nature structural & molecular biology

Check for updates

Evolution of a histone variant involved in compartmental regulation of NAD metabolism

Iva Guberovic [1], Sarah Hurtado-Bagès[1], Ciro Rivera-Casas[2], Gunnar Knobloch[3,4], Roberto Malinverni[1], Vanesa Valero[1], Michelle M. Leger [5], Jesús García [6], Jerome Basquin[7], Marta Gómez de Cedrón [8], Marta Frigolé-Vivas[6], Manjinder S. Cheema[9], Ainhoa Pérez[1], Juan Ausió[9], Ana Ramírez de Molina[8], Xavier Salvatella [6,10,11], Iñaki Ruiz-Trillo[5,11], Jose M. Eirin-Lopez[2], Andreas G. Ladurner[3,4] and Marcus Buschbeck [1,12] ✉

NAD metabolism is essential for all forms of life. Compartmental regulation of NAD$^+$ consumption, especially between the nucleus and the mitochondria, is required for energy homeostasis. However, how compartmental regulation evolved remains unclear. In the present study, we investigated the evolution of the macrodomain-containing histone variant macroH2A1.1, an integral chromatin component that limits nuclear NAD$^+$ consumption by inhibiting poly(ADP-ribose) polymerase 1 in vertebrate cells. We found that macroH2A originated in premetazoan protists. The crystal structure of the macroH2A macrodomain from the protist *Capsaspora owczarzaki* allowed us to identify highly conserved principles of ligand binding and pinpoint key residue substitutions, selected for during the evolution of the vertebrate stem lineage. Metabolic characterization of the *Capsaspora* lifecycle suggested that the metabolic function of macroH2A was associated with nonproliferative stages. Taken together, we provide insight into the evolution of a chromatin element involved in compartmental NAD regulation, relevant for understanding its metabolism and potential therapeutic applications.

NAD metabolism plays an essential role in all domains of life[1,2]. NAD functions as a redox cofactor and a signaling molecule. As a redox cofactor in catabolic reactions, it enables ATP production in mitochondria[3]. As a signaling molecule in its oxidized form, NAD$^+$, serves as a donor of ADP-ribose moieties for effector enzymes such as poly(ADP-ribose) polymerases (PARPs), sirtuins and the CD38 family of hydrolases[4]. The biosynthetic enzymes and effectors involved in NAD metabolism are differentially distributed in the cell, making NAD metabolism highly compartmentalized[5,6]. In particular, the balance between NAD$^+$ consumption in the nucleus and its availability for cytosolic and mitochondrial redox reactions is essential for energy homeostasis[6,7]. A major challenge for the field is to understand how NAD metabolism is regulated on the compartmental level, and to dissect the relevance of this regulation for health and disease[5].

PARP1 is the major NAD$^+$-consuming enzyme in the nucleus, best known for its function as a sensor in the DNA-damage response[8]. The inhibition of PARP1 results in increased global NAD$^+$ levels in both cultured cells and mice, indicating that PARP1 also consumes NAD$^+$ under basal physiological conditions[9–11]. Conversely, nuclear NAD$^+$ levels regulate PARP1 activity[12,13]. A picture is emerging in which the nuclear NAD$^+$ consumption by PARP1 in differentiated cells needs to be kept low to maintain NAD-dependent functions in other compartments[14]. In particular, during myogenic differentiation this is achieved by the transcriptional downregulation

of PARP1 expression[15] and the simultaneous upregulation of its endogenous inhibitor, macrodomain-containing histone variant macroH2A1.1 (ref. [11]).

Macrodomains are ancient globular protein modules that have emerged as key players in NAD$^+$-dependent ADP-ribose signaling[16]. They bind ADP-ribose as a free molecule, as an oligomer or, when covalently bound to proteins, as a post-translational modification[17–19], whereas in some cases they have hydrolyzing activity[20,21]. The macroH2A family of histone variants is the only structural chromatin component containing macrodomains. In vertebrates, two genes and one event of alternative splicing give rise to three macroH2A proteins that differ in their macrodomains[22]. The capacity to bind ADP-ribose is limited to the macrodomain of the splice variant macroH2A1.1 (refs. [23,24]). As a consequence, macroH2A1.1, but not macroH2A1.2 or macroH2A2, binds auto-ADP-ribosylated PARP1 (ref. [19]) and contributes to its enrichment on specific chromatin regions[25]. MacroH2A1.1 can inhibit PARP1 activity and thus interfere with PARP1-dependent processes[24,26]. In differentiated muscle cells, macroH2A1.1 reduces nuclear NAD$^+$ consumption by PARP1 and increases the availability of NAD$^+$ in other compartments, thus indirectly promoting respiration and ATP production in mitochondria[11].

Hence, macroH2A1.1 is a chromatin component that takes part in the compartmentalized regulation of NAD metabolism in vertebrates. In the present study, we addressed the question of when

macroH2A histone variants emerged on an evolutionary scale and whether their implication in metabolism was an ancient trait. For this, we performed a phylogenetic analysis to determine the order of events in the evolution of macroH2A histone variants and characterized the function of a macroH2A macrodomain from one of the most divergent macroH2A-containing species in comparison to vertebrates.

Results

MacroH2A first appeared in unicellular protists. The amino acids encoded by the mutually exclusive exon 5 determine the capacity of macroH2A1.1 to bind ADP-ribose, inhibit PARP1 and thus affect NAD metabolism (Fig. 1a). To understand when this role of macroH2A1.1 evolved, we first aimed to determine when the fusion between the histone fold and the metabolite-binding macrodomain occurred. For this, we analyzed genomic and transcriptomic sequencing data representing a wide diversity of eukaryotes and identified a *MACROH2A* gene in 330 of them (Supplementary Tables 1 and 2). Previously, we reported the presence of macroH2A in a unicellular filasteran, *Capsaspora* spp.[27]. Together with animals and fungi, filastereans belong to the group of opisthokonts[28]. In the present study, we report the presence of macroH2A sequences in protists that diverged earlier than opisthokonts, such as the haptist *Choanocystis* sp. and the breviate *Pygsuia biforma* (Fig. 1b and Supplementary Fig. 1). It is interesting that we did not find macroH2A in fungi or other unicellular opisthokonts. The presence of macroH2A became more prevalent with the emergence of animal multicellularity. Indeed, we found macroH2A-encoding genes in diverse species of most animal phyla and all vertebrates. On the other hand, and as previously reported[16,27], macroH2A was absent from several nonvertebrate species, such as *Drosophila melanogaster* and *Caenorhabditis elegans*, as well as tunicates (Fig. 1b), which is indicative of lineage-specific losses. The second macroH2A gene corresponding to human *MACROH2A2* appeared after the whole-genome duplication in the last common ancestor of vertebrates, followed by the appearance of the alternative splicing variant of the ancestral *MACROH2A1* in a common ancestor of jawed vertebrates (Fig. 1b). All three vertebrate macrodomains show a substantial level of conservation that placed them closer to the highly conserved, replication-coupled histone H2B than the fast-evolving histone variant H2A.Bbd (Fig. 1c). Comparison of the amino acid sequences corresponding to exon 5 indicated that nonvertebrate macroH2A is most similar to macroH2A1.1 (Fig. 1d). Importantly, this included a high level of conservation of amino acids known to be required for ADP-ribose binding in human macroH2A1.1 such as Asp203 and Gly224 (refs. [17,23]).

Taken together, these results allowed us to describe the evolutionary order of events that resulted in the three different macroH2A histone variants present in vertebrates (Fig. 1e). Importantly, we found that macroH2A is much older than previously reported and originated in premetazoan protists. It was exciting to find that the first macroH2A gene is similar to vertebrate macroH2A1.1. This suggests that the function of the macroH2A histone variant in nuclear NAD[+] metabolism might be ancient. The high conservation of the macrodomains of macroH2A1.2 and macroH2A2 suggests that they have acquired new, and yet unknown, binding functions.

A protist macrodomain has higher ADP-ribose affinity. Given the origin of macroH2A before metazoans (animals), we sought to determine its potential ancestral metabolic implication by determining the biochemical properties of macroH2A in one of the protist organisms. As a model system we used *Capsaspora owczarzaki*, one of the closest unicellular relatives of animals[29]. The macrodomain of *C. owczarzaki* interacted with ADP-ribose in a similar fashion to the murine macroH2A1.1 macrodomain, as suggested by nuclear magnetic resonance (NMR)-based binding spectra

(Fig. 2a and Extended Data Fig. 1a,b). However, the ADP-ribose binding by the *Capsaspora* macrodomain was 8× stronger than by the macrodomain of mouse macroH2A1.1, with equilibrium dissociation constants (K_d) of 1.3 μM and 11.3 μM, respectively (Fig. 2b and Extended Data Fig. 1c–e). The thermodynamic profiles indicated that the two macrodomains bind ADP-ribose using a different binding mode. The *Capsaspora* macrodomain bound ADP-ribose through favorable enthalpic and entropic contributions, whereas the mouse macroH2A1.1 macrodomain bound it mainly in an enthalpy-driven manner, which was partially compensated by an unfavorable entropic contribution (Fig. 2c). Furthermore, *Capsaspora* macroH2A macrodomain showed high selectivity for ADP-ribose binding, because its affinity toward ADP was ~50-fold lower, and no interaction was observed with related nucleotides (ATP, AMP, GDP) or ribose (Extended Data Fig. 1f).

