Journal Of Proteomics
February 2021, Volume 232 Pages 104044 (11p.)
https://doi.org/10.1016/j.jprot.2020.104044
https://archimer.ifremer.fr/doc/00657/76938/

Lysine-specific acetylated proteome from the archaeon *Thermococcus gammatolerans* reveals the presence of acetylated histones

Alpha-Bazin Béatrice [1], Gorlas Aurore [2], Lagorce Arnaud [2, 3], Joulié Damien [1], Boyer Jean-Baptiste [1], Dutertre Murielle [2], Gaillard Jean-Charles [1], Lopes Anne [2], Zivanovic Yvan [2], Dedieu Alain [1], Confalonieri Fabrice [2], Armengaud Jean [1, *]

[1] Université Paris-Saclay, CEA, INRAE, Département Médicaments et Technologies pour la Santé (DMTS), SPI, 30200 Bagnols-sur-Cèze, France.
[2] Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198 Gif-sur-Yvette, France
[3] IHPE, Université de Montpellier, CNRS, Ifremer, Université de Perpignan, Via Domitia, Perpignan, France.

* Corresponding author : Jean Armengaud, email address : jean.armengaud@cea.fr

Abstract :

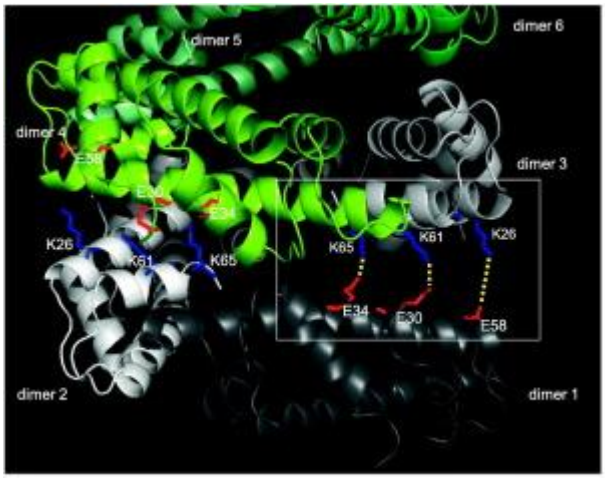
Thermococcus gammatolerans EJ3 is an extremophile archaeon which was revealed as one of the most radioresistant organisms known on Earth, withstanding up to 30 kGy gamma-ray radiations. While its theoretical proteome is rather small, T. gammatolerans may enhance its toolbox by post-translational modification of its proteins. Here, we explored its extent of Nε-acetylation of lysines. For this, we immunopurified with two acetylated-lysine antibodies the acetylated peptides resulting from a proteolysis of soluble proteins with trypsin. The comparison of acetylated proteomes of two archaea highlights some common acetylation patterns but only 4 out of 26 orthologous proteins found to be acetylated in both species, are acetylated on the same lysine site. We evidenced that histone B is acetylated in T. gammatolerans at least at two different sites (K27 and K36), and a peptide common at the C-terminus of histones A and B is also acetylated. We verified that acetylation of histones is a common trait among Thermococcales after recording data on Thermococcus kodakaraensis histones and identifying three acetylated sites. This discovery reinforces the strong evolutionary link between Archaea and Eukaryotes and should be an incentive for further investigation on the extent and role of acetylation of histones in Archaea.

Significance

Acetylation is an important post-translational modification of proteins that has been extensively described in Eukaryotes, and more recently in Bacteria. Here, we report for the first time ever that histones in Archaea are also modified by acetylation after a systematic survey of acetylated peptides in Thermococcus gammatolerans. Structural models of histones A and B indicates that acetylation of the identified modified residues may play an important role in histone assembly and/or interaction with DNA.

The in-depth protein acetylome landscape in T. gammatolerans includes at least 181 unique protein sequences, some of them being modified on numerous residues. Proteins involved in metabolic processes, information storage and processing mechanisms are over-represented categories in this dataset, highlighting the ancient role of this protein post-translational modification in primitive cells.

Graphical abstract



Highlights

► We identified 338 acetylated sites by mass spectrometry. ► Some common acetylation patterns are found among archaea. ► Histones from two different archaea were evidenced for the first time to be acetylated. ► Structural models of histones A and B indicates that acetylation may play a key role.

Keywords : Proteome, Post-translational modification, Acetylome, Archaea, Histone

Introduction

Reversible protein N^{ε}-Lysine acetylation, a post-translational modification (PTM), was first discovered on histones [1, 2]. Since these first studies, lysine acetylation has been recognized as an abundant and important PTM and found conserved in the three domains of life [3, 4]. It influences a wide variety of essential biological processes [5]. In prokaryotes, the reversibility of this PTM is catalyzed by a KAT family acetyltransferase and different deacetylases (KDACs, sirtuins or serine hydrolases). In addition, a non-reversible N^{α}-lysine acetylation catalyzed by N-terminal acetyltransferases may also occur at the N-terminal of proteins [6, 7]. However, the functional implication of N^{α}-lysine acetylation remains unclear for prokaryotes [8, 9].

Based on shotgun proteomics, in-depth exploration of the acetylome from several organisms belonging to the different domains of life has been reported [8]. In these studies, most bacterial acetylated proteins are involved in metabolic processes, particularly the control of central metabolism, but also in transcription, translation, virulence, adaptation, and stress responses [10-12]. All these results indicate that protein acetylation is a common trait in bacteria and is involved in a broad range of cellular functions.

While acetylomes in *Bacteria* have been well documented, only a few studies have been reported for *Archaea*. A [2Fe-2S] ferredoxin from *Halobacterium salinarum* (previously named *Halobacterium halobium*) [13] and the chromatin protein Alba (standing for “acetylation lowers binding affinity”) from *Sulfolobus solfataricus* [14] were the first archaeal proteins identified with an internally acetylated lysine. The residue 16 of Alba can be acetylated by the PatA

acetyltransferase and deacetylated by an archaeal Sir2 homologue, thereby mediating transcriptional repression [14]. It has been suggested that the acetylation state of Alba influences the level of chromatin packaging in *Sulfolobus* [4]. Nevertheless, recent studies have shown that this lysine residue is not acetylated but extensively methylated [15]. The same applies for the Alba homolog from the closely-related archaeon *Sulfolobus islandicus* [7]. The substitution of this lysine by an arginine residue did not modify cell growth but triggered transcription of a few genes [15]. Recent surveys of different archaeal acetylomes from *Halobacterium salinarum*, *Natronomonas pharaonis*, *Haloferax volcanii* and *Sulfolobus islandicus* revealed that N^{α}-lysine acetylation is an abundant modification in these organisms and that methylation on lysine side chains is the most frequent PTM found in *S. islandicus* [7, 16, 17]. By contrast, using an immunoaffinity enrichment and tandem mass spectrometry, Liu *et al.* [18] identified 1,017 acetylated N^{ε}-lysine sites in 643 proteins from *Haloferax mediterranei*. Among the proteins detected as acetylated, some enzymes are involved in glycolysis, tricarboxylic acid (TCA) cycle, poly(3-hydroxybutyrate-co-3-hydroxyvalerate) biosynthesis from acetyl-CoA and propionyl-CoA. Many other proteins are involved in key biological processes, such as transcription and replication. Interestingly, the mutation of acetylated site in *H. mediterranei* replication initiation protein Cdc6A destroyed the Autonomous Replicating Sequence (ARS) activity of its adjacent origin oriC1 [18].

Thermococcus gammatolerans EJ3 was isolated from samples collected from hydrothermal chimneys located in the mid-Atlantic Ridge and the Guyamas basin [19]. *T. gammatolerans* is one of the most radioresistant organisms known on Earth, as it withstands 5 kGy of radiation without any detectable lethality [20]. *T. gammatolerans* genome is composed of a circular chromosome of 2.045 Mbp without any extra-chromosomal element, coding for 2,157 proteins [21]. *T. gammatolerans* was described as an obligatory anaerobic heterotroph organism that grows optimally at 88°C in the presence of sulfur or cystine on yeast extract, tryptone and

peptone, producing H_2S. Here, we established the protein lysine-specific acetylome survey from exponentially growing phase *T. gammatolerans* cells taking into account acetylated-lysine peptide enrichment by two different antibodies. We detected in *T. gammatolerans* 316 N^{ε}-Lysine peptides, representing 338 different acetylation sites, from 181 unique proteins. We report here for the first time that *Thermococcus* histones may be acetylated on different positions.

Materials and Methods

Strains, media and growth conditions. *T. gammatolerans* EJ3 was grown in serum bottles under anaerobic conditions at 85°C in complex organic medium (VSM) supplemented with S° (2 g/l) as previously described [20]. Serum bottles were inoculated at a cellular density of 5 × 10^5 cells/ml and incubated at 85°C for 16 h. Cells were harvested at a cellular density of 8-9 × 10^7 cells/ml corresponding to the transition phase between the late exponentially growing phase and the beginning of the stationary phase.

Protein extraction. A lysis buffer containing a complete, EDTA-free, protease inhibitor cocktail (Roche, 1 tablet per 50 ml), phosphatase inhibitor cocktails 2 and 3 (Sigma-Aldrich, 1: 100 vol/vol, each), 8 M urea, 75 mM NaCl, 50 mM Tris (pH 8.2), 50 mM β-glycerophosphate, 1 mM sodium orthovanadate, 10 mM sodium pyrophosphate, 1 mM $MgCl_2$, and Benzonase 25 KU (1: 10,000, vol/vol) was prepared. Four independent biological samples were prepared as follows. *T. gammatolerans* cells were resuspended in 3.5 ml of lysis buffer per gram of cells, disrupted by sonication and centrifugated at 13,000 g for 20 min at 4°C. Supernatants were collected, and protein concentrations were measured using the Bradford method.

Immunodetection of acetylated lysines by Western Blotting. Each protein lysate (20 µg) was resolved by SDS-PAGE on a 4-12% Nu-PAGE gel in 5% MOPS buffer (Novex) and transferred to a 0.45-µm polyvinylidene difluoride membrane (Thermo Scientific Pierce) by means of a MiniTrans-Blot, Electrophoretic Transfert Cell (Bio-Rad). Blocking of non-specific binding was achieved by placing the membrane in a 4% (w/v) BSA solution in TBST buffer (Trizma Base 25mM, Glycine 192 nM, EtOH 10%, pH 8.5, Tween20 0.1%) for 6 h. Acetylated lysines were detected using either an anti-acetyl-lysine rabbit polyclonal antibody (ICP380, ImmuneChem) at dilution 1:1000 or an anti-acetyl-lysine rabbit monoclonal antibody (Cell Signalling Technologies) at dilution 1:4000 overnight at 4°C. After washing with TBST buffer three times for 5 min, an horseradish peroxidase conjugated anti-rabbit IgG (dilution 1:2000) was added for 1 h incubation at room temperature. All antibody solutions were prepared in TBST buffer with 4% bovine serum albumin. Immobilon Western Chemiluminescent HRP Substrate (Millipore) was used for membrane revelation. Finally, detection was performed by means of a 45 sec exposure in a Hyperprocessor developing apparatus (Amersham).

