Antarctic notothenioid fish: what are the future consequences of 'losses' and 'gains' acquired during long-term evolution at cold and stable temperatures?. *Journal of Experimental Biology*, 218(12), 1834-1845.
- Bruno M, Flaus A, Stockdale C, Rencurel C, Ferreira H, Owen-Hughes T. 2003. Histone H2A/H2B dimer exchange by ATP-dependent chromatin remodeling activities. *Molecular Cell*. 12:1599–1606.
- de Smit, M. H., & Van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proceedings of the National Academy of Sciences*, 87(19), 7668-7672.
- DORAN, J (2013). *Histones and other nucleosomes proteins*. Prof. Andrew Flaus, Phd. Galway: National University of Ireland Galway
- Dubendorff, J. W., & Studier, F. W. (1991). Creation of a T7 autogene: Cloning and expression of the gene for bacteriophage T7 RNA polymerase under control of its cognate promoter. *Journal of molecular biology*, 219(1), 61-68.
- Flaus A, Luger K, Tan S, Richmond TJ. 1996. Mapping nucleosome position at single base-pair resolution by using site-directed hydroxyl radicals. *Proceedings of the National Academy of Sciences*. 93:1370–1375.
- Flaus, A., & Owen-Hughes, T. (2011). Mechanisms for ATP-dependent chromatin remodelling: the means to the end. *Febs Journal*, 278(19), 3579-3595.
- Germond, J. E., Bellard, M., Oudet, P., & Chambon, P. (1976). Stability of nucleosomes in native and reconstituted chromatins. *Nucleic acids research*, 3(11), 3173-3192.
- Henaut, A., & Danchin, A. (1996). Analysis and predictions from Escherichia coli sequences, or E. coli in silico. *Escherichia coli and Salmonella: cellular and molecular biology*. ASM Press, Washington, DC, 2047-2066.
- Kane, J. F., & Hartley, D. L. (1988). Formation of recombinant protein inclusion bodies in Escherichia coli. *Trends in Biotechnology*, 6(5), 95-101.

- Kellner, N., Schwarz, J., Sturm, M., Fernandez-Martinez, J., Griesel, S., Zhang, W., ... & Hurt, E. (2016). Developing genetic tools to exploit *Chaetomium thermophilum* for biochemical analyses of eukaryotic macromolecular assemblies. *Scientific reports*, 6, 20937.
- Klosin, A., Casas, E., Hidalgo-Carcedo, C., Vavouri, T., & Lehner, B. (2017). Transgenerational transmission of environmental information in *C. elegans*. *Science*, 356(6335), 320-323.
- Krenek, S., Petzoldt, T., & Berendonk, T. U. (2012). Coping with temperature at the warm edge—patterns of thermal adaptation in the microbial eukaryote *Paramecium caudatum*. *PLoS One*, 7(3), e30598.
- Kudla, G., Murray, A. W., Tollervey, D., & Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in *Escherichia coli*. *science*, 324(5924), 255-258.
- Lanzer, M., & Bujard, H. (1988). Promoters largely determine the efficiency of repressor action. *Proceedings of the National Academy of Sciences*, 85(23), 8973-8977.
- Lopez, P. J., Marchand, I., Joyce, S. A., & Dreyfus, M. (1999). The C-terminal half of RNase E, which organizes the *Escherichia coli* degradosome, participates in mRNA degradation but not rRNA processing in vivo. *Molecular microbiology*, 33(1), 188-199.
- Luger, K., Dechassa, M. L., & Tremethick, D. J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?. *Nature reviews Molecular cell biology*, 13(7), 436-447.
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648), 251-260.
- Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B., & Müller-Hill, B. (1994). Quality and position of the three lac operators of *E. coli* define efficiency of repression. *The EMBO journal*, 13(14), 3348.
- Oudet, P., Gross-Bellard, M., & Chambon, P. (1975). Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell*, 4(4), 281-300.
- PATWAL, I (2014). Properties of divergent histones. Prof. Andrew Flaus, PhD. Galway: National University of Ireland Galway
- Pennings, S., Meersseman, G., & Bradbury, E. M. (1991). Mobility of positioned nucleosomes on 5 S rDNA. *Journal of molecular biology*, 220(1), 101-110.
- Pierce BA. 2003. *Genetics: A conceptual approach*. New York : W.H. Freeman.
- Pinto DMS, Flaus A. 2010. Structure and function of histone H2AX. In: *Genome Stability and Human Diseases*, New York: Springer, volume 50, pp. 55–78.
- Pollard TD, Earnshaw WC, Lippincott-Schwartz J. 2007. *Cell Biology*. Elsevier Health Sciences, 2nd edition.
- Rustad, D. P. S. (1930). *Euphausiacea: With Notes on Their Biogeography and Development*. I kommisjon hos Jacob Dybwad.

Sadler, J. R., Sasmor, H., & Betz, J. L. (1983). A perfectly symmetric lac operator binds the lac repressor very tightly. *Proceedings of the National Academy of Sciences*, 80(22), 6785-6789.

Shin, S. C., Ahn, D. H., Kim, S. J., Pyo, C. W., Lee, H., Kim, M. K., ... & Edwards, D. (2014). The genome sequence of the Antarctic bullhead notothen reveals evolutionary adaptations to a cold environment. *Genome biology*, 15(9), 468.

Van den Brink, J., Facun, K., De Vries, M., & Stielow, J. B. (2015). Thermophilic growth and enzymatic thermostability are polyphyletic traits within Chaetomiaceae. *Fungal biology*, 119(12), 1255-1266.

van Noort, V., Bradatsch, B., Arumugam, M., Amlacher, S., Bange, G., Creevey, C., ... & Bork, P. (2013). Consistent mutational paths predict eukaryotic thermostability. *BMC evolutionary biology*, 13(1), 7.

Wu, Z., Irizarry, R. A., Gentleman, R., Martinez-Murillo, F., & Spencer, F. (2004). A model-based background adjustment for oligonucleotide expression arrays. *Journal of the American statistical Association*, 99(468), 909-917.

Yager, T. D., McMurray, C. T., & Van Holde, K. E. (1989). Salt-induced release of DNA from nucleosome core particles. *Biochemistry*, 28(5), 2271-2281.