To characterize the *Capsaspora* macroH2A macrodomain and its interactions with ADP-ribose at the atomic level, we solved the structure of the protein in the presence and absence of ADP-ribose by protein X-ray crystallography (Fig. 2d,e and Supplementary Table 3). The unliganded macroH2A macrodomain crystallized in the space group P12$_1$1 and could be refined to 1.4-Å (0.14-nm) resolution (Protein Data Bank (PDB) accession no. 7NY6), whereas the ADP-ribose-bound protein crystallized in the space group P3$_2$21 and was refined to 2.0-Å resolution (PDB accession no. 7NY7). The obtained globular structures, with seven central β-sheets in the characteristic 1276354 order surrounded by α-helices (Extended Data Fig. 2a,b), showed high structural similarity to the previously described macrodomains[17,23,30]. Although the C-terminal α-helix was not resolved in the ADP-ribose bound form of *C. owczarzaki*, its root mean square deviation was only 0.5 Å to the human ADP-ribose-bound macroH2A1.1 macrodomain (PDB accession no. 3IID). The electron density of ADP-ribose clearly revealed that the ligand situates itself within the canonical binding pocket of macroH2A macrodomain (Fig. 2f). The residues important for ligand binding included Asp203, which established a H bond with the adenine amino group, and Phe352, which stabilized the adenine ring via π-electron stacking of the aromatic rings. Furthermore, the amino group of the side chain of Asn316 established a H bond with the distal ribose of ADP-ribose (Extended Data Fig. 2c and Supplementary Table 4), which can explain the selectivity for ADP-ribose over ADP (Extended Data Fig. 1f). On ADP-ribose binding, the side chains of Gln225 and Asn316 move toward each other and establish a H bond, resulting in a conformation that encloses the central diphosphate moiety of ADP-ribose (Fig. 2g).

In summary, the capacity of the macroH2A macrodomain to bind ADP-ribose is conserved in the protist *C. owczarzaki*. Indeed, the *Capsaspora* macroH2A macrodomain bound ADP-ribose stronger than its vertebrate counterpart. This raises the possibility that evolution has selected for decreased ADP-ribose affinity along the vertebrate stem lineage.

Two evolutionarily divergent residues close the binding pocket. Next, we determined the importance of the conservation of the amino acid sequence for ADP-ribose binding and for the structural integrity of the macrodomain. *Capsaspora* macroH2A and human macroH2A1.1 macrodomains share only 50% identity at the level of the amino acid sequence (Extended Data Fig. 3a). However, the multiple sequence alignment of >300 macroH2A macrodomain sequences delineated two well-conserved regions of the protein, one located toward the N terminus, overlapping with the region encoded by exon 5, and one toward the C terminus of the macrodomain fold (Fig. 3a). It is of interest that, with the exception of Ser275, all ADP-ribose-interacting residues were located in the conserved N- and C-terminal regions. Comparison of these two regions between *Capsaspora* macroH2A and human macroH2A1.1 indicated that most amino acids involved in ADP-ribose binding
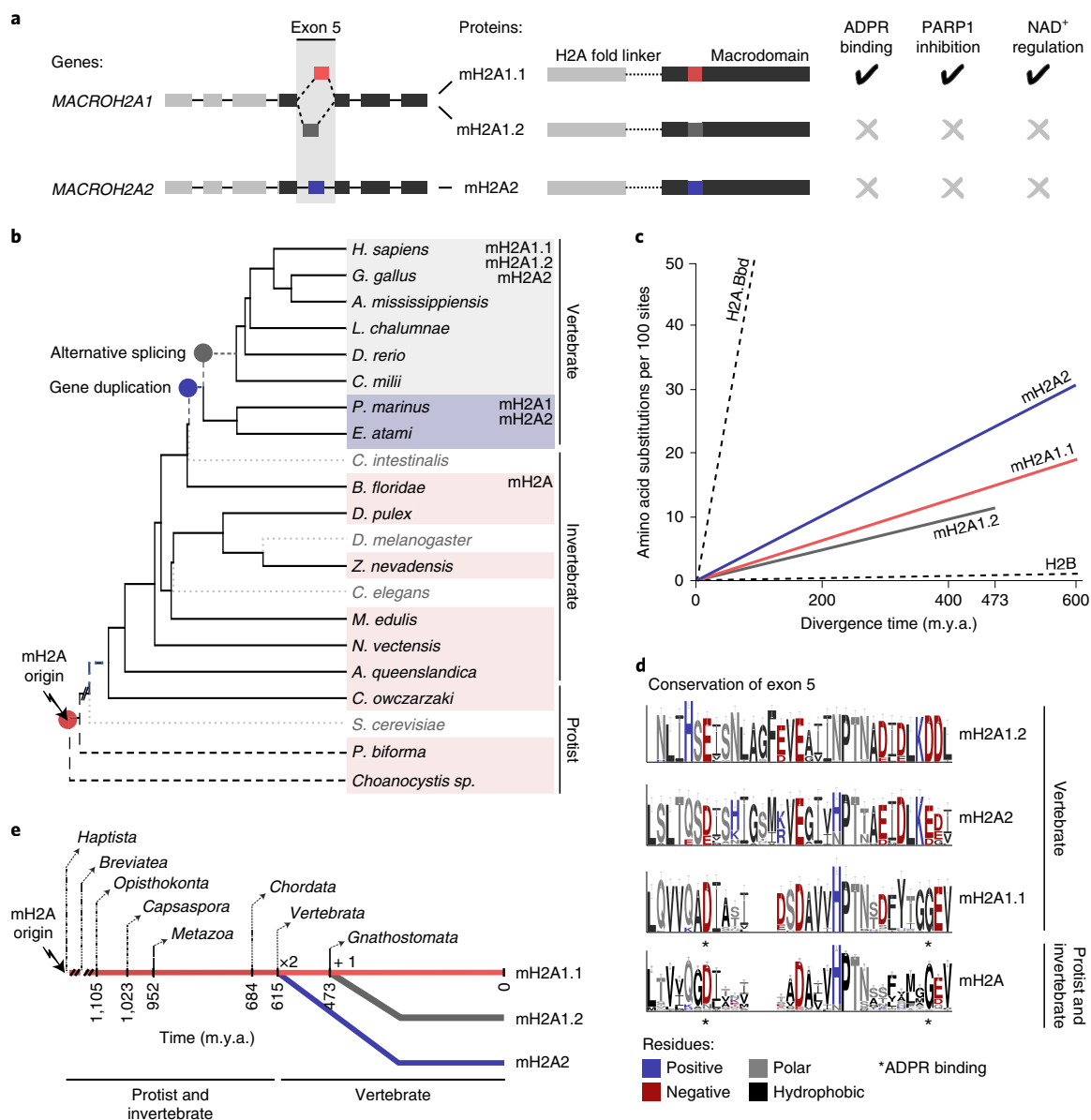
Fig. 1 | Evolutionary history of macroH2A with origin in protists. a, Schematic representation of the two genes and the alternative exon in mammals leading to three macroH2A (mH2A) proteins, of which only the macrodomain of macroH2A1.1 is able to bind ADP-ribose, inhibit PARP1 and thus reduce nuclear NAD+ consumption. **b**, Schematic tree based on the estimated divergence times of selected species. The events of macroH2A origin (arrow), gene duplication and splicing isoform emergence are represented in the tree. Dotted gray lines depict events of macroH2A loss. A full protein phylogeny based on macroH2A macrodomain sequences is shown in Supplementary Fig. 1. **c**, Estimated rates of evolution for macroH2A proteins in vertebrates. For each macroH2A variant, the evolutionary rate of the macrodomain is depicted. Pairwise amino acid identities are represented based on the species divergence times as defined in the TimeTree database. Evolutionary rates for fast-evolving histone H2A.Bbd and slow-evolving histones H2B were included as references. m.y.a., million years ago. **d**, Logo plots comparing the amino acid sequence encoded by exon 5 of macroH2A from vertebrate macroH2A1.1, macroH2A1.2 and macroH2A2, and invertebrate macroH2A (including protists). Residue colors are based on their biochemical properties: positively charged in blue, negatively charged in red, hydrophobic in black and polar in gray. Residues important for ADP-ribose (ADPR) binding in vertebrate macroH2A1.1 are marked by an asterisk. **e**, Schematic overview of the evolutionary history of macroH2A.

were invariant or structurally related, although the *Capsaspora* macroH2A macrodomain established a lower number of bonds with ADP-ribose (Fig. 3a, Extended Data Fig. 3b and Supplementary Table 4). Mapping the conservation rate on the crystal structure of the apo-form of the *Capsaspora* macroH2A macrodomain indicated a high conservation of the inner part of the ADP-ribose-binding pocket (Fig. 3b), whereas the surface regions were more variable (Fig. 3c, Extended Data Fig. 3c and Supplementary Video 1). The only two residues involved in ADP-ribose binding and with a low degree of conservation were Gln225 and Asn316 in the *Capsaspora*