Microwave-assisted in solution proteolysis. For each *T. gammatolerans* proteome sample, 0.8 mg of protein material per microcentrifuge tube was reduced by 10 mM dithiothreitol (DTT) at 37 °C for 1 h. A 50 % formic acid aqueous solution was then added at a 6 % (v/v) final concentration. Tubes were sealed, placed in a glass water bath, and exposed to 750 W microwave energy in an oven for 2.5 min. After denaturation the tubes were placed on melting ice. A volume of 0.5 M Tris-Base was added per tube to bring pH to 8.0 - 8.5. The fractions were then alkylated by 20 mM iodoacetamide by incubation for 30 min in the dark at room temperature. Proteolysis was performed by adding trypsin solution (0.5 µg/µl in 0.01 % trifluoroacetic acid (TFA)) at 2:100 enzyme to protein mass ratio (w/w) and an overnight incubation at 37°C. Proteolysis was stopped by lowering the pH to 4.0 with a 50% formic acid

aqueous solution. The proteolysis yields were monitored by SDS-PAGE carried out on a 4-12 % Bis-Tris gradient NuPAGE gel run as previously described [22]. Tryptic digests were desalted using Sep-Pak Plus Cartridge C18 (Waters), eluted in 65% CH_3CN / 35% H_2O / 0.1% TFA. The eluates were dried using a SpeedVac and then dissolved in 50 mM Tris-HCl buffered at pH 8.0, and containing 150 mM NaCl and 1 mM EDTA (ETN buffer) at 5 mg protein digest/ml.

Enrichment of acetylated-lysine peptides by immunocapture. Two different antibodies were used to enrich acetylated-lysine containing peptides. An affinity-purified anti-acetyl-lysine antibody was used from a commercial preparation (ICP0388; ImmunoChem Pharmaceuticals Inc.) consisting of the pan-specific antibody covalently bound to agarose beads. Tryptic digests were added to 40 µl of pan antibody-agarose beads and incubated overnight at 4°C. The beads were washed three times with ETN buffer consisting of 50 mM Tris-HCl (pH 8.0), 100 mM NaCl, and 0.5% NP40. The bound peptides were eluted from the beads with 2 x 30 µl of Glycine-HCl 50mM, pH 2.5. The acetylated lysine monoclonal antibody (#9814; Cell Signaling Technologies) was incubated with Protein A agarose beads (ICP1001; ImmuneChem Pharmaceuticals Inc.) for 4h at 4°C. Supernatant was removed and the beads were washed twice with ETN buffer. Tryptic digest was then added to these beads and incubated overnight at 4°C. After washing three times with ETN buffer the bound peptides were eluted from the beads with 2 x 30 µl of Glycine-HCl 50mM, pH 2.5 buffer. Combined eluates were desalted on ZipTip C18 (Millipore) according to the manufacturer's instructions and dried in a SpeedVac. Immuno-enriched lysine acetylated peptides were solubilized in 40 µL of TFA 0,1% prior to nano-liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS) analysis.

NanoLC-MS/MS analysis of the immune-enriched peptides. The identification was performed by nanoLC-MS/MS using a LTQ-Orbitrap XL hybrid mass spectrometer (ThermoFisher) coupled to an UltiMate 3000 LC system (Dionex-LC Packings), essentially as previously described [23]. Immune-enriched peptides (10 µl) were loaded and desalted online on a reverse phase Acclaim Pepmap 100 C18 precolumn (5 µm bead size, 100 Å pore size, 300 µm i.d. × 5 mm; LC Packings). Peptides were resolved on a nanoscale Acclaim Pepmap100 C18 reverse phase capillary column (3 µm bead size, 100-Å pore size, 75 µm i.d. × 15 cm**;** LC Packings) operated at a flow rate of 0.3 µl.min^{-1} along a 120-min CH_3CN gradient from 0 to 40 % solvent B (100 % CH_3CN, 0.1 % HCO_2H), solvent A being 0.1 % HCO_2H. Full-scan mass spectra were measured from m/z 300 to 1800 with the LTQ-Orbitrap XL mass spectrometer operated in the data-dependent mode. A TOP7 strategy was used, consisting in a very accurate scan in the Orbitrap analyzer (30,000 resolution; internal calibration) while the 7 most abundant precursor ions detected in the Orbitrap pre-scan were fragmented in the linear ion trap with dynamic exclusion of previously selected ions. A normalized collision energy of 35% was used for the fragmentation.

NanoLC-MS/MS characterization of chromatin-bound proteins. The detection of peptide mixtures resulting from trypsin digestion of chromatin proteins extracted from *T. kodakaraensis* was performed on a Q-Exactive HF mass spectrometer (ThermoFisher) coupled to an UltiMate 3000 LC system (Dionex-LC Packings) operated at a flow rate of 0.3 µL.min^{-1} essentially as described [24]. Peptides were resolved on a 50 cm nano scale AcclaimPepmap100 C18 reverse phase capillary column with the same 120 min gradient described here above. A Top 20 strategy was applied in data dependent acquisition mode. Full scan mass spectra were acquired from m/z 350 to 1800 with an AGC target set at 3×10^6 ions and a resolution of 60,000. MS/MS scan was initiated when the ACG target reached 10^5 ions with a threshold intensity of

1.7 x10^5 and potential charge states of 2^+ and 3^+. Ion selection was performed applying a dynamic exclusion window of 10 sec.

Peptide-to-MS/MS spectrum assignation. Peak lists were generated as described previously [23]. Using the MASCOT (2.2.06) search engine (Matrix Science), MS/MS spectra were searched against the protein coding sequences from *T. gammatolerans* (2,157 entries totaling 636,564 residues, [21]). MS/MS assignments were performed with the following parameters: i) mass tolerance of 5 ppm on the parent ion, ii) mass tolerance of 0.5 or 0.02 Da for fragment ions from MS/MS acquired with the LTQ orbitrap or the Q Exactive HF, respectively; iii) carbamidomethylated Cys (+57.0215) as static modifications; and iv) oxidized methionine (+15.9949) and acetylation of amino group (+42.0105) as variable modification. The maximum number of missed cleavages was set at 2. All peptide matches with a peptide score with a p-value below 0.05 were filtered.

Extraction of chromatin proteins from *T. kodakaraensis*. *T. kodakaraensis* cells were cultured at 85°C in VSM medium supplemented with sodium butyrate (NaBu) at 5 mM. Cultures in mid-exponential phase or late stationary phase were centrifuged at 4,000 g for 20 min. The pellets were treated following the protocol developed by Maruyama et al [25] with slight modification. Briefly, cells were frozen at -80°C for 1 h and then lysed in a solution containing 25 mM HEPES (pH 7.0), 15 mM $MgCl_2$, 100 mM NaCl, 0.4 M Sorbitol, 0.5% Triton X100, and 0.5 mM NaBu during 20 min at 4°C. The lysates were centrifuged at 2,350 g for 20 min. The resulting pellets containing the chromatin associated proteins were washed with lysis buffer, then centrifuged at 12,000 g for 20 min and stored at -20°C. Chromatin associated proteins were then resolved by SDS-PAGE on NuPAGE 4-12% Bis-Tris gel (Thermofisher) with MES as running buffer (Invitrogen). Gels were stained with Coomassie Blue. After

destaining, the low-molecular weight band of each gel lane, containing histone-sized proteins of the sample, was sliced. The resulting polyacrylamide gel pieces were processed by in-gel proteolysis with trypsin Gold (Promega) in the presence of 0.01% ProteaseMax additive (Promega), as previously described [26]. The resulting peptide digests were analyzed by nanoLC-MS/MS.

T. gammatolerans **and** ***H. mediterranei*** **orthologous protein inference.** *T. gammatolerans* (NC_012804) and *H. mediterranei* (NC_017941) protein coding sequences were all blasted without SEG filtering (BLASTP) against each other. Among the various alignment statistics, bit score was chosen in order to obtain significant similarity information. Hence, maximum bit score (max bitscore) was defined as the bit score obtained by aligning one sequence with itself, setting the 100% of max bitscore value. Plotting the distribution of percent of max bitscore (% max bitscore) obtained by BLASTing one population of sequences against another (not shown), allows generally to identify regions of orthologous and paralogous (multicopy genes, gene families) peaks on the distribution plot. A threshold of 35% for (reciprocal, in both direction) max bitscore to accept orthologous relation between two sequences was defined. Sequence pairs of orthologous proteins, as defined above, from *T. gammatolerans* and *H. mediterranei,* in which at least one member is known to be acetylated, were aligned with MultAlin [27]. Shared positions of un-acetylated, hemi-acetylated (lysine residue position conserved in two homologous proteins only found acetylated in one protein) and acetylated lysine residues at equivalent position on both protein sequences were identified.

Histone structural models. Structural models of homodimers (HTkA/HTkA and HTkB/HTkB) were generated with Swiss-Model [28] using as template the X-ray structure of HPhA from *P. horikoshii,* (PDB code: 1ku5). The ratios of sequence identity shared by HPhA

with HTkA and HTkB are 89% and 85%, respectively. The quality of both models is very high (GMQE score of 0.94 and QMEAN score of 2.01). In order to generate a model of the histone superhelix (involving either homodimers of HTkA or HTkB), the 3D models of HTkA or HTkB homodimers were aligned onto the two layers of the HMfB superhelix from *M. fervidus* (i.e. six consecutive homodimers of HMfB; PDB code: 5t5k). The sequence identity ratio between histones from *M. fervidus* and *T. kodakarensis* is 54%. The RMSD after superposition of homodimers of HMfB with either homodimers of HTkA or HTkB is 0.73Å in both cases.

MS/MS data repository. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [29] partner repository with the dataset identifier PXD016987 and 10.6019/PXD016987.

Results and Discussion

Massive lysine acetylation evidenced in *T. gammatolerans* by immunodetection.

We first investigated the occurrence of acetylation on proteins by immunodetection on Western blot using two antibodies raised against acetylated lysines. **Figure 1** shows the western-blot immunostaining of proteins extracted from *T. gammatolerans* cells. Acetylated proteins were revealed with either acetylated-lysine rabbit polyclonal antibodies (**Figure 1,** lanes 1-3) or an acetylated-lysine rabbit monoclonal antibody (**Figure 1**, lanes 4-6). In both cases, different bands spanning a wide mass range of proteins were detected, signing massive lysine acetylation of proteins. Noteworthy, the polyclonal antibodies revealed the presence of four main protein bands, abundantly detected (Lanes 1 and 2), while the monoclonal antibody did not reveal strong differences between protein profiles.