macroH2A, which correspond to Glu225 and Arg315 in human macroH2A1.1. Both pairs of residues close the binding pocket on ADP-ribose binding. However, the polar uncharged side chains of Gln225 and Asn316 establish an H-bond (Fig. 2g), whereas the corresponding residues Glu225 and Arg315 in human macroH2A1.1 establish a salt bridge to close the binding pocket on ADP-ribose binding[23]. Despite the slight difference in the orientation of these side chains in the *Capsaspora* apo form from the human apo form (Fig. 3d), both macrodomains adopt an almost identical conformation on ADP-ribose binding (Fig. 3e).
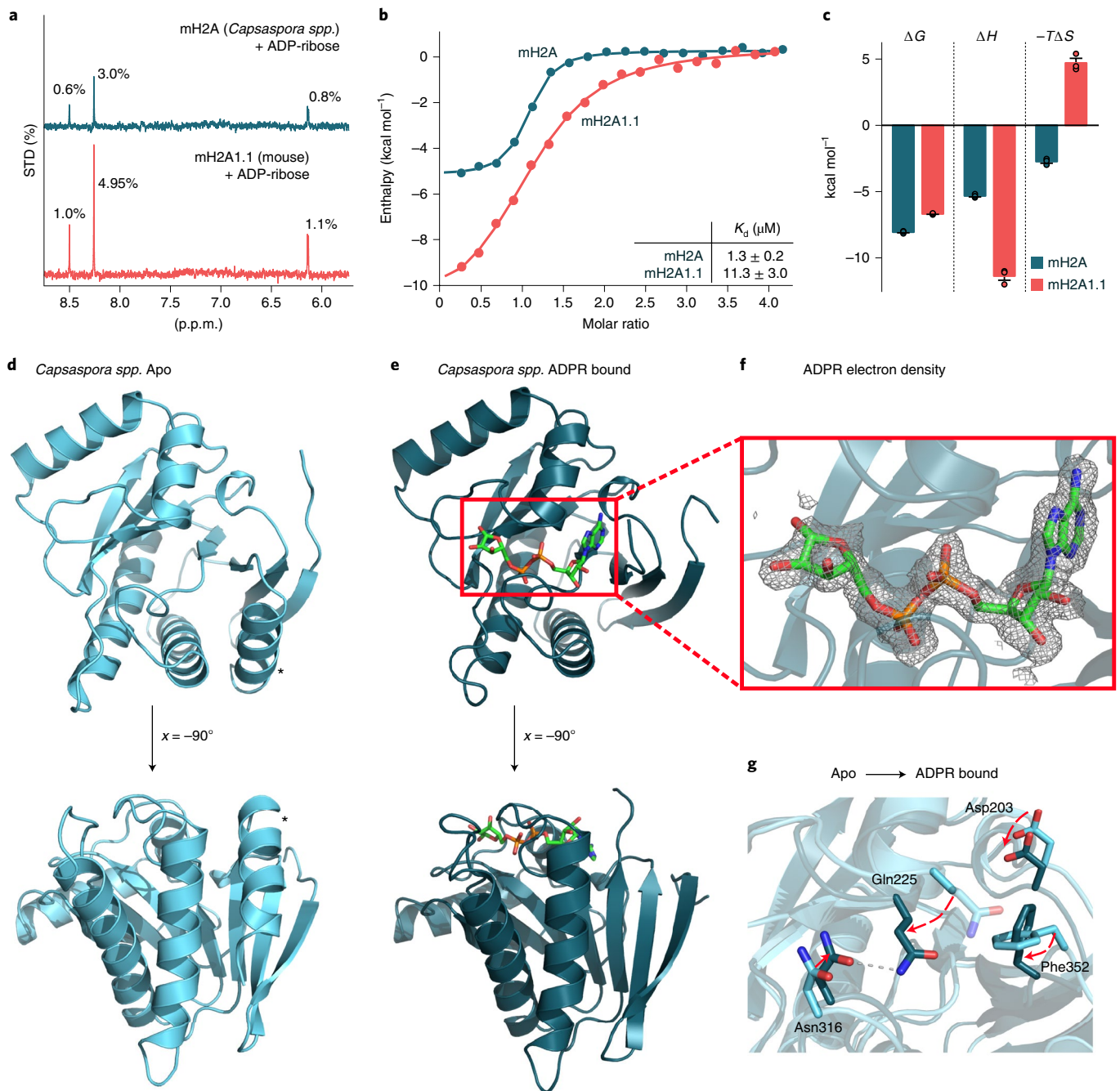
Fig. 2 | Structure and ADP-ribose binding of the protist macroH2A macrodomain. a, STD-NMR spectra indicate the presence of the interaction between purified macrodomains of *Capsaspora* macroH2A (mH2A) and mouse macroH2A1.1 (mH2A1.1) with ADP-ribose. **b**, The ITC of *Capsaspora* mH2A and mouse mH2A1.1 with ADP-ribose. The calculated K_d values are indicated. A representative graph of four independent experiments is shown. **c**, Thermodynamic parameters of **b**, shown as signature plots, indicating different contributions of enthalpy (ΔH) and entropy ($-T\Delta S$) to Gibb's free energy (ΔG). Data represent mean ($n \geq 3$) $\pm$ s.d. **d**, Crystal structure of the apo-form of the *Capsaspora* mH2A macrodomain showing entrance to the ADP-ribose-binding pocket, and rotated by $-90°$ around the *x* axis. A C-terminal α-helix is marked by an asterisk. **e**, Crystal structure of *Capsaspora* mH2A macrodomain in complex with ADP-ribose in same orientation as in **d**. **f**, Composite $2F_o - F_c$ electron density map (gray mesh, contoured at 1σ) of ADP-ribose bound to the *Capsaspora* macroH2A macrodomain. **g**, Zoom into the binding pocket of superimposed crystal structures of apo- (light-blue) and ADP-ribose-bound (deep-teal) forms of *Capsaspora* macroH2A macrodomains showing significant reorganization on ligand binding. Arrows indicate the change in position from apo to bound for residues Asp203, Gln225, Asn316 and Phe352. The side chain of Gln225 is not modeled due to ambiguous electron density. One possible conformation is shown as a transparent stick.

Taken together, the low deduced conservation at the level of primary sequence contrasts with the highly conserved three-dimensional structure of *C. owczarzaki* and human macrodomains. The residues forming the inner ADP-ribose-binding pocket are highly conserved between *C. owczarzaki* and vertebrates. Two nonconserved residues corresponding to *Capsaspora* Gln225 and Asn316 close the binding pocket of the macroH2A macrodomain after an ADP-ribose-induced fit.
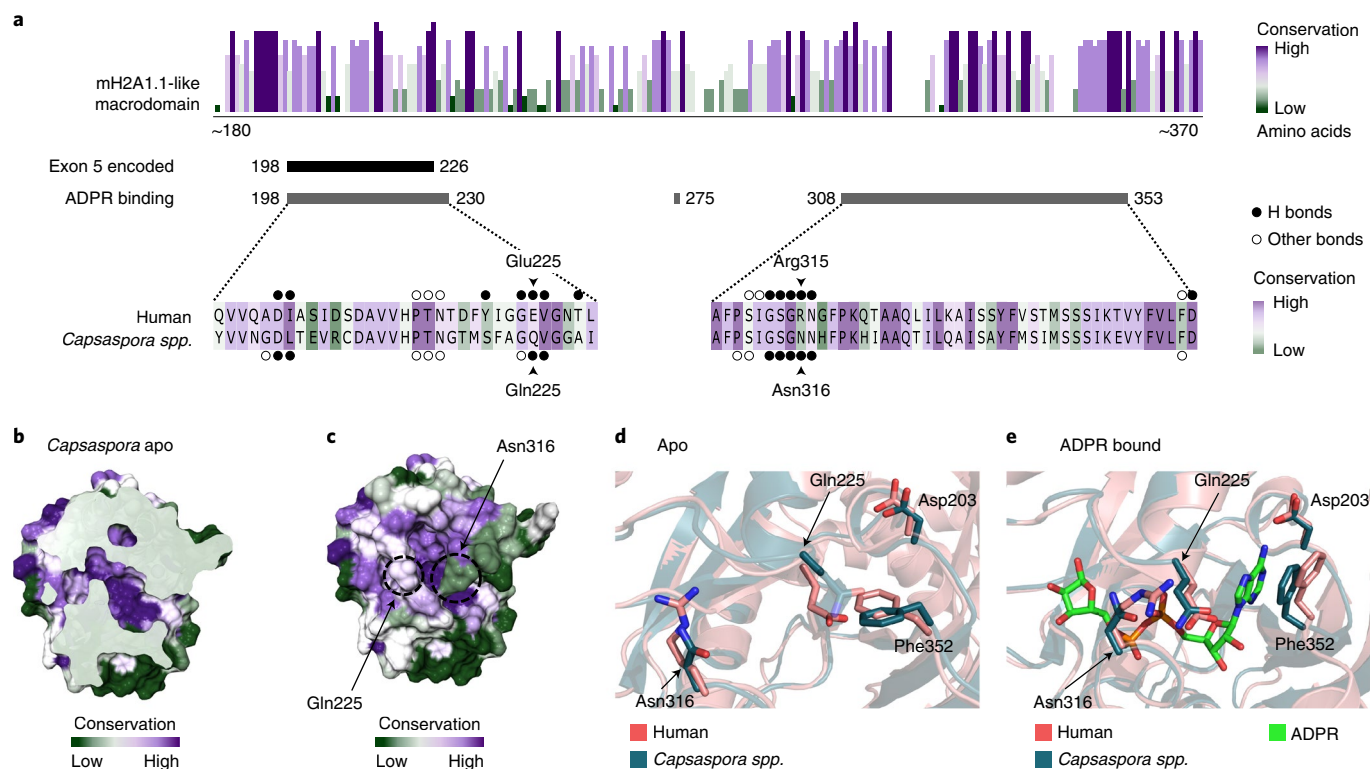
Fig. 3 | Evolutionary change in the binding mode of ADP-ribose. a, Multiple sequence alignment of >300 macroH2A1.1 (mH2A1.1)-like macrodomains indicating two highly conserved regions that highly overlap with ADP-ribose-binding residues. Amino acids encoded by exon 5 and residues establishing bonds with ADP-ribose in the *Capsaspora* macroH2A (mH2A) macrodomain (Extended Data Figs. 2c and 3b) or human mH2A1.1 are indicated. Arrows point to the two key less-conserved residues involved in ADP-ribose binding, namely Gln225 and Asn316 in *Capsaspora* macroH2A, which correspond to residues Glu225 and Arg315 in human mH2A1.1. **b**,**c**, Consurf analysis of structural conservation of mH2A1.1-like macrodomain mapped on the apo-form of the *Capsaspora* mH2A macrodomain. The horizontal cut shows strong conservation of the interior of the binding pocket (**b**), which is closed by the less-conserved amino acids Gln225 and Asn316 (**c**). **d**, Overlay of apo-forms of the macrodomains of *Capsaspora* macroH2A (dark-teal) and human macroH2A1.1 (light-red, PDB accession no. 2FXK), highlighting residues participating in the binding of ADP-ribose. The indicated residue numbers refer to the amino acids in *Capsaspora* protein. **e**, Same as in **d** for the ADP-ribose-bound forms (human macroH2A1.1, PDB accession no. 3IID). ADP-ribose is shown in green.

Succeeding substitutions of Gln225 and Asn316 reduce affinity. We hypothesized that the specific substitution of amino acids Gln225 and Asn316 drove the evolution toward the decreased ADP-ribose-binding affinity from *C. owczarzaki* to vertebrates. To test this, we reconstructed the protein phylogeny with a subset of 260 sequences and inferred the ancestral states of these two sites (Fig. 4a, Extended Data Fig. 4 and Supplementary Table 5). The obtained results revealed that the replacement of the protist-characteristic residue at position 225 (Gln225Glu) occurred early during macroH2A evolution, most likely in the ancestor of bilaterians and cnidarians, and was subsequently maintained in most animals (Fig. 4a). On the other hand, the transition to arginine at position 316 (Asn316Arg) shows more variability, as revealed by its prevalence only at specific protostome lineages, while becoming almost ubiquitous in hemichordates and chordates (Fig. 4a and Extended Data Fig. 5a–c).

To understand the physiological consequence of this evolutionary course, we individually introduced Gln225Glu and Asn316Arg mutations in the *Capsaspora* macrodomain and confirmed that the mutant proteins were folded (Extended Data Fig. 5d). It is interesting that the Gln225Glu substitution resulted in a sevenfold decrease of the affinity toward ADP-ribose compared with wild-type (WT) *Capsaspora* protein (Fig. 4b). Thus, the affinity and the thermodynamic properties of ADP-ribose binding by the Gln225Glu *Capsaspora* mutant were more similar to those of the mouse macroH2A1.1 macrodomain (Extended Data Fig. 5e,f). Strikingly, the

Asn316Arg substitution resulted in loss of ADP-ribose binding (Fig. 4b and Extended Data Fig. 5e). It is interesting that we observed alternative splicing of exon 7 affecting position 316 in Mollusca (Fig. 4c and Extended Data Fig. 6a). The alternatively spliced exon 7 was present in >25% of analyzed species (Fig. 4c and Extended Data Fig. 6b) and strikingly determined either a protist- or a vertebrate-characteristic configuration of amino acid 316 (Fig. 4d and Extended Data Fig. 6c).

Taken together, these results shed light on the evolutionary events that affected the affinity of macroH2A macrodomains toward ADP-ribose. The exchange of Gln225Glu proved to be a determining factor for the decrease in ADP-ribose affinity over the course of evolution and preceded the epistatic change of Asn316Arg.

Dynamic regulation of *Capsaspora* metabolism. To understand how the protist macroH2A is related to metabolism, we used *C. owczarzaki* as a model organism. *C. owczarzaki* has a dynamic lifecycle composed of three stages: filopodial, aggregative and cystic (Fig. 5a). As a response to environmental cues, single cells from the filopodial stage transition to the aggregative stage, one of the simple forms of multicellularity. In both filopodial and aggregative stage, cells are proliferative. In less advantageous environments, *C. owczarzaki* can transition to a spore-like cystic stage, which represents a resistance form with much smaller, nonproliferative cells[29]. Using previously generated, stage-specific RNA-sequencing (RNA-seq) data[29], we analyzed the expression levels of a curated list
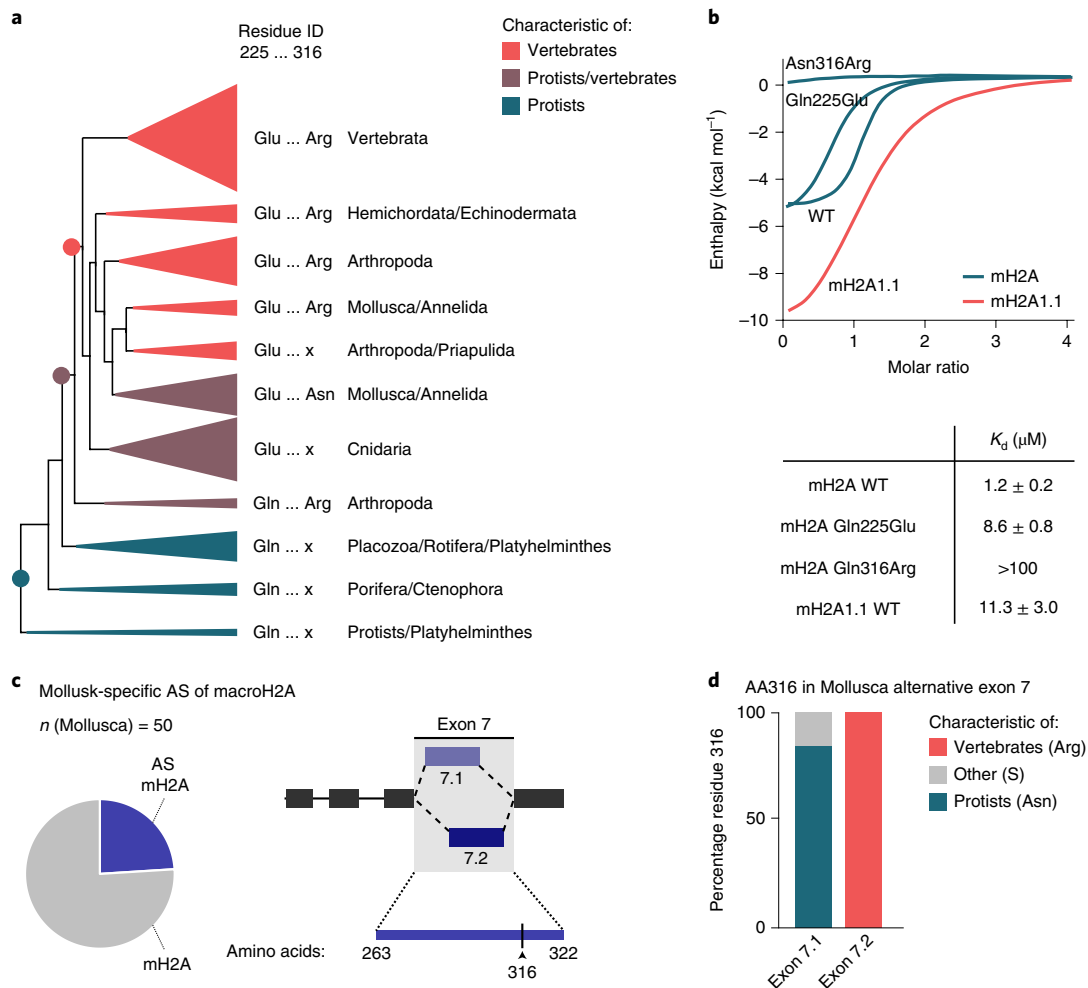
Fig. 4 | Transition from protist-characteristic to vertebrate-characteristic key residues. a, Simplified phylogenetic tree of macroH2A representing ancestral sequence reconstruction (ASR) with the focus on the identity of the amino acid residues corresponding to positions 225 and 316 in *C. owczarzaki*. For more detail on the phylogeny please see Supplementary Fig. 1 in poster format with ASR and Extended Data Fig. 4. The circles indicate three key nodes. The vertebrate-like or protist-like identity of residues 225 and 316 is indicated. **b**, ITC of *Capsaspora* macroH2A macrodomain mutants (Gln225Glu and Asn316Arg) with ADP a priori. Data of WT *Capsaspora* macroH2A macrodomain and the macrodomain of mouse macroH2A1.1 (mH2A1.1) from Fig. 1b were included as reference. The calculated K_d values are indicated. A representative graph of four independent experiments is shown. **c**, Pie chart showing that almost 25% of analyzed Mollusca have mollusk-specific alternative splicing (AS) of macroH2A. The scheme in the right panel shows the alternatively spliced exon 7, specifying the amino acid residues that this splicing affects. Residue 316 is outlined by an arrowhead. **d**, For the subset of Mollusca with AS of exon 7, the identity of the amino acid corresponding to 316 is plotted as a percentage of species for both alternative exons. In most cases, isoform 7.1 contains the protist-characteristic residue Asn316 and the isoform 7.2 has exclusively the vertebrate-characteristic Arg316. AA, amino acid.

of >400 metabolic genes, including genes encoded by mitochondria (Supplementary Table 6). The large majority of these metabolic genes were differentially expressed in at least one of the three life stages. Differentially expressed genes could be grouped in two larger and two smaller clusters by unsupervised clustering (Fig. 5b and Extended Data Fig. 7a). The larger groups contained 103 and 192 genes that were specifically up- or downregulated, respectively, in the cystic stage compared with filopodial and aggregative stages. We refer to these groups as Cys high and Cys low. Of the metabolic genes, 25 and 29 formed the smaller clusters that were differentially expressed between the filopodial and aggregative stages, placing the cystic stage in an intermediate position. These results suggested that the three life stages differ in their metabolic activity, with the cystic stage having the most divergent metabolic state.