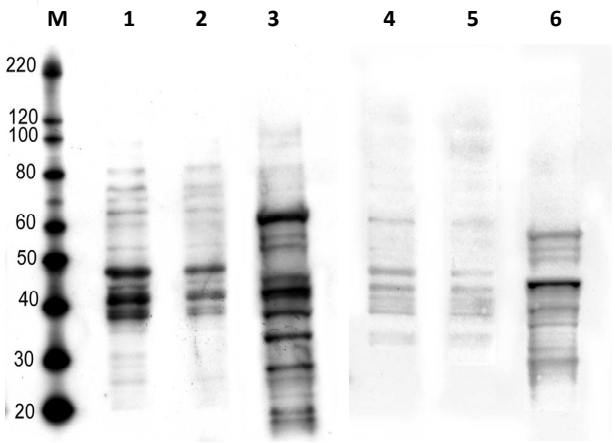


Figure 1. Detection of lysine-acetylated proteins. Western blotting analysis using either anti-acetylated-lysine rabbit polyclonal antibody (PAB) with 20 µg (Lane 1), 10 µg (Lane 2) of extracted proteins of T. gammatolerans and 15 µg (Lane 3) of extracted proteins of Saccharomyces cerevisiae, or anti-acetylated-lysine rabbit monoclonal antibody (MAB) with 20 µg (Lane 4), 10 µg (Lane 5) of extracted proteins of T. gammatolerans and 15µg (Lane 6) of extracted proteins of S. cerevisiae. The molecular weights of the protein markers are indicated in kDa (Lane M).

247

248 **Shotgun proteomic screening of lysine acetylome.**

249 A shotgun proteomic strategy was implemented in order to establish the catalogue of lysine-
250 specific acetylated peptides from four independent *T. gammatolerans* biological cultures. We
251 experienced that *T. gammatolerans* proteome is refractory to classical trypsin proteolysis due
252 to the thermophilic trait of this organism that led to selection of numerous ionic interactions at
253 the surface of proteins to rigidify their structures. For this reason, we applied an *in*-solution
254 digestion protocol with trypsin including microwave denaturation of proteins prior to
255 proteolysis [30]. A systematic immune-based affinity purification of acetyl-lysine containing
256 peptides from the four samples was performed. As it has been already suggested, recognition
257 of acetylated peptides is more efficient from peptide digest than from a whole protein mixture
258 because the modified residue will not be buried in peptides [31]. Two different antibodies were
259 used for enrichment of acetylated-lysine peptides as specificities between antibodies may be
260 different [32]. Finally, peptides from enriched fractions from four biological replicates were
261 identified by means of nanoLC-MS/MS with a high-resolution mass spectrometer, and lysine

acetylation sites were pointed by querying the MS/MS dataset for a 42.01057 Da addition on lysine which is indicative of an acetyl moiety on the ε-amino group of the lysine side chain. High accuracy mass spectrometry with our set-up allows identification of the peptide parent mass with a precision below 5 ppm, which is enough to exclude the possibility of a tri-methylation of the lysine residue (+ 42.04695 Da). Most of the identifications matched their theoretical mass within 1 ppm highlighting the quality of the dataset obtained with the LTQ-orbitrap XL mass spectrometer. As our objective was not to compare amongst different conditions but get a large overview of acetylome, we analyzed hereafter the whole dataset. Amongst the 66,036 MS/MS spectra recorded for the four biological samples and the two immunopurification protocols, we unambiguously assigned 11,142 MS/MS spectra leading to the identification of 1,324 peptide sequences (**Supplementary Tables S1 & S2**). Amongst these, 2,227 MS/MS spectra correspond to 339 acetylated lysine sites from 181 unique acetylated proteins (**Supplementary Tables S3 & S4)**. The number of acetylated proteins found in this dataset represents 8.4 % of the *T. gammatolerans* theoretical proteome. This ratio is lower than what was reported for the haloarchaeon *H. mediterranei* (17.3%), but higher than for most bacteria (*Erwinia amylovora* 2.7%, *Thermus thermophilus* 5.7%, *Geobacillus kaustophilus* 3.1%, *Bacillus subtilis* 4.4%, and *Mycobacterium abscessus* 2.3-5.9%). The percentage of acetylated proteins found in these prokaryotes seems not to be directly in accordance with the complexity of the theoretical proteome since larger genomes often exhibited smallest acetylome extents and may be a consequence of the experimental set up. When comparing the number of acetylated peptides identified from immunoprecipitation experiments performed with either the polyclonal antibodies or the monoclonal antibody, a total of 33 peptides were found common, while 198 and 127 were identified specifically for the polyclonal antibodies or the monoclonal antibody immune-enrichment, respectively. This result confirms that a better coverage of the acetylome can be achieved by increasing the efforts for

287 the sample purification with a broader array of affinity reagents. As previously shown in other
288 prokaryotes, a significant number of *T. gammatolerans* proteins (45%) contained multiple
289 acetylation sites (**Figure 2**). Remarkably, the translation elongation factors EF-1-Alpha
290 (TGAM_1054) and EF-2 (fusA, TGAM_1397), and the phosphoenolpyruvate synthase
291 (TGAM-1043) were found with 10, 8 and 7 modifications, respectively. Interestingly, two
292 peptides isolated from TGAM_0191 and TGAM_0460 are acetylated twice. The peptide
293 KPIKVETVLK isolated from TGAM_0460 which is an acetyl-CoA C-acetyltransferase
294 converting two acetyl-CoA molecules into an acetoacetyl-CoA, a substrate for lipid synthesis,
295 was found to be acetylated on the two lysines located at positions +1 and +4 (**Supplementary**
296 **Table S5**). These data strongly suggested that several *T. gammatolerans* proteins may be poly-
297 acetylated, at least in exponentially growing phase cells.
298 In a previous work [21], we have identified a large ratio of the soluble proteome of *T.*
299 *gammatolerans* validating 951 proteins by nanoLC-MS/MS analysis. We estimated the relative
300 abundance of these proteins by calculating the normalized abundance spectral count factor
301 (NSAF) for each protein. We compared the list of acetylated proteins with these previous data
302 in order to evaluate whether they belong to the most abundant proteins. Among the 181
303 acetylated proteins found here, 175 were detected in this previous study and most of them (62%)
304 were shown to belong to the most abundant proteins (global NASF greater than 0.09% in the
305 previous study). A limited number of acetylated proteins were found to be amongst the less
306 abundant entities of the proteome from the standard condition. Among these, a poorly
307 characterized conserved hypothetical protein (TGAM_2047) and the pre-mRNA splicing,
308 snoRNA binding protein (TGAM_0464) represent only 0.0075% and 0.0045% of the proteome,
309 respectively.

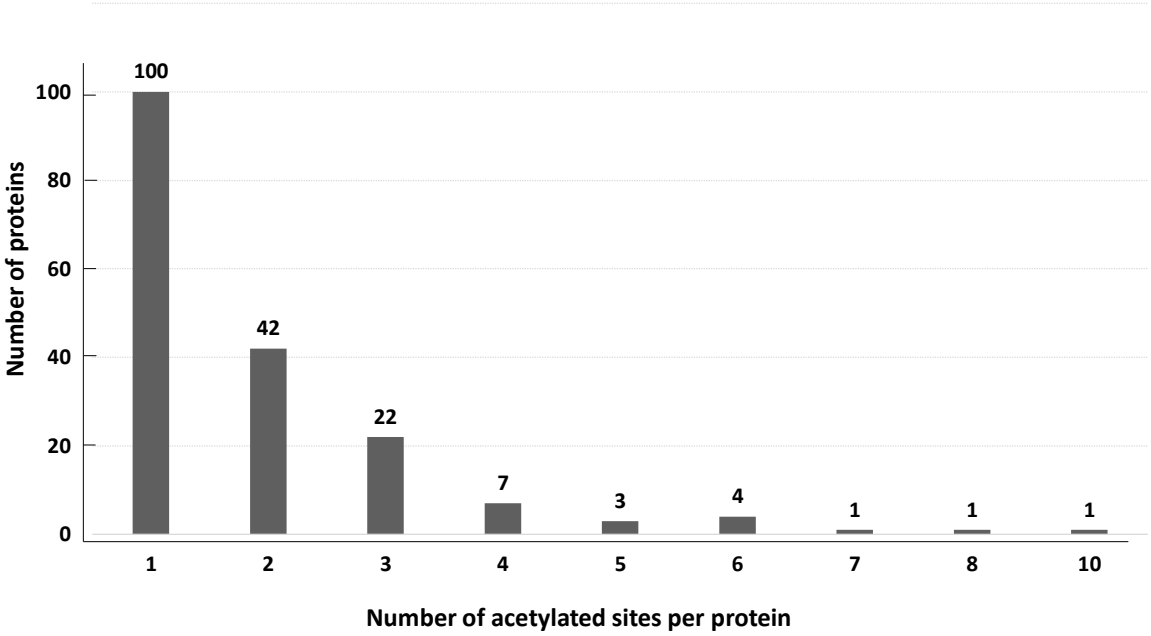


Figure 2. Distribution of T. gammatolerans proteins along the number of acetylated sites. The proteins with 7, 8 and 10 acetylated sites are the PEP synthase, EF-2, and EF-1-alpha, respectively.

Functional classification of acetylated proteins from *T. gammatolerans*.

The acetylated proteins were classified into different groups according to their biological process and molecular function established based on annotations, COG categories and KEGG pathways (Supplementary Table S4). As shown in **Figure 3**, the "metabolic processes" (45%) and the "information storage and processing mechanisms" (29%) are over-represented categories in this dataset.