It is interesting that almost one-fifth of the genes in the Cys high cluster were not related to anabolism or catabolism, but instead to other metabolic processes, including autophagy-related V-ATPase proton pumps (Fig. 5c and Extended Data Fig. 7b). Although the

genes involved in catabolism had a relatively lower representation in the Cys high cluster than in the Cys low one (Fig. 5c), they seemed to be more directed toward efficient mitochondrial ATP generation. For instance, mitochondrial genes encoding components of the respiratory chain were exclusively found in Cys high (Fig. 5d). In terms of anabolism, the relative proportion of anabolism-related genes in the Cys high and Cys low gene clusters was the same (Fig. 5c), but they had a different metabolic implication. NAD and NADP biosynthesis pathways were strikingly overrepresented in the Cys high cluster (Fig. 5d). Intrigued by this observation, we examined the dynamic expression of an extended set of genes related to NAD$^+$ metabolism, including effectors such as PARP1 and macroH2A. Unsupervised clustering distinguished four clusters related to their differential expression in the three life stages (Fig. 5e). Clusters 1 and 2 contained genes with increased expression in the two proliferative stages and decreased expression in the nonproliferative cystic stage (Fig. 5e). In line with the proliferative characteristics, they included most of the genes encoding
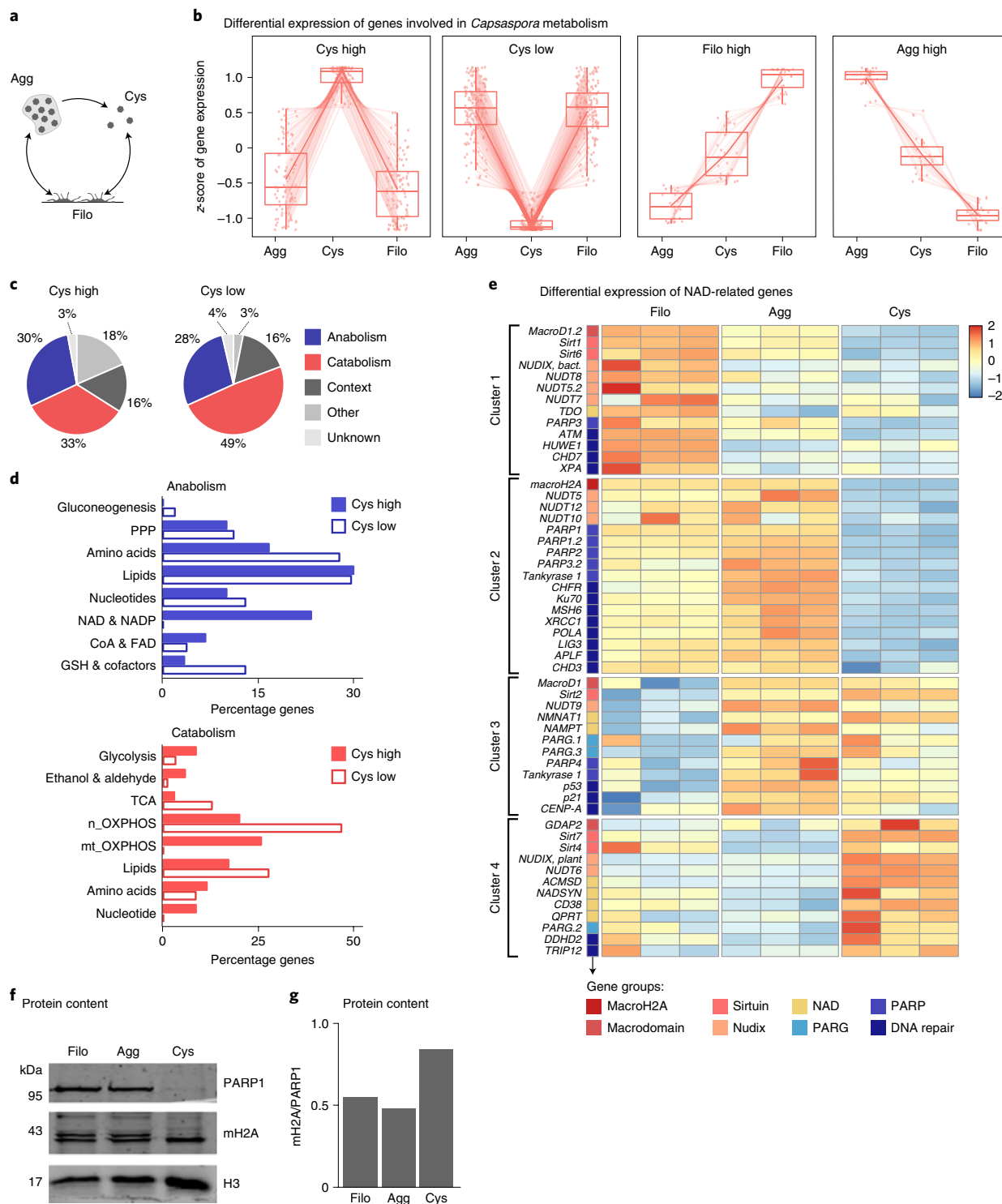
Fig. 5 | A high macroH2A:PARP1 ratio is associated with *C. owczarzaki* 's cystic stage. a, The *Capsaspora* lifecycle has three life stages: aggregative (Agg), filopodial (Filo) and cystic (Cys). The Cys stage is spore-like and nonproliferative. **b**, Most metabolic genes are differentially expressed in the three life stages of *C. owczarzaki* and fall into four major groups: Cys high (103), Cys low (192), Filo high (25) and Agg high (29). Data represent a distribution of z-scores of expressed genes ($n = 3$ biologically independent samples). Box plot parameters are detailed in Statistical analysis. **c**, Metabolic genes were classified as anabolic, catabolic, context dependent, other and unknown. Pie charts indicate their proportion in Cys high and Cys low clusters identified in **b**. **d**, Column charts showing the relative contribution of different pathways to the anabolic (top) and catabolic (bottom) component of the group of genes in Cys high and Cys low. The total number of anabolic and catabolic genes shown in **c** was set to 100%. CoA, coenzyme A; FAD, flavin adenine dinucleotide; GSH, glutathione. **e**, Differentially expressed genes related to NAD$^+$ metabolism group into four clusters. Genes were classified in eight groups and color coded as shown: macroH2A, other macrodomain proteins, sirtuins, Nudix proteins, PARPs, PARGs, DNA repair and NAD biosynthesis. **f**, Western blots of total *Capsaspora* cell extracts from the three life stages using *C. owczarzaki*-specific antibodies for macroH2A and PARP1, and histone H3 as a loading control. A representative western blot is shown ($n = 3$ independent assays). An uncropped blot image is available as Source data. **g**, Changes in the relative ratio of macroH2A (mH2A) and PARP1 as determined by mass spectrometry. The value in Filo has been set to 1.
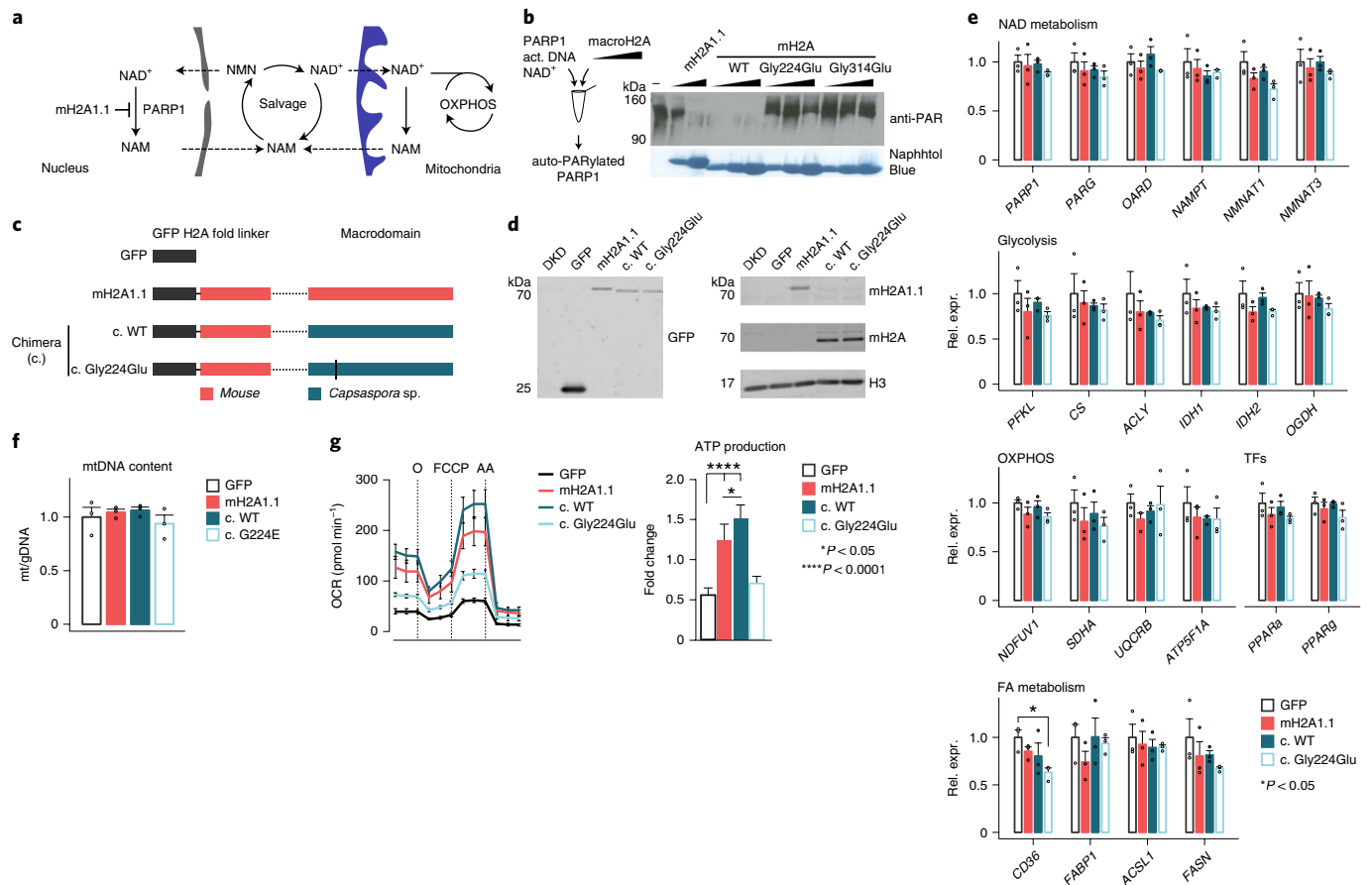
Fig. 6 | Differential impact of a protist and a vertebrate macroH2A macrodomain on PARP1 activity and cell metabolism. a, Scheme showing how the inhibition of PARP1 by macroH2A1.1 (mH2A1.1) in the nucleus is connected with mitochondrial respiration through NAD$^+$ metabolism in vertebrates. NAM, Nicotinamide; NMN, Nicotinamide mononucleotide; OXPHOS, Oxidative phosphorylation. **b**, In vitro PARP1 auto-PARylation activity (act.) induced by nicked DNA measured by anti-PAR western blotting. Naphthol Blue staining shows the increasing amounts of macrodomains that were titrated into the reaction. A representative western blot is shown ($n = 3$ independent assays). **c**, Schematic overview of the constructs that have been introduced into macroH2A-deficient HepG2 cells (referred to as DKD cells). **d**, The exogenous expression of constructs illustrated in **c** shown by western blotting using anti-macroH2A1.1, anti-*Capsaspora* macroH2A and anti-GFP. Anti-histone H3 is included as a loading control. A representative western blot is shown ($n = 3$ independent assays). **e**, The expression of genes encoding key components and regulators of indicated metabolic pathways analyzed by RT-qPCR in the cell lines described in **c** and **d**. Data represent mean ($n = 3$) ± s.d. A two-tailed Student's t-test was used to make the indicated pairwise comparisons ($^*P < 0.05$). Rel. expr., Relative expression. **f**, Mitochondrial content was assessed by measuring the relative amount of unique sequences in mitochondrial and genomic DNA (mt/gDNA). Mitochondrial DNA content normalized to nuclear DNA (genomic DNA) with MT-ND2 and NDUFV1. Data represent mean ($n = 3$) ± s.d. A two-tailed Student's t-test is used to analyze the presence of statistical differences. **g**, The oxygen consumption rate (OCR) of stable cell lines described in **c** and **d**, measured in routine culture and after subsequent addition of the ATPase inhibitor oligomycin (O), the uncoupling compound FCCP and the electron transport chain inhibitors rotenone and antimycin A (AA) (left). The bar plot shows the resulting ATP production (right). Data represent mean ($n \geq 4$) ± s.d. Ordinary one-way ANOVA was used to make the indicated comparisons between different groups of samples: $^*P < 0.05$, $^{****}P < 0.0001$. Uncropped blot images for **b** and **d** are available as Source data.

PARP enzymes and other components of the DNA-repair machinery, which might allow cells to cope with replication-associated DNA damage. Clusters 3 and 4 contained genes more prominently expressed in aggregative and cystic stages. This included genes encoding components of biosynthetic NAD pathways such as the salvage enzyme nicotinamide phosphoribosyltransferase (NAMPT) and the de novo synthesis enzyme quinolinate phosphoribosyl transferase (QPRT) (Fig. 5e).

It is noteworthy that the RNA expression patterns of PARP1 and macroH2A in cluster 2 were highly similar, suggesting coregulation at the transcriptional level (Fig. 5e). At the protein level, PARP1 was readily detected in both proliferative stages by western blotting, but its levels dropped in the cystic stage (Fig. 5e and Extended Data Fig. 7c). For macroH2A, we detected a doublet at the expected size, which collapsed into a single band in the cystic stage (Fig. 5f

and Extended Data Fig. 7d). To confirm this change in ratio, we extracted the information on PARP1 and macroH2A from available shotgun mass proteomic data[31]. Although this approach is not quantitative at the absolute level, it demonstrated that the relative ratio of macroH2A and PARP1 was the highest in the cystic stage (Fig. 5g), consistent with our western blotting results.

Taken together, gene expression data indicated that metabolism is dynamically regulated between the three life stages of the unicellular filasterean *C. owczarzaki*. It is possible that the cells in the cystic stage use salvaged and de novo synthesized NAD$^+$ for life-sustaining oxidative phosphorylation and ATP production primarily fueled by a catabolic metabolism. The relative ratio of macroH2A and PARP1 was the highest in the cystic stage, suggesting that the macroH2A-dependent inhibition of PARP1's nuclear NAD$^+$ consumption is most likely to occur in the nonproliferative stage.

Together, these metabolic adaptations can explain how *C. owczarzaki* survives in adverse conditions.

Affinity correlates with PARP1 inhibition and respiration. The inhibition of PARP1 by macroH2A1.1 in vertebrate cells reduces nuclear consumption of NAD[+], thereby increasing NAD[+] availability in the mitochondria necessary for respiration[11] (Fig. 6a). In accordance with its increased affinity toward ADP-ribose, the macrodomain of *Capsaspora* macroH2A had an increased inhibitory capacity toward PARP1 in vitro than the murine macroH2A1.1 macrodomain. This was abolished by mutations in the ADP-ribose-binding pocket, such as Gly224Glu and Gly314Glu (Fig. 6b).

Next, we sought to determine the impact of ADP-ribose binding by the *Capsaspora* macrodomain on metabolism in vivo. Currently, there are no available tools that would allow for the genetic manipulation of *C. owczarzaki*. Therefore, we used an orthogonal approach and introduced the *Capsaspora* protein into human HepG2 cells, stably depleted of all macroH2A isoforms[32]. To avoid confounding influences caused by any differences in histone-fold or linker sequences, we fused the WT and mutant macrodomains of *Capsaspora* macroH2A to the histone-fold and linker region of mouse macroH2A1.1, respectively (Fig. 6c). The expression levels of the WT reached approximately half the level of full-length mouse macroH2A1.1 and were in the range of the endogenous levels of macroH2A proteins in parental HepG2 cells (Fig. 6d and Extended Data Fig. 8a–c). We confirmed that the green fluorescent protein (GFP)-tagged chimeric and murine proteins were fully incorporated into chromatin (Extended Data Fig. 8d) and in contact with PARP1 (Extended Data Fig. 8e). Key metabolic genes were largely unaffected by the expression of the different macroH2A constructs (Fig. 6e). Furthermore, the mitochondrial content was similar across the four cell lines (Fig. 6f). However, when analyzing the oxygen consumption, we found that both the basal and the maximal mitochondrial respiration increased in the presence of mouse macroH2A1.1 (Fig. 6g). Strikingly, despite its lower expression level, this was even more pronounced in the case of the chimeric protein containing the *Capsaspora* WT macrodomain and translated into a significantly increased calculated ATP production (Fig. 6g). The Gly224Glu mutant macrodomain was inert, further substantiating the requirement for a functional and intact ADP-ribose pocket. In sum, our results show that the increased ADP-ribose affinity of *Capsaspora* macroH2A translated into stronger PARP1 inhibitory capacity and a more pronounced impact on mitochondrial respiration, when compared with mouse macroH2A1.1.

Taken together, the results of the present study suggest that the capacity of macroH2A to bind ADP-ribose, inhibit PARP1 and dampen its nuclear NAD[+] consumption is an ancient trait that was already present in protists. During evolution, a reduction in ADP-ribose affinity, mediated by changes in the residues that close the binding pocket on ligand binding, fine-tuned the stringency of this mechanism.

Discussion

The origin of macrodomain-containing histone variants. The macrodomain is the defining feature of all macroH2A histone variants. By focusing our evolutionary analysis on the amino acid sequence of macroH2A macrodomains, we were able to delineate the evolutionary history of this atypical histone variant. MacroH2A first appeared in protists ancestral to modern animals, filastereans and breviates, with the original gene resembling that of macroH2A1.1. The presence of the macroH2A gene was retained and further diversified in vertebrates, whereas it was sporadically lost in some invertebrates with accelerated evolution, such as *Drosophila* spp. It is of interest that the loss of macroH2A in these species correlates with a reduced number of *PARP* genes[16]. Gene duplication in a common ancestor of vertebrates resulted in the appearance of

macroH2A2, an isoform deficient in ADP-ribose binding. Consecutively, the alternatively spliced exon encoding macroH2A1.2 appeared in a common ancestor of jawed vertebrates, adding to a second example of an NAD signaling-inert macroH2A histone variant that can be incorporated into chromatin. The presence of macroH2A-encoding genes in haptist and breviate suggest that the fusion of a macrodomain and a histone fold occurred before the split between fungi and animals, and thus much earlier than previously thought. Due to the scarcity of the data, we cannot fully rule out the possibility that the appearance of macroH2A in the haptist or breviate was caused by independent fusion events as an example of convergent evolution.

Macrodomains are present in all forms of life. Some viral and bacterial macrodomains bind ADP-ribose[33], whereas the first extensively characterized ADP-ribose-binding macrodomain was archaeal[17]. Histone proteins, including the histone variant macroH2A, have a longer half-life than average cellular proteins[34–37]. It is intriguing to speculate that the fusion of a macrodomain to a histone provided cells with an abundant and strictly nuclear element with an increased half-life for the regulation of NAD-dependent reactions. The ability to influence NAD[+]-dependent reactions in a compartmentalized manner might have provided an advantage to eukaryotes, allowing them to adapt to changes in their environment by adopting different states, consistent with our analysis of *Capsaspora* life stages. In addition, early macroH2A might have had functions in ADP-ribose signaling.

Evolution reduced ADP-ribose-binding affinity. A major conclusion of the present study is that ADP-ribose binding is the most ancient trait of the histone variant macroH2A. The comparative analysis of *C. owczarzaki* and vertebrate macroH2A macrodomains provides us with a better understanding of how the ADP-ribose- and NAD metabolism-related functions of macroH2A were shaped through evolution. The *Capsaspora* macrodomain bound ADP-ribose with almost ten times higher affinity than its mouse counterpart and, consequently, was a much more potent PARP1 inhibitor. We were able to map this functional difference to the substitution of only two residues that close the binding pocket in the *Capsaspora* macrodomain, Gln225 and Asn316. The ancestral sequence reconstruction is consistent with the Gln225Glu replacement occurring as early as in the common ancestor of cnidarians and bilaterian metazoans, leading to a decreased ADP-ribose affinity, which was maintained in most animal groups. Similarly, the Asn316Arg replacement seems to have occurred early during macroH2A evolution, but was sporadically lost in several protostome groups. Asn316Arg is strongly represented among deuterostomes, although the physiological reason remains less clear. It is of interest that a similar course of evolution was recently reported for hemoglobin, where only two historical substitutions in the ancestral protein decrease oxygen affinity, while enabling tetramerization and cooperativity[38].