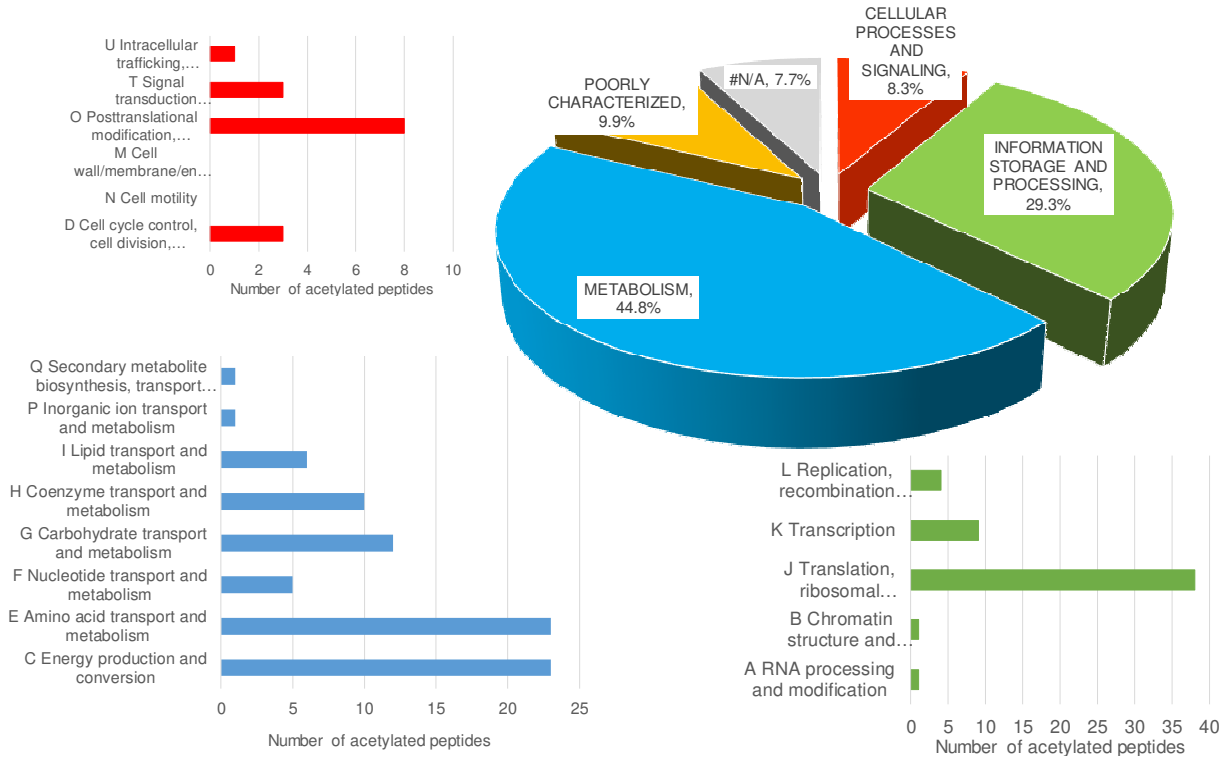


Figure 3. Functional classification of T. gammatolerans acetylated proteins based on COG classification and KEGG pathways. Detailed categories are presented for the three main groups: "metabolism" (blue), "information, storage and processing" (green), and "cellular processes and signaling" (red). N/A means not assigned.

Central metabolism - T. gammatolerans is able to grow on complex organic compounds or a mixture of 20 amino acids in presence of elemental sulfur (S°) but also without S° in the presence of sugars as carbon sources [21]. Here, cells were grown in a rich medium containing complex organic compounds and S°. Under these experimental conditions, *T. gammatolerans* catabolism is mainly based on peptides and amino-acid degradation and produces H_2S [21]. As shown in **Figure 4**, numerous peptidases are found to be acetylated, including two aminopeptidases (TGAM_0800, TGAM_0861), two 20S proteasome beta-subunits (TGAM_1721, TGAM_2052), the pyrolysin (TGAM_1044) and putative amino-, carboxy- or dipeptidases (TGAM_0929, TGAM_0186, and TGAM_0106). The proteinous substrates are assimilated in the subsequent steps, *i.e.* deamination of the resulting amino-acids into 2 oxo-acids and transformation into the corresponding acids by acetyl-CoA synthetases and succinyl-CoA synthetases (POR, IOR, VOR, ACS, SCS), respectively that produces energy through

concomitant ADP phosphorylation. These enzymes also provide the substrates required for other central pathways [33], such as the archaeal modified Embdem-Meyerhof pathway [34], the non-oxidative pentose phosphate pathway [35], as well as for the metabolism of several aminoacids. Interestingly, among the acetylated proteins involved in the modified Embden-Meyerhof pathway, no enzyme involved in glycolysis (GAPOR, PYK) and only several bidirectional enzymes or the GAPDH enzyme involved in neoglucogenesis pathway are found, suggesting that acetylation could also play a role in the regulation of the flux between glycolysis and neoglucogenesis.

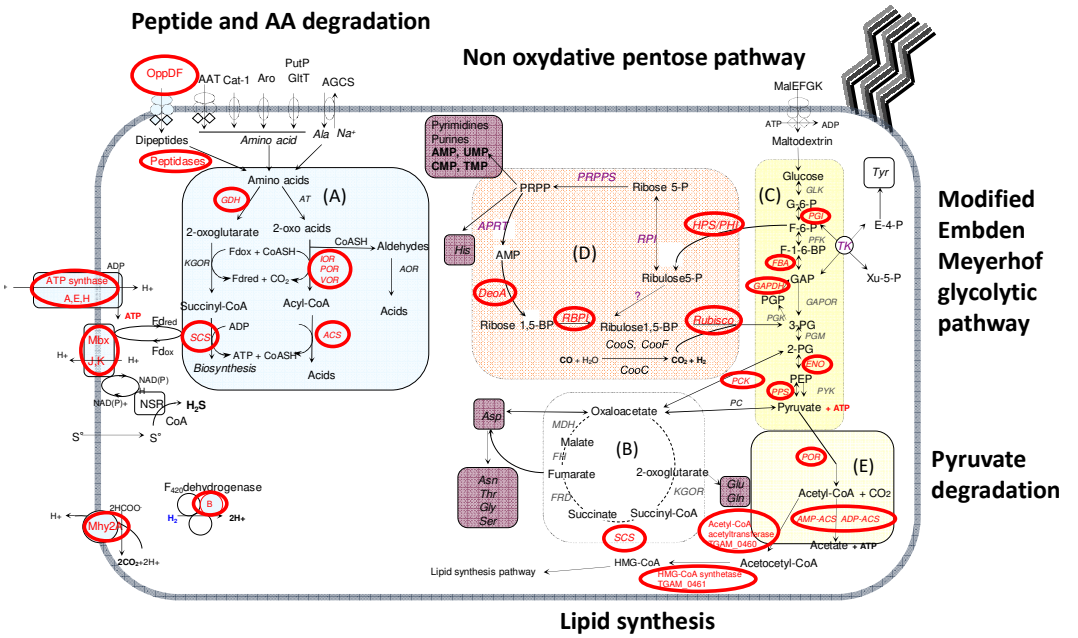


Figure 4. Role of acetylated enzymes in T. gammatolerans central metabolic pathways. Acetylated proteins are indicated in red and circled in red. Figure adapted from 42 showing the main metabolic pathways: Peptide and amino acid degradation, Pseudo TCA cycle, Modified Embden Meyerhof glycolytic pathway, Non oxidative pentose pathway, Pyruvate degradation and Lipid synthesis.

An interesting feature of *T. gammatolerans* among *Thermococcales* is the presence of an abundant acetate CoA ligase (TGAM_0230) that may produce ATP and acetate from acetyl-CoA and CO_2 or reversely could transform acetate into acetyl-CoA accompanied by AMP formation [21]. Alternatively, acetyl-CoA molecules may also be used for synthesis of lipids. In archaea, the two first steps are catalyzed by the acetyl-CoA synthetase which transforms two acetyl-CoA molecules into an acetoacetyl-CoA that is subsequently transformed into an HMG-

342 CoA by the HMG-CoA synthase. These two enzymes are also found acetylated in the present
343 dataset.
344 A large variety of hydrogenase complexes exist in *T. gammatolerans*. The F_{420} deshydrogenase
345 is supposed to adjust proton concentration into the cells. Like in all other sequenced
346 *Thermococcales* species, *T. gammatolerans* encodes orthologues of the membrane-bound Mbh
347 and Mbx complexes [36] that pump protons across the membrane, the resulting proton gradient
348 being used by the ATP synthase to form ATP. Based on our previous semi-quantitative
349 estimation of the amount of proteins in the cells harvested during the exponential phase [21],
350 Mbx components seemed more abundant than Mbh subunits indicating that Mbx is probably
351 preferentially used in the tested culture condition. One subunit of F_{420} deshydrogenase, the
352 MbxJ (TGAM_0712) and MbxK (TGAM_0713) subunits, as well as atpH (TGAM_0140), aptE
353 (TGAM_0143) and atpA (TGAM_0146), and three subunits of the ATP synthase are found to
354 be acetylated.
355 *Transcription and translation* - Amongst the proteins involved in transcription, the RNA
356 polymerase subunit (rpoB), the initiation factor TBP, the Tbp interacting protein
357 (TGAM_0640), six putative transcriptional regulators, and several other proteins were found
358 acetylated (Supplementary Table S4). The homolog of rpoB subunit in *E. coli* has been already
359 described as being acetylated [37], showing the conservation of this mechanism amongst
360 prokaryotes. Regarding translation, among the 64 ribosomal proteins, 9 belonging to the large
361 ribosome and one to the small ribosome subunits were found acetylated, as well as several
362 initiation and elongation factors, and 14 out of the 20 aminoacyl tRNA synthetases. The
363 Elongation factor 1-alpha (TGAM_1054) and the translation elongation factor EF-2
364 (TGAM_1397) are the two proteins which exhibited the highest number of acetylation sites: 10
365 and 8, respectively. An important number of proteins involved in translation were found

366 acetylated both in bacteria and in the archaeon *H. mediterranei* [18]. It has been proposed that
367 polyacetylation could participate in regulation of ribosome function and protein biosynthesis.

368 **Conservation of acetylated sites between two distantly related *Euryarchaeota*.**
369 The specific patterns of amino acid residues surrounding the lysine acetylation sites observed
370 with the two different antibodies exhibit some differences (Supplemental Data 1). More
371 importantly, we compared the position of lysine acetylation in orthologous proteins from the
372 two archaeal species for which we have now comprehensive experimental data
373 (**Supplementary Figure S1, Supplementary Table S6**). From this set of proteins, 99
374 acetylated proteins from *H. mediterranei* have an ortholog in *T. gammatolerans*, whereas 46 *T.*
375 *gammatolerans* acetylated proteins have an ortholog in *H. mediterranei* and only 26
376 orthologous proteins were found to be acetylated in both species (**Supplementary Table S7**).
377 These proteins are mainly involved in translation and in different metabolic pathways. They
378 may comprise several acetylation sites that could be conserved or not in their respective
379 orthologs. We checked whether a lysine acetylated site in one protein at a given position was
380 also found acetylated in the orthologous protein at the same position. From a total of 199
381 common lysine positions in the 26 pairs of orthologous proteins, 67 in *T. gammatolerans* and
382 58 in *H. mediterranei* are acetylated, but only three of them are acetylated at the same position
383 in both orthologs (**Supplementary Table S8, Supplementary Figure S1**). Interestingly, while
384 the translation elongation factor EF-1-Alpha (TGAM_1054) contains 10 acetylation sites and
385 the *H. mediterranei* ortholog (HFX_0346) encompasses six sites, only one site is found to be
386 acetylated in both species amongst the 13 putative sites conserved in both proteins. Since this
387 protein sequence is very well conserved in these archaea (E= $3.0\ E^{-137}$ and $8.0\ E^{-144}$, with a max
388 bitscore > 0.6), these results suggest that in archaea N^{ε}-Lysine acetylation is either conserved

for a given protein but not at the same sites, or a highly dynamic mechanism occurs which is not comprehensively detected by the current shotgun procedure.

Nucleoid structure and dynamics.

It was previously shown that mutation on the acetylated site of one of the 13 Cdc6A encoded by *H. mediterranei* destroyed the ARS activity of its adjacent origin oriC1 [18]. *T. gammatolerans* only encodes one Cdc6 (TGAM_0128) that was not found to be acetylated in our studies. In contrast, we report herein, among the acetylated proteins, several proteins described as involved in cell division and replication: the FtsZ-like protein (TGAM_1841), a cell division GTPase, a parB-like nuclease (TGAM_1907), the RPA41 subunit of replication factor A complex (TGAM_1760), a PP-loop ATPase, mrp-like protein (TGAM_0229), a recJ-like single-stranded exonuclease (TGAM_1460), the Gins 23 protein (TGAM_2023), as well as a primase-related protein (TGAM_1776). The reverse gyrase, a topoisomerase I that introduces positive supercoiling into DNA, is only encoded by hyperthermophilic organisms [38]. This protein was also found to be acetylated on two adjacent sites (Lys79 and Lys80). These lysines are conserved in the protein encoded by *Thermotoga maritima* which has been crystalized (PDB ID: 4ddu). They are located in the RecA-like H1 domain of the protein [39]. *T. gammatolerans* is able to resist to massive doses of ionizing radiations, but only a few DNA repair proteins, a Rad55-like protein (TGAM_0280) and a putative AP endonuclease (TGAM_1637), were found acetylated. However, we cannot exclude that acetylation on DNA repair proteins occurs when cells are exposed to DNA damaging agents. Altogether, these results showed that acetylation may be an important partner for essential mechanisms of DNA integrity maintenance in *T. gammatolerans*.

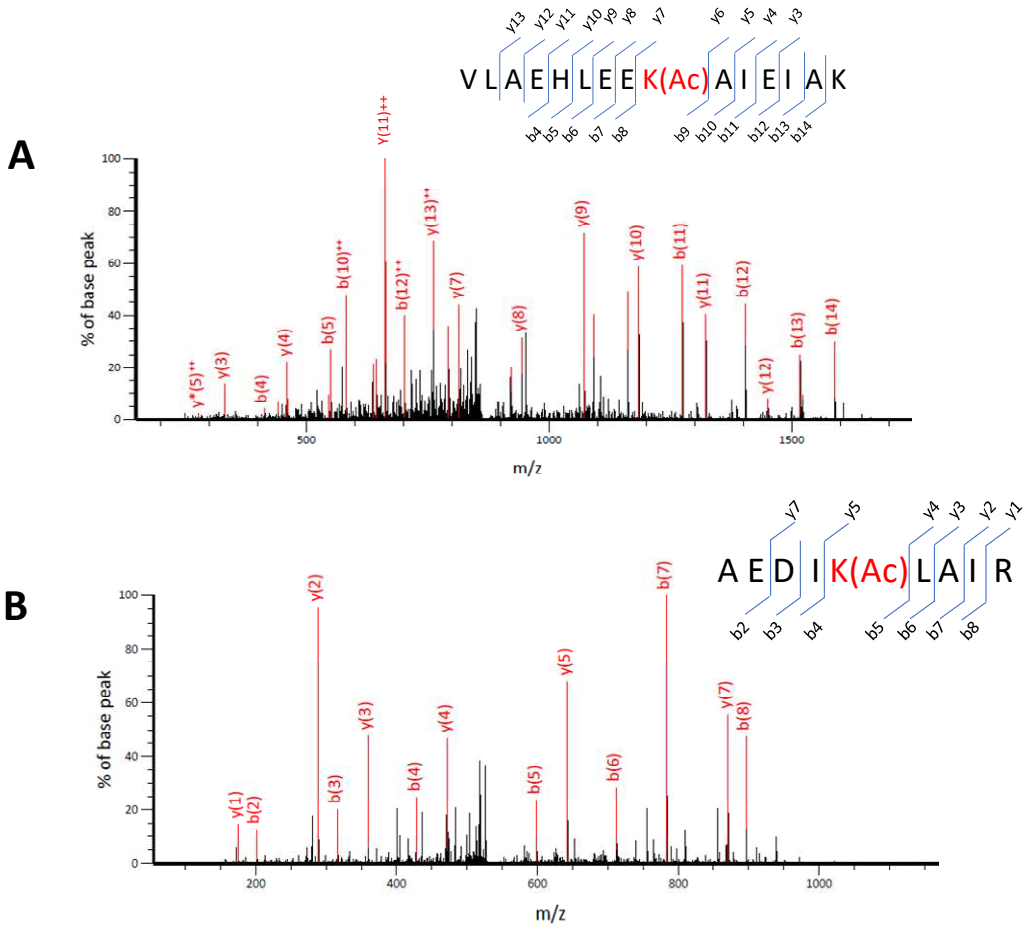


Figure 5. Examples of acetylation site identification in histones by tandem mass spectrometry. A. MS/MS spectrum of the acetylated [VLAEHLEEKacAIEIAK] tryptic peptide from T. gammatolerans Histone B (TGAM_0477). B. MS/MS spectrum of the acetylated [AEDIKacLAIR] peptide which is common in T. gammatolerans histone A or B protein sequences. In both spectra N-terminal fragment ions (b-type ions) and C-terminal fragment ions (y-type ions) are indicated.

Interestingly, we found that the histone A2 (also named histone B, TGAM_0477), composed of 67 amino acids, is acetylated in *T. gammatolerans* at least at two different sites (**Supplementary Table S3**): K_{27} and K_{36}, (numbering assuming that the N-terminal positions methionine is not processed). Another peptide common to histones A and B and located in the C-terminus was found acetylated at K_{62}. **Figure 5** shows the MS/MS spectra of two histone acetylated peptides. In panel A, the MS/MS spectrum (*m/z* at 867.98627, doubly charged ion, Δmass 1 ppm) corresponds to the [28–42] tryptic peptide of Histone B that is acetylated. In this case, the peptide fragments obtained after collisional-induced dissociation clearly indicate acetylation on Lys36. The y_7^+ fragment (*m/z* at 814.5033) and b_9^{2+} fragment (*m/z* at 546.2902) as well as y_8^+ fragment (*m/z* at 943.5459) and b_{10}^{2+} fragment (*m/z* at 581.8088) are the most

422 informative ions. In panel B, the MS/MS spectrum (*m/z* at 535.8137, doubly charged ion, Δmass
423 -0.3 ppm) corresponds to the [258–66] tryptic acetylated peptide common to histones A and B.
424 In this case, the peptide fragments obtained after collisional-induced dissociation clearly
425 indicate acetylation on Lys62. The y_5^+ fragment (*m/z* at 642.4297) and b_5^+ fragment (*m/z* at
426 546.2902) are the most informative ions. Archaeal histones are homologous to their H3-H4
427 eukaryotic counterparts and are well conserved among *Euryarchaeota*, but are not reputed to
428 be post-translationally modified since archaeal histones lack the N-terminal domain of
429 eukaryotic histones, on which most of the PTMs occur in eukaryotes [40]. Alba, which plays
430 an important role for chromatin structure of *Sulfolobales* that do not encode histones, is also
431 post translationally modified but is methylated on K_{16} [36].
432 The chromatin structure of *T. kodakaraensis*, a phylogenetically closely-related species has
433 been investigated, highlighting a classical beads-of-the-string structure mainly formed by DNA
434 and histones and also a thick fibrous structure composed of DNA, histones, Alba and TrmBL2
435 proteins [25]. In order to check whether acetylation also occurs in histones from *T.*
436 *kodakaraensis*, we extracted its chromatin from mid-log phase cells as well as from late
437 stationary cells, and directly analyzed the presence of acetylated peptides from the proteins with
438 a molecular weight of 15 kDa. As shown in **Figure 6** the histone A and B sequences from *T.*
439 *kodakaraensis* and *T. gammatolerans* are well conserved. Six lysine positions are perfectly
440 conserved among the four sequences while three other lysine positions are not conserved. The
441 analysis of acetylated peptides in *T. kodakaraensis* species shows that both histones A and B
442 can be acetylated on several acetylation sites. We identified up to six acetylated sites in the
443 histone B of *T. kodakaraensis* (**Figure 6, Supplementary Table S9**). Histone A from *T.*
444 *kodakaraensis* displays three acetylated sites, two of them being common to histone B. The
445 same peptides were also found unacetylated in the tryptic digestion of chromatin-enriched
446 15kDa proteins, supporting the idea that all the histones extracted from the chromatin are not

447 acetylated at each site. The number of acetylated peptides increased when chromatin was
448 extracted from late stationary phase cells (**Supplementary Table S9**). Altogether these results
449 suggest this post-translational modification on histones is a dynamic mechanism in
450 *Thermococcales* species.

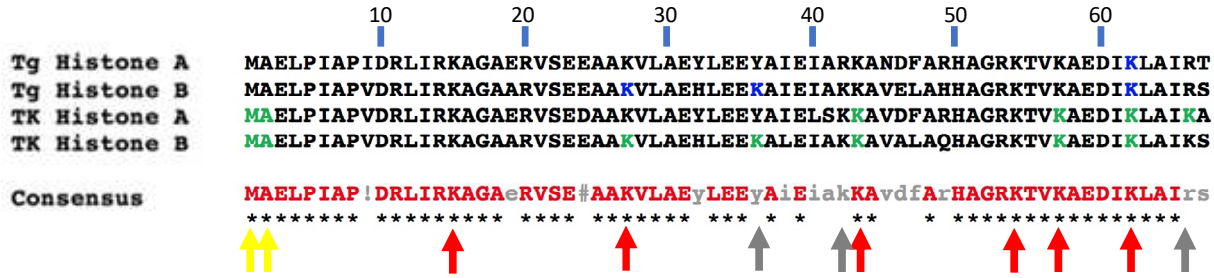


Figure 6. Position of the acetylation sites found in the 4 histone proteins. On the multi-alignment of the histones A and B sequences from T. gammatolerans (Tg) and T. kodakaraensis (Tk) the conserved and not conserved lysines (K) are indicated with red and grey arrows, respectively. The N-terminal acetylation position is illustrated by a yellow arrow. Acetylated lysine from peptides trapped by antibodies by colored in blue. Acetylated lysines detected from chromatin enriched histones are colored in green.