NAD[+] and macroH2A sustain nonproliferative life stages. The decreased stringency of macroH2A-dependent regulation was probably selected for along the vertebrate stem lineage. In vertebrates, macroH2A1.1 takes part in crosscompartmental regulation of NAD metabolism by inhibiting PARP1 activity in the nucleus. This function scales with the expression of the macroH2A1.1 isoform and was particularly prominent in nonproliferative myotubes where macroH2A1.1 is expressed at slightly higher levels than PARP1 (refs. [11,14]). At the present, it is unclear how the inhibitory effect of macroH2A1.1 is mediated at the molecular level. We speculate that the binding of the macrodomain to mono-ADP-ribosylated PARP1 or PARP1 modified with short-chain PARylation could interfere with conformational changes required for PARP1 activity[14].

C. owczarzaki has three different life stages: two proliferative stages and a nonproliferative, spore-like, cystic stage, to which it

transitions in unfavorable environmental conditions[29]. We found that the ratio between macroH2A and PARP1 was the highest in the cystic stage, suggesting that the putative macroH2A-dependent compartmental regulation of NAD metabolism might mostly operate in the nonproliferative stage of *C. owczarzaki*, similar to what was observed in myotubes. The cystic stage was further characterized by high expression of catabolic, mitochondrially encoded genes encoding components of the respiratory chain. Although anabolic pathways were overall downregulated in the cystic stage, genes involved in biosynthesis of NAD and precursors were upregulated. Curiously, several reports indicate the importance of ADP-ribose and NAD during sporulation and germination of bacterial spores[39,40]. Bacteria might rely on their NAD-based redox potential for germination; more specifically the accumulation of the reduced forms may have an important role in the initiation of germination[41]. It will be interesting to test whether a similar mechanism enables *C. owczarzaki* to re-enter the proliferative stages of its lifecycle. Taken together, our results indicate that *C. owczarzaki* uses a combination of both eukaryotic and bacterial mechanisms for survival in nutrient-poor environments. Our data indicate that NAD biosynthesis is channeled to life-sustaining catabolic reactions in the cystic stage. This coincides with a high macroH2A:PARP1 ratio, which has the potential to limit nuclear NAD+ consumption by PARP1. The experimental proof is pending the development of genetic tools.

The need for compartmental regulation of NAD metabolism. NAD homeostasis is vital for optimal cell function and, by extension, for organismal health[1,2]. The NAD+ pools of independent compartments communicate and are connected through the shared NAMPT reaction of the salvage pathway, and transport of NAD+ and its precursors[5], and thus creating a requirement for communicating and regulating local needs. But why is the compartmental regulation of NAD levels essential? NAD+-dependent enzymes differ in their Michaelis–Menten constant (K_m) as much as 100-fold, from $2\,\mu M$ to $1,000\,\mu M$[7,42]. Thus, their activities are controlled by the local NAD+ concentrations, which differ between cellular compartments[13,43]. This is well illustrated by the example of nuclear NAD+-consuming enzymes PARP1, sirtuin 1 (SirT1) and PARP2, which have decreasing NAD+ affinity, respectively. They form a regulatory loop, whereby SirT1 can inhibit PARP1, whereas PARP2 regulates the activity of SirT1, depending on the nuclear NAD+ levels[7]. Furthermore, some of the nuclear NAD+-consuming enzymes, such as SirT6, have an even higher NAD+ affinity than PARP1. It is interesting that macroH2A1.1 has been shown to interact with ADP-ribosylated SirT7 in an ADP-ribose-binding pocket-dependent manner[44]. This raises the possibility that macroH2A1.1 may be involved in a more general regulation of nuclear NAD+ metabolism and ADP-ribose signaling.

Compartmental NAD+ regulation in the context of evolution. The fine-tuned regulation of NAD+ consumption and localization is particularly relevant during cell differentiation, when the requirements for NAD+ compartmentalization change[13]. It is conceivable that nonproliferative states require less NAD+ for nuclear functions, such as replication-associated DNA repair, and thus benefit from prioritizing NAD usage for life-sustaining functions, such as ATP production through mitochondrial respiration. The function of macroH2A1.1 as a nuclear NAD+ regulator was first demonstrated in differentiated myotubes[11]. However, macroH2A1.1 is expressed in a wide array of tissues[24]. In addition, its upregulation is also observed during the differentiation of other tissues apart from muscle, such as skin and colon[45,46]. This suggests that its function is more widespread and generally related to cell differentiation and increased cellular plasticity in animals. In this regard, it is worth noting that *C. owczarzaki*, one of the closest relatives of animals with a complex lifecycle, shares several mechanisms of spatial cell differentiation with animals[31,47].

Furthermore, the consolidation of macroH2A with decreased ADP-ribose affinity in vertebrates coincided with the diversification of NAD biosynthesis pathways, which provided additional elements of regulation[48]. The increased complexity of higher organisms requires intricate fine-tuning of cell processes. This is often achieved by increasing the number of proteins in a regulatory network to allow for efficient sensing of subtle changes in the environment, thus enabling a fast response to environmental cues. It has been suggested that changes at the periphery of metabolic networks, possibly encoded by nonessential genes, are more likely to endow the system with the high probability of gaining beneficial changes than the changes in the rigid core of the pathway encoded by essential genes[49]. Our data suggest that the emergence of macroH2A in protists could be such a peripheral change in the network of NAD metabolism, and that it has been selected for during the evolution of metazoans. However, macroH2A is not essential for multicellular life itself because several animal species have lost macroH2A and macroH2A-deficient mice are viable[50].

In the present study, we described the evolution of a histone variant that can act as an inhibitor of nuclear NAD+ consumption, adding a unique mechanism for compartmental metabolic regulation. A better understanding of these regulatory mechanisms will be informative for ongoing development of the therapies targeting NAD metabolism and signaling[1,2]. Future work will have to further elucidate how the metabolic function of macroH2A integrates with its other molecular functions, such as the regulation of higher-order chromatin architecture[32], DNA repair[51] and transcription[52].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-021-00692-5.

References

1. Covarrubias, A. J., Perrone, R., Grozio, A. & Verdin, E. NAD+ metabolism and its roles in cellular processes during ageing. *Nat. Rev. Mol. Cell Biol.* **22**, 119–141 (2021).
2. Rajman, L., Chwalek, K. & Sinclair, D. A. Therapeutic potential of NAD-boosting molecules: the in vivo evidence. *Cell Metab.* **27**, 529–547 (2018).
3. Xiao, W., Wang, R. S., Handy, D. E. & Loscalzo, J. NAD(H) and NADP(H) redox couples and cellular energy metabolism. *Antioxid. Redox Signal* **28**, 251–272 (2018).
4. Palazzo, L., Mikolčević, P., Mikoč, A. & Ahel, I. ADP-ribosylation signalling and human disease. *Open Biol.* https://doi.org/10.1098/rsob.190041 (2019).
5. Cambronne, X. A. & Kraus, W. L. Location, location, location: compartmentalization of NAD+ synthesis and functions in mammalian cells. *Trends Biochem. Sci.* **45**, 858–873 (2020).
6. Strømland, Ø. et al. Keeping the balance in NAD metabolism. *Biochem. Soc. Trans.* **47**, 119–130 (2019).
7. Cantó, C., Menzies, K. J. & Auwerx, J. NAD+ metabolism and the control of energy homeostasis: a balancing act between mitochondria and the nucleus. *Cell Metab.* **22**, 31–53 (2015).
8. Altmeyer, M. & Hottiger, M. O. Poly(ADP-ribose) polymerase 1 at the crossroad of metabolic stress and inflammation in aging. *Aging* **1**, 458–469 (2009).
9. Bai, P. et al. PARP-1 inhibition increases mitochondrial metabolism through SIRT1 activation. *Cell Metab.* **13**, 461–468 (2011).
10. Pirinen, E. et al. Pharmacological inhibition of poly(ADP-ribose) polymerases improves fitness and mitochondrial function in skeletal muscle. *Cell Metab.* **19**, 1034–1041 (2014).
11. Posavec-Marjanović, M. et al. MacroH2A1.1 regulates mitochondrial respiration by limiting nuclear NAD+ consumption. *Nat. Struct. Mol. Biol.* **24**, 902–910 (2017).
12. Luo, X. et al. PARP-1 controls the adipogenic transcriptional program by PARylating C/EBPβ and modulating its transcriptional activity. *Mol. Cell* **65**, 260–271 (2017).

13. Ryu, K. W. et al. Metabolic regulation of transcription through compartmentalized NAD+ biosynthesis. *Science* **360**, eaan5780 (2018).

14. Hurtado-Bagès, S. et al. The histone variant macroH2A1 regulates key genes for myogenic cell fusion in a splice-isoform dependent manner. *Cells* **9**, 1109 (2020).

15. Oláh, G. et al. Differentiation-associated downregulation of poly(ADP-ribose) polymerase-1 expression in myoblasts serves to increase their resistance to oxidative stress. *PLoS ONE* **10**, e0134227 (2015).

16. Rack, J. G. M., Perina, D. & Ahel, I. Macrodomains: structure, function, evolution and catalytic activities. *Annu. Rev. Biochem.* **85**, 431–54 (2016).

17. Karras, G. I. et al. The macro domain is an ADP-ribose binding module. *EMBO J.* **24**, 1911–1920 (2005).

18. Singh, H. R. et al. A poly-ADP-ribose trigger releases the auto-inhibition of a chromatin remodeling oncogene. *Mol. Cell* **68**, 860–871.e7 (2017).

19. Timinszky, G. et al. A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation. *Nat. Struct. Mol. Biol.* **16**, 923–929 (2009).