451 Recently, the structure of a 90 pb DNA fragment bound to three homodimers of histone B
452 (HMfB) from the archaeon *Methanothermus fervidus* has been published [41]. Interaction
453 modes with DNA are very similar between the archaeal and eukaryotic structures as well the
454 dimer:dimer arrangements while the surface of the archaeal histone complex/dimer displays a
455 less positive charge [41]. Nevertheless, the HMfB dimers are symmetric and polymerize to
456 form with the wrapped DNA an endless left-handed rod which Henneman *et al* proposed to call
457 "hypernucleosome" [42].
458 The 3D structures of the histones from *T. gammatolerans* or *T. kodakaraensis* have not been
459 characterized experimentally. *T. kodakaraensis* histone homodimers TkHA/TkHA,
460 TkHB/TkHB and heterodimer TkHA/TkHB structural models were built using as template the
461 X-ray structure of *P. horikoshii,* (PDB code; 1ku5), a *Thermococcale* species (**Supplementary**
462 **Figure S2**). The 3D structures of the dimer models (TkHA/TkHA, TkHB/TkHB and
463 TkHA/TkHB from *T. kodakaraensis*) are very similar to the histone B dimer MfHB from
464 *Methanothermus fervidus* (all RMSD values ~ 0.73Å). Using the X-ray structure of the polymer

of *M. fervidus* (PDB code: 5t5k), we built models of hexamers from three pairs of histones TkHA/TkHA, TkHB/TkHB and TkHA/TkHB. All models display a very good overlay with the X-ray structure of *M. fervidus* as shown in **Supplementary Figure S3** which represents the structure alignment of three consecutive TkHB dimers with three dimers of MfHB. One should notice that the resolution of the X-ray structure of the polymer of HMfB is 4Å, thus, the position of side-chains in the X-ray structure remains imprecise. However, it was proposed that interactions of four residue pairs (E30-K61, E34-K65, K14-Q48, K26-E58, positions based on the sequences starting with the second amino acid residue), are involved in hypernucleosome stability and compactness [42]. Interestingly three out of the four lysines (K26, K61, K65) are found to be acetylated (**Supplementary Table S9**). In this model, acetylation could affect lysine interaction with acidic residues (E30, E54 and E58 residues, respectively). It could destabilize dimer:dimer stacking and thus the compactness of two adjacent histone layers (**Figure 7**). All these acetylation sites are conserved between HtkA and HTkB, as well as the three glutamates (E31, E35 and E59 residues), suggesting that the impact of these lysine acetylations is the same in the three dimer forms (e.g. HTkA/HTkA, HTkB/HTkB and HTkA/HTkB). Strikingly, K42 is in proximity to the carboxyl-terminus of both the other monomer and of the adjacent dimer (N+1) (**Supplementary Figure S4**). We can hypothesize that its acetylation could also affect the stability of the dimer and/or the polymerization process. In addition, we also propose that the K57 residue interact with DNA (**Supplementary Figure S5**) and suggest that acetylation might reduce the affinity for DNA. It was already reported that lysine acetylation of HU, a histone-like protein, alters its DNA binding properties and may modulate the chromosome organization of *Mycobacterium tuberculosis* [43]. The impact of K35 acetylation is unclear as this lysine is not conserved and is replaced by a tyrosine in histone B (**Figure 6**). Finally, the N-terminus of HTkB has also been found acetylated on an alanine residue. We already reported that several cytosolic and membrane proteins from *T.*

490 *gammatolerans* were acetylated at their N-terminal residue after excision of the N-terminal
491 methionine [21]. From the N-terminal peptidic signatures that were recorded, we deduced that
492 *T. gammatolerans* encodes a functional analogue of NatA because several acetylation occurred
493 on an alanine located at position 2 [21]. It has been shown that, H2A and H4 histones exhibited
494 acetylated serine residues located at the N-terminal of these proteins in *S. cerevisiae* and
495 maintenance of N-terminal acetylation of histone H4 reduced the extension of lifespan mediated
496 by calorie restriction [44, 45].

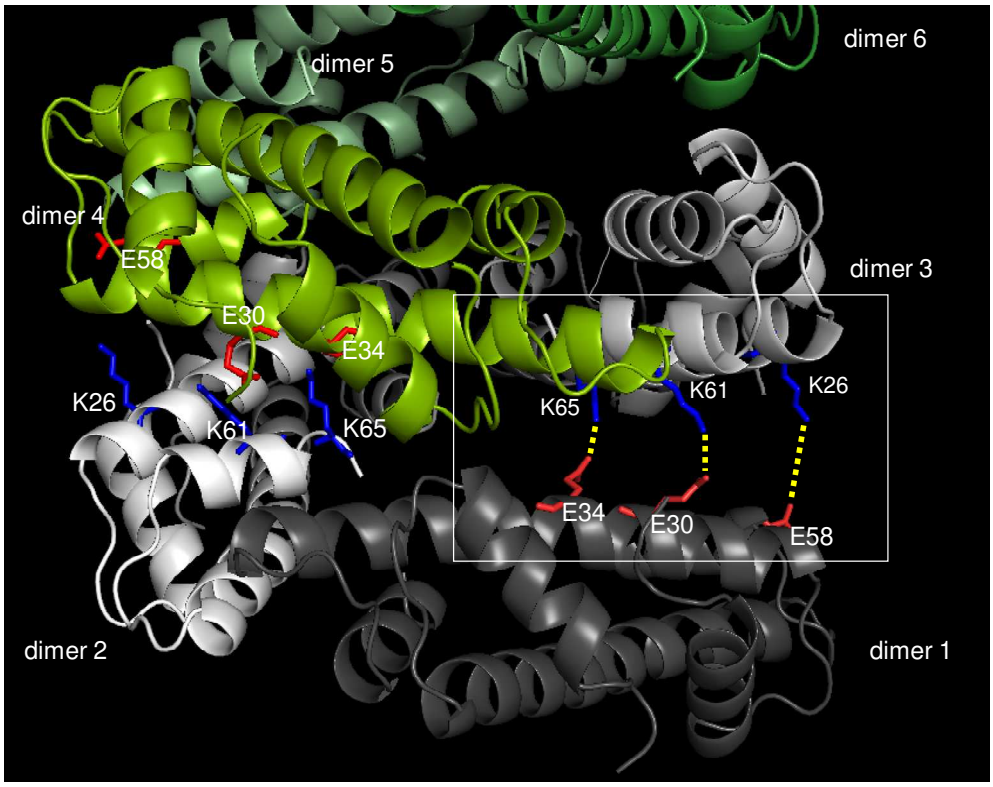


Figure 7. Interactions between TkHB dimers in the hypernucleosome model. The i+2 dimer interactions involving K26, K61 and K65 are highlighted (white box). Each dimer is colored differently and represented in cartoon. Lysines 26, 61 and 65 in dimers 2 and 3 are represented in blue sticks while glutamates 30, 34 and 58 in dimers 1 and 4 are represented in red sticks.

497 Altogether, the acetylome data along with the structural models of histone polymers of *T.*
498 *kodakarensis* strongly suggest that lysine acetylation in histone A and B may play an important
499 role in histone assembly and/or interaction with DNA. The regulation of gene expression in
500 eukaryotic species is mainly driven by post-translational modifications of histone tails that are
501 lacking in most of the archaeal ones. We can hypothesize that the acetylation of lysines involved

in dimer:dimer or histone:DNA interactions could ensure the regulation of gene expression by disrupting interaction with DNA or hypernucleosome assembly, thus destabilizing the chromatin structure. Mattiroli *et al* [41] showed that perturbation of hypernucleosome function *in vivo* driven by a mutation on G17 in *T. kodakarensis* modifies the response to nutrient change, suggesting a role in transcription. These results could also explain why histone acetylated peptides were mostly found in an enriched-chromatin protein extracts from stationary phase cells. Finally, we also found that not only histones but also TrmBL2 and Alba from proteins are acetylated in these *Thermococcus* species. TrmBL2 (TGAM_1678) from *T. gammatolerans* was found to be acetylated on K192 and K198, and Alba from *T. kodakaraensis* is acetylated on the N-terminal of the protein. This acetylation has already also been recently reported for Alba from *Sulfolobus islandicus* [15]. The lack of structure of TrmBL2 prevents the ability to model the impact of acetylation on TrmBL2 activity but all these results suggested that acetylation of other chromatin proteins may also modulate the dynamics of the genome structure.

Conclusions

The aim of the present study was to verify whether lysine-specific acetylation of proteins occurs at a genome-wide scale in the archaeon *Thermococcus gammatolerans.* We established an in-depth acetylome panorama for *T. gammatolerans*. A total of 316 Nε-Lysine peptides were confidently identified, with 338 distinct acetylation sites listed from 181 unique protein sequences. We compared the acetylome patterns of *T. gammatolerans* and *H. mediterranei* archaea which belong to the same *Euryarchaeota* lineage and evidenced strong differences that call for further experimental comparative studies. We evidenced for the first time that histones from two different archaea, namely *T. gammatolerans* and *T. kodakaraensis*, can be acetylated

at various positions. This discovery reinforces the strong evolutionary link between Archaea and Eukaryotes and pleads for further investigation on the extent and role of acetylation of histones in Archaea.

Credit author statement

Béatrice Alpha-Bazin: Conceptualization, Investigation, Validation, Data Curation, Writing **Aurore Gorlas**: Investigation, Resources **Arnaud Lagorce**: Investigation, Resources **Damien Joulié**: Investigation, **Jean-Baptiste Boyer**: Investigation **Murielle Dutertre**: Investigation, Resources **Jean-Charles Gaillard**: Investigation **Anne Lopes**: Methodology, Formal analysis **Yvan Zivanovic**: Methodology, Formal analysis; **Alain Dedieu**: Conceptualization, Validation, Supervision **Fabrice Confalonieri**: Conceptualization, Investigation, Validation, Formal analysis, Supervision, Writing, Project administration, Funding acquisition **Jean Armengaud**: Conceptualization, Validation, Formal analysis, Supervision, Writing, Project administration, Funding acquisition

Acknowledgements

We thank the Agence Nationale de la Recherche ("GAMMATOLERANS" project, ANR-12-BSV6-0012) for financial support.

REFERENCES

[1] V.G. Allfrey, R. Faulkner, A.E. Mirsky, Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis, Proc Natl Acad Sci U S A 51 (1964) 786-94.
[2] D.M. Phillips, The presence of acetyl groups of histones, Bochem J 87 (1963) 258-63.
[3] A. Drazic, L.M. Myklebust, R. Ree, T. Arnesen, The world of protein acetylation, Biochim Biophys Acta 1864(10) (2016) 1372-401.
[4] J. Soppa, Protein acetylation in archaea, bacteria, and eukaryotes, Archaea 2010 (2010).

[5] V.J. Carabetta, I.M. Cristea, Regulation, Function, and Detection of Protein Acetylation in Bacteria, J Bacteriol 199(16) (2017).
[6] S. Bonissone, N. Gupta, M. Romine, R.A. Bradshaw, P.A. Pevzner, N-terminal protein processing: a comparative proteogenomic analysis, Mol Cell Proteomics 12(1) (2013) 14-28.
[7] E.A. Vorontsov, E. Rensen, D. Prangishvili, M. Krupovic, J. Chamot-Rooke, Abundant Lysine Methylation and N-Terminal Acetylation in Sulfolobus islandicus Revealed by Bottom-Up and Top-Down Proteomics, Mol Cell Proteomics 15(11) (2016) 3388-3404.
[8] D.G. Christensen, J.T. Baumgartner, X. Xie, K.M. Jew, N. Basisty, B. Schilling, M.L. Kuhn, A.J. Wolfe, Mechanisms, Detection, and Relevance of Protein Acetylation in Prokaryotes, mBio 10(2) (2019).
[9] B. Macek, K. Forchhammer, J. Hardouin, E. Weber-Ban, C. Grangeasse, I. Mijakovic, Protein post-translational modifications in bacteria, Nat Rev Microbiol 17(11) (2019) 651-664.
[10] V. Bernal, S. Castano-Cerezo, J. Gallego-Jara, A. Ecija-Conesa, T. de Diego, J.L. Iborra, M. Canovas, Regulation of bacterial physiology by lysine acetylation of proteins, N Biotechnol 31(6) (2014) 586-95.
[11] J. Ren, Y. Sang, J. Lu, Y.F. Yao, Protein Acetylation and Its Role in Bacterial Virulence, Trends Microbiol 25(9) (2017) 768-779.
[12] C.M. VanDrisse, J.C. Escalante-Semerena, Protein Acetylation in Bacteria, Annu Rev Microbiol 73 (2019) 111-132.
[13] T. Hase, S. Wakabayashi, H. Matsubara, L. Kerscher, D. Oesterhelt, K.K. Rao, D.O. Hall, Complete amino acid sequence of Halobacterium halobium ferredoxin containing an Nepsilon-acetyllysine residue, J Biochem 83(6) (1978) 1657-70.
[14] S.D. Bell, C.H. Botting, B.N. Wardleworth, S.P. Jackson, M.F. White, The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation, Science 296(5565) (2002) 148-51.
[15] J. Cao, Q. Wang, T. Liu, N. Peng, L. Huang, Insights into the post-translational modifications of archaeal Sis10b (Alba): lysine-16 is methylated, not acetylated, and this does not regulate transcription or growth, Mol Microbiol (2018).
[16] M. Aivaliotis, K. Gevaert, M. Falb, A. Tebbe, K. Konstantinidis, B. Bisle, C. Klein, L. Martens, A. Staes, E. Timmerman, J. Van Damme, F. Siedler, F. Pfeiffer, J. Vandekerckhove, D. Oesterhelt, Large-scale identification of N-terminal peptides in the halophilic archaea Halobacterium salinarum and Natronomonas pharaonis, J Proteome Res 6(6) (2007) 2195-204.
[17] P.A. Kirkland, M.A. Humbard, C.J. Daniels, J.A. Maupin-Furlow, Shotgun proteomics of the haloarchaeon Haloferax volcanii, J Proteome Res 7(11) (2008) 5033-9.
[18] J. Liu, Q. Wang, X. Jiang, H. Yang, D. Zhao, J. Han, Y. Luo, H. Xiang, Systematic Analysis of Lysine Acetylation in the Halophilic Archaeon Haloferax mediterranei, J Proteome Res 16(9) (2017) 3229-3241.
[19] E. Jolivet, S. L'Haridon, E. Corre, P. Forterre, D. Prieur, Thermococcus gammatolerans sp. nov., a hyperthermophilic archaeon from a deep-sea hydrothermal vent that resists ionizing radiation, Int J Syst Evol Microbiol 53(Pt 3) (2003) 847-51.
[20] A. Tapias, C. Leplat, F. Confalonieri, Recovery of ionizing-radiation damage after high doses of gamma ray in the hyperthermophilic archaeon Thermococcus gammatolerans, Extremophiles 13(2) (2009) 333-43.
[21] Y. Zivanovic, J. Armengaud, A. Lagorce, C. Leplat, P. Guerin, M. Dutertre, V. Anthouard, P. Forterre, P. Wincker, F. Confalonieri, Genome analysis and genome-wide proteomics of Thermococcus gammatolerans, the most radioresistant organism known amongst the Archaea, Genome Biol 10(6) (2009) R70.
[22] A. de Groot, R. Dulermo, P. Ortet, L. Blanchard, P. Guerin, B. Fernandez, B. Vacherie, C. Dossat, E. Jolivet, P. Siguier, M. Chandler, M. Barakat, A. Dedieu, V. Barbe, T. Heulin, S.

601 Sommer, W. Achouak, J. Armengaud, Alliance of proteomics and genomics to unravel the
602 specificities of Sahara bacterium Deinococcus deserti, PLoS Genet 5(3) (2009) e1000434.
603 [23] G. Clair, J. Armengaud, C. Duport, Restricting fermentative potential by proteome
604 remodeling: an adaptive strategy evidenced in Bacillus cereus, Mol Cell Proteomics 11(6)
605 (2012) M111 013102.
606 [24] G. Klein, C. Mathe, M. Biola-Clier, S. Devineau, E. Drouineau, E. Hatem, L. Marichal, B.
607 Alonso, J.C. Gaillard, G. Lagniel, J. Armengaud, M. Carriere, S. Chedin, Y. Boulard, S. Pin,
608 J.P. Renault, J.C. Aude, J. Labarre, RNA-binding proteins are a major target of silica
609 nanoparticles in cell extracts, Nanotoxicology 10(10) (2016) 1555-1564.
610 [25] H. Maruyama, M. Shin, T. Oda, R. Matsumi, R.L. Ohniwa, T. Itoh, K. Shirahige, T.
611 Imanaka, H. Atomi, S.H. Yoshimura, K. Takeyasu, Histone and TK0471/TrmBL2 form a novel
612 heterogeneous genome architecture in the hyperthermophilic archaeon Thermococcus
613 kodakarensis, Mol Biol Cell 22(3) (2011) 386-98.
614 [26] E.M. Hartmann, F. Allain, J.C. Gaillard, O. Pible, J. Armengaud, Taking the shortcut for
615 high-throughput shotgun proteomic analysis of bacteria, Methods Mol Biol 1197 (2014) 275-
616 85.
617 [27] F. Corpet, Multiple sequence alignment with hierarchical clustering, Nucleic Acids Res
618 16(22) (1988) 10881-90.
619 [28] M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T.
620 Gallo Cassarino, M. Bertoni, L. Bordoli, T. Schwede, SWISS-MODEL: modelling protein
621 tertiary and quaternary structure using evolutionary information, Nucleic Acids Res 42(Web
622 Server issue) (2014) W252-8.
623 [29] E.W. Deutsch, N. Bandeira, V. Sharma, Y. Perez-Riverol, J.J. Carver, D.J. Kundu, D.
624 Garcia-Seisdedos, A.F. Jarnuczak, S. Hewapathirana, B.S. Pullman, J. Wertz, Z. Sun, S.
625 Kawano, S. Okuda, Y. Watanabe, H. Hermjakob, B. MacLean, M.J. MacCoss, Y. Zhu, Y.
626 Ishihama, J.A. Vizcaino, The ProteomeXchange consortium in 2020: enabling 'big data'
627 approaches in proteomics, Nucleic Acids Res 48(D1) (2020) D1145-D1152.
628 [30] Y.S. Yang, B. Fernandez, A. Lagorce, V. Aloin, K.M. De Guillen, J.B. Boyer, A. Dedieu,
629 F. Confalonieri, J. Armengaud, C. Roumestand, Prioritizing targets for structural biology
630 through the lens of proteomics: the archaeal protein TGAM_1934 from Thermococcus
631 gammatolerans, Proteomics 15(1) (2015) 114-23.
632 [31] K.L. Guan, W. Yu, Y. Lin, Y. Xiong, S. Zhao, Generation of acetyllysine antibodies and
633 affinity enrichment of acetylated peptides, Nat Protoc 5(9) (2010) 1583-95.
634 [32] P.G. Shaw, R. Chaerkady, Z. Zhang, N.E. Davidson, A. Pandey, Monoclonal antibody
635 cocktail as an enrichment tool for acetylome analysis, Anal Chem 83(10) (2011) 3623-6.
636 [33] C. Brasen, D. Esser, B. Rauch, B. Siebers, Carbohydrate metabolism in Archaea: current
637 insights into unusual enzymes and pathways and their regulation, Microbiol Mol Biol Rev 78(1)
638 (2014) 89-175.
639 [34] T. Sato, H. Atomi, Novel metabolic pathways in Archaea, Curr Opin Microbiol 14(3)
640 (2011) 307-14.
641 [35] I. Orita, T. Sato, H. Yurimoto, N. Kato, H. Atomi, T. Imanaka, Y. Sakai, The ribulose
642 monophosphate pathway substitutes for the missing pentose phosphate pathway in the archaeon
643 Thermococcus kodakaraensis, J Bacteriol 188(13) (2006) 4698-704.
644 [36] G.J. Schut, E.S. Boyd, J.W. Peters, M.W. Adams, The modular respiratory complexes
645 involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and
646 their evolutionary implications, FEMS Microbiol Rev 37(2) (2013) 182-203.
647 [37] B.P. Lima, H. Antelmann, K. Gronau, B.K. Chi, D. Becher, S.R. Brinsmade, A.J. Wolfe,
648 Involvement of protein acetylation in glucose-induced transcription of a stress-responsive
649 promoter, Mol Microbiol 81(5) (2011) 1190-204.

650 [38] C. Brochier-Armanet, P. Forterre, Widespread distribution of archaeal reverse gyrase in
651 thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene
652 transfers, Archaea 2(2) (2007) 83-93.
653 [39] M.G. Rudolph, Y. del Toro Duany, S.P. Jungblut, A. Ganguly, D. Klostermeier, Crystal
654 structures of Thermotoga maritima reverse gyrase: inferences for the mechanism of positive
655 DNA supercoiling, Nucleic Acids Res 41(2) (2013) 1058-70.
656 [40] K. Luger, T.J. Richmond, The histone tails of the nucleosome, Curr Opin Genet Dev 8(2)
657 (1998) 140-6.
658 [41] F. Mattiroli, S. Bhattacharyya, P.N. Dyer, A.E. White, K. Sandman, B.W. Burkhart, K.R.
659 Byrne, T. Lee, N.G. Ahn, T.J. Santangelo, J.N. Reeve, K. Luger, Structure of histone-based
660 chromatin in Archaea, Science 357(6351) (2017) 609-612.
661 [42] B. Henneman, C. van Emmerik, H. van Ingen, R.T. Dame, Structure and function of
662 archaeal histones, PLoS Genet 14(9) (2018) e1007582.
663 [43] S. Ghosh, B. Padmanabhan, C. Anand, V. Nagaraja, Lysine acetylation of the
664 Mycobacterium tuberculosis HU protein modulates its DNA binding and genome organization,
665 Mol Microbiol 100(4) (2016) 577-88.
666 [44] D. Molina-Serrano, V. Schiza, C. Demosthenous, E. Stavrou, J. Oppelt, D. Kyriakou, W.
667 Liu, G. Zisser, H. Bergler, W. Dang, A. Kirmizis, Loss of Nat4 and its associated histone H4
668 N-terminal acetylation mediates calorie restriction-induced longevity, EMBO Rep 17(12)
669 (2016) 1829-1843.
670 [45] O.K. Song, X. Wang, J.H. Waterborg, R. Sternglanz, An Nalpha-acetyltransferase
671 responsible for acetylation of the N-terminal residues of histones H4 and H2A, J Biol Chem
672 278(40) (2003) 38109-12.
673
674

675 **FIGURE LEGENDS**

676

677 **Figure 1. Detection of lysine-acetylated proteins.** Western blotting analysis using either anti-
678 acetylated-lysine rabbit polyclonal antibody (PAB) with 20 µg (Lane 1), 10 µg (Lane 2) of
679 extracted proteins of *T. gammatolerans* and 15 µg (Lane 3) of extracted proteins of
680 *Saccharomyces cerevisiae,* or anti-acetylated-lysine rabbit monoclonal antibody (MAB) with
681 20 µg (Lane 4), 10 µg (Lane 5) of extracted proteins of *T. gammatolerans* and 15µg (Lane 6)
682 of extracted proteins of *S. cerevisiae*. The molecular weights of the protein markers are
683 indicated in kDa (Lane M).

684

Figure 2. Distribution of *T. gammatolerans* proteins along the number of acetylated sites. The proteins with 7, 8 and 10 acetylated sites are the PEP synthase, EF-2, and EF-1-alpha, respectively.

Figure 3. Functional classification of *T. gammatolerans* acetylated proteins based on COG classification and KEGG pathways. Detailed categories are presented for the three main groups: "metabolism" (blue), "information, storage and processing" (green), and "cellular processes and signaling" (red).

Figure 4. Role of acetylated enzymes in *T. gammatolerans* central metabolic pathways. Acetylated proteins are indicated in red and circled in red. Figure adapted from [42] showing the main metabolic pathways: Peptide and amino acid degradation, Pseudo TCA cycle, Modified Embden Meyerhof glycolytic pathway, Non oxidative pentose pathway, Pyruvate degradation and Lipid synthesis.

Figure 5. Examples of acetylation site identification in histones by tandem mass spectrometry. A. MS/MS spectrum of the acetylated [VLAEHLEEKacAIEIAK] tryptic peptide from *T. gammatolerans* Histone B (TGAM_0477). **B.** MS/MS spectrum of the acetylated [AEDIKacLAIR] peptide which is common in *T. gammatolerans* histone A or B protein sequences. In both spectra N-terminal fragment ions (b-type ions) and C-terminal fragment ions (y-type ions) are indicated.

Figure 6. Position of the acetylation sites found in the 4 histone proteins. On the multi-alignment of the histones A and B sequences from *T. gammatolerans* (Tg) and *T. kodakaraensis* (Tk) the conserved and not conserved lysines (K) are indicated with red and grey arrows,

710 respectively. The N-terminal acetylation position is illustrated by a yellow arrow. Acetylated
711 lysine from peptides trapped by antibodies by colored in blue. Acetylated lysines detected from
712 chromatin enriched histones are colored in green.
713
714 **Figure 7. Interactions between TkHB dimers in the hypernucleosome model.** The i+2
715 dimer interactions involving K26, K61 and K65 are highlighted (white box). Each dimer is
716 colored differently and represented in cartoon. Lysines 26, 61 and 65 in dimers 2 and 3 are
717 represented in blue sticks while glutamates 30, 34 and 58 in dimers 1 and 4 are represented in
718 red sticks.
719
720 **SUPPLEMENTARY DATA**
721
722 **Supplementary Data 1. Acetylated lysine consensus motives (Figure A and comments).**
723
724 **Supplementary Figure S1. Alignment of orthologous acetylated proteins in *H.***
725 ***mediterranei* and *T. gammatolerans*.** The data from *H. mediterranei* have been extracted
726 from [18]. The acetylated lysine is boxed in red.
727
728 **Supplementary Figure S2. HA/HA, HB/HB, HA/HB dimer models superimposed on the**
729 **X-ray structure of the dimer HMfB (5t5k).**
730
731 **Supplementary Figure S3. Model of the hexamer:DNA complex for *T. kodakarensis***
732 **according to *M. fervidus* complex structure.** Model of TkHB/TkHB hexamer structure
733 (yellow) threaded on MfHB hexamer (gray and black) and wrapped by a 90bp DNA fragment.
734

Supplementary Figure S4. Interactions between K42 and the carboxyl-terminus of the other chains in the TkHB hexamer model. A. Each histone dimer is colored differently and represented in cartoon. K42 and the two carboxyl-terminus (S66) belonging either to the other monomer within the same dimer (gray chain) or to the adjacent dimer (green or orange chain) are represented in sticks and colored in blue (K42) and red (carboxyl-terminus). **B.** Zoom-in on the interactions between K42 and the two carboxyl-terminus group in proximity. The S66 in grey belongs to the other monomer forming the histone dimer. The S66 in green belongs to the adjacent dimer.

Supplementary Figure S5. Interactions between TkHB dimers and DNA. A. Model of a hexamer of TkHB in interaction with DNA calculated from X-ray structure of *M. fervidus* nucleosome (PDB code: 5t5k). Each histones dimer is colored differently and represented in cartoon. Lysines 56 of each monomer are represented in blue sticks. **B.** Zoom-in on K56-DNA interactions.

Supplementary Table S1. List of MS/MS spectra assigned to *T. gammatolerans* peptides.

Supplementary Table S2. List of all detected proteins from *T. gammatolerans* and their spectral counts.

Supplementary Table S3. List of MS/MS spectra assigned to acetylated peptides from *T. gammatolerans*.

Supplementary Table S4. List of *T. gammatolerans* acetylated proteins, their spectral count and COG classification.

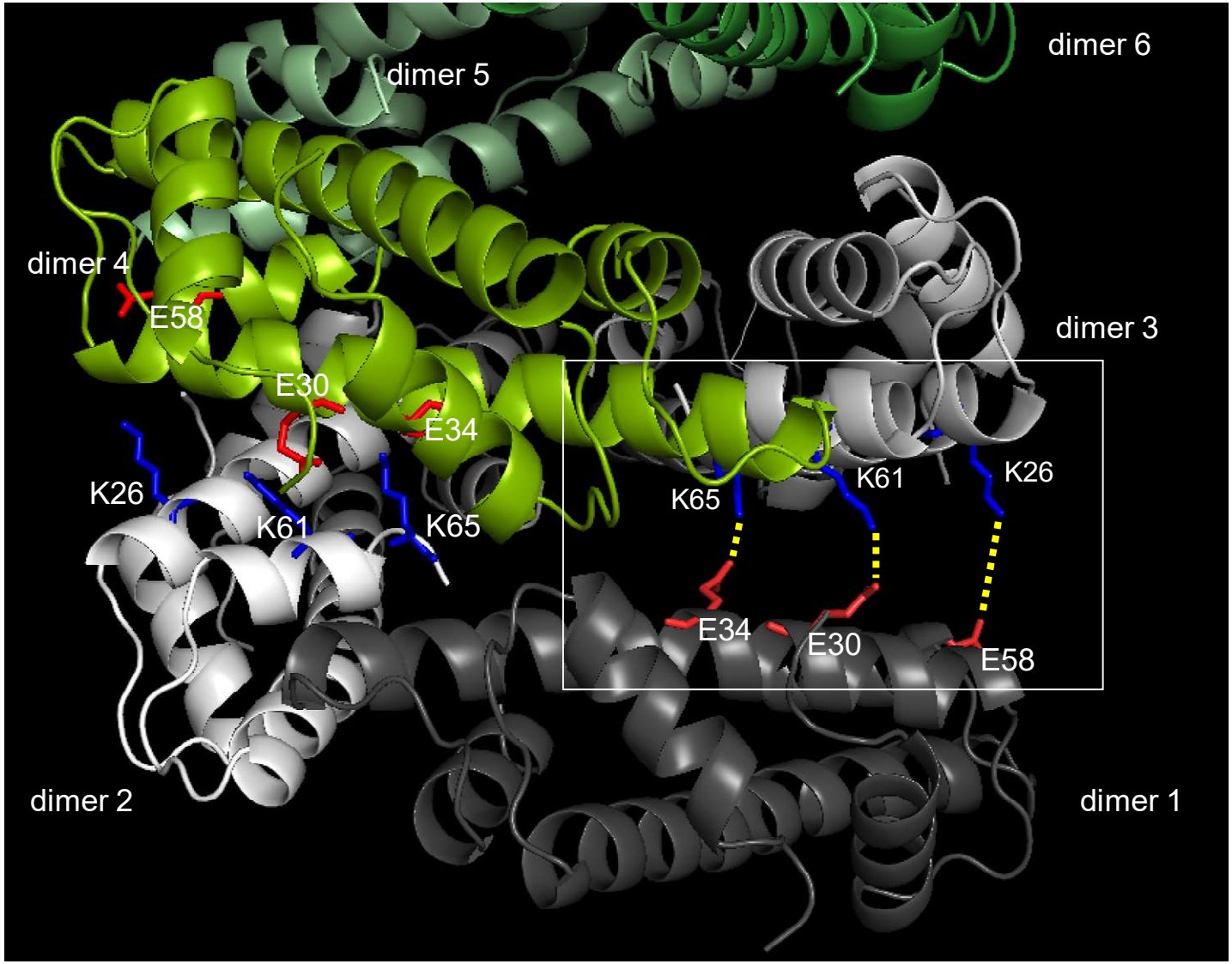
Supplementary Table S5. List of acetylated peptides, their characteristics and spectral counts per sample.

Supplementary Table S6. Reciprocal orthologs of *T. gammatolerans* and *H. mediterranei.*

Supplementary Table S7. Acetylated proteins and acetylated orthologs in *T. gammatolerans* and *H. mediterranei.*

Supplementary Table S8. Number of acetylated K-sites in orthologous proteins.

Supplementary Table S9. Peptides and spectral counts for histones in chromatin bound 15 kDa proteins from *T. kodakaraensis*.


dimer 6
dimer 5
dimer 4
E58
dimer 3
E30
E34
K26
K61
K65
K65
K61
K26
E34
E30
E58
dimer 2
dimer 1