20. Jankevicius, G. et al. A family of macrodomain proteins reverses cellular mono-ADP-ribosylation. *Nat. Struct. Mol. Biol.* **20**, 508–514 (2013).

21. Rosenthal, F. et al. Macrodomain-containing proteins are new mono-ADP-ribosylhydrolases. *Nat. Struct. Mol. Biol.* **20**, 502–507 (2013).

22. Buschbeck, M. & Hake, S. B. Variants of core histones and their roles in development, stem cells and cancer. *Nat. Publ. Gr.* https://doi.org/10.1038/nrm.2016.166 (2017).

23. Kustatscher, G., Hothorn, M., Pugieux, C., Scheffzek, K. & Ladurner, A. G. Splicing regulates NAD metabolite binding to histone macroH2A. *Nat. Struct. Mol. Biol.* **12**, 624–625 (2005).

24. Kozlowski, M. et al. MacroH2A histone variants limit chromatin plasticity through two distinct mechanisms. *EMBO Rep.* **19**, e44445 (2018).

25. Chen, H. et al. MacroH2A1.1 and PARP-1 cooperate to regulate transcription by promoting CBP-mediated H2B acetylation. *Nat. Struct. Mol. Biol.* **21**, 981–989 (2014).

26. Ouararhni, K. et al. The histone variant mH2A1.1 interferes with transcription by down-regulating PARP-1 enzymatic activity. *Genes Dev.* **20**, 3324–3336 (2006).

27. Rivera-Casas, C., Gonzalez-Romero, R., Cheema, M. S., Ausió, J. & Eirín-López, J. M. The characterization of macroH2A beyond vertebrates supports an ancestral origin and conserved role for histone variants in chromatin. *Epigenetics* **11**, 415–425 (2016).

28. Torruella, G. et al. Phylogenomics reveals convergent evolution of lifestyles in close relatives of animals and fungi. *Curr. Biol.* **25**, 2404–2410 (2015).

29. Sebé-Pedrós, A. et al. Regulated aggregative multicellularity in a close unicellular relative of metazoa. *eLife* **2013**, e01287 (2013).

30. Allen, M. D., Buckle, A. M., Cordell, S. C., Löwe, J. & Bycroft, M. The crystal structure of AF1521 a protein from *Archaeoglobus fulgidus* with homology to the non-histone domain of macroH2A. *J. Mol. Biol.* **330**, 503–511 (2003).

31. Sebé-Pedrós, A. et al. High-throughput proteomics reveals the unicellular roots of animal phosphosignaling and cell differentiation. *Dev. Cell* **39**, 186–197 (2016).

32. Douet, J. et al. MacroH2A histone variants maintain nuclear organization and heterochromatin architecture. *J. Cell Sci.* **130**, 1570–1582 (2017).

33. Catara, G., Corteggio, A., Valente, C., Grimaldi, G. & Palazzo, L. Targeting ADP-ribosylation as an antimicrobial strategy. *Biochem. Pharmacol.* **167**, 13–26 (2019).

34. Chen, W., Smeekens, J. M. & Wu, R. Systematic study of the dynamics and half-lives of newly synthesized proteins in human cells. *Chem. Sci.* **7**, 1393–1400 (2016).

35. Commerford, S. L., Carsten, A. L. & Cronkite, E. P. Histone turnover within nonproliferating cells. *Proc. Natl Acad. Sci. USA* **79**, 1163–1165 (1982).

36. Fornasiero, E. F. et al. Precisely measured protein lifetimes in the mouse brain reveal differences across tissues and subcellular fractions. *Nat. Commun.* https://doi.org/10.1038/s41467-018-06519-0 (2018).

37. Mathieson, T. et al. Systematic analysis of protein turnover in primary cells. *Nat. Commun.* **9**, 689 (2018).

38. Pillai, A. S. et al. Origin of complexity in haemoglobin evolution. *Nature* **581**, 480–485 (2020).

39. Huh, J. W., Shima, J. & Ochi, K. ADP-ribosylation of proteins in *Bacillus subtilis* and its possible importance in sporulation. *J. Bacteriol.* **178**, 4935–4941 (1996).

40. Setlow, P. & Kornberg, A. Biochemical studies of bacterial sporulation and germination. *J. Biol. Chem.* **245**, 3637–3644 (1970).

41. Setlow, R. & Setlow, P. Levels of oxidized and reduced pyridine nucleotides in dormant spores and during growth, sporulation, and spore germination of *Bacillus megaterium*. *J. Bacteriol.* **129**, 857–865 (1977).

42. Berger, F., Lau, C., Dahlmann, M. & Ziegler, M. Subcellular compartmentation and differential catalytic properties of the three human nicotinamide mononucleotide adenylyltransferase isoforms. *J. Biol. Chem.* **280**, 36334–36341 (2005).

43. Cambronne, X. A. et al. Biosensor reveals multiple sources for mitochondrial NAD+. *Science* **352**, 1474–1477 (2016).

44. Simonet, N. G. et al. SirT7 auto-ADP-ribosylation regulates glucose starvation response through mH2A1. *Sci. Adv.* https://doi.org/10.1101/719559 (2020).

45. Creppe, C. et al. MacroH2A1 regulates the balance between self-renewal and differentiation commitment in embryonic and adult stem cells. *Mol. Cell. Biol.* **32**, 1442–1452 (2012).

46. Sporn, J. C. & Jung, B. Differential regulation and predictive potential of macroH2A1 isoforms in colon cancer. *Am. J. Pathol.* **180**, 2516–2526 (2012).

47. Sebé-Pedrós, A. et al. The dynamic regulatory genome of capsaspora and the origin of animal multicellularity. *Cell* **165**, 1224–1237 (2016).

48. Bockwoldt, M. et al. Identification of evolutionary and kinetic drivers of NAD-dependent signaling. *Proc. Natl Acad. Sci. USA* **116**, 15957–15966 (2019).

49. Morowitz, H. J. A theory of biochemical organization, metabolic pathways, and evolution. *Complexity* **4**, 39–53 (1999).

50. Pehrson, J. R., Changolkar, L. N., Costanzi, C. & Leu, N. A. Mice without macroH2A histone variants. *Mol. Cell. Biol.* **34**, 4523–4533 (2014).

51. Sebastian, R. et al. Epigenetic regulation of DNA repair pathway choice by MacroH2A1 splice variants ensures genome stability. *Mol. Cell* **79**, 836–845.e7 (2020).

52. Lavigne, M. D. et al. Composite macroH2A/NRF-1 nucleosomes suppress noise and generate robustness in gene expression. *Cell Rep.* https://doi.org/10.1016/j.celrep.2015.04.022 (2015).

Methods

Molecular data mining. MacroH2A sequences were collected from the GenBank database by Blast searches using human sequences as a query. For better representation of species, especially in the transition to the vertebrate lineage, de novo assembly of transcriptomes for jawless fish (hagfish and lamprey; BioProject accession nos. PRJDB4902 and PRJNA292033, respectively) and the bowfin (BioProject accession no. PRJNA292033) were carried out using Trinity software, v.2.2.0 in the Galaxy web platform[53]. Briefly, paired-end Search Read Archive Fastq files were uploaded from the European Nucleotide Archive to the Galaxy platform and their quality analyzed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). All left and right reads were concatenated in two separate files and used as input for Trinity with default parameters. Assembled sequences were used to create local nucleotide databases and Blast searches were performed as described above.

Overall, 467 sequences encompassing 327 metazoan (58 vertebrate and 269 nonvertebrate) and 3 nonmetazoan species were retrieved. Three macroH2A variants (macroH2A1.1, macroH2A1.2 and macroH2A2) were collected for all vertebrate species except for species displaying only macroH2A1.1. For all macroH2A proteins, only the globular part of the macrodomain (amino acids 182 to the end in human sequences) was used in the analyses, unless stated otherwise. Manual Blast searches were performed for curating the data of underrepresented species, specifically for filasterean and ichthyosporean species using available information[28,54–59].

Sequence alignments. Multiple sequence alignments were performed using MAFFT v.7 (ref. [60]) and Jalview v.2 (ref. [61]), and edited for potential errors in BioEdit (v.7.2). Logo plots were generated based on the aligned sequences using WebLogo3 (ref. [62]). Multiple sequence alignments of the macroH2A1.1-like macrodomain sequences from 305 species with 1 macroH2A1.1-like isoform was generated using the alignment tool PRALINE[63]. The conservation score indicates the conservation of amino acid biochemical properties on a scale from 1 to 6, and is represented by height and color of the bar.

For pairwise sequence homology analysis, we used blast homology search and macroH2A macrodomain sequences of representative organisms as a query. The resulting percentage homology rate is represented in the homology matrix visualized using Morpheus matrix analysis and visualization software (software.broadinstitute.org/morpheus).

Phylogenetic and evolutionary analyses. If not stated otherwise, phylogenetic and molecular evolutionary analyses were conducted using MEGA X v.10.1.7 (ref. [64]). MacroH2A phylogeny and ancestral sequence states were inferred by using the maximum likelihood method with the LG substitution model[65], including gamma-distributed variation among sites. Positions with <95% site coverage were eliminated, so the analysis involved 260 amino acid sequences and a total of 181 positions in the final dataset. The reliability of the reconstructed topology was contrasted by a nonparametric bootstrap method (1,000 replicates).

Protein sequence divergence was estimated using uncorrected differences (*P* distances, partial deletion 95%), and the rates of evolution were estimated by correlating pairwise protein divergences between pairs of taxa with their corresponding divergence times as defined by the TimeTree database[66].

Metabolism-focused analysis of high-content data from *C. owczarzaki*. Previously generated raw RNA-seq data[29] were realigned using STAR v.2.7.3a[67]. The function genomeGenerate was used to include the mitochondrial transcriptome data in the described assembly of the *Capsaspora* genome[68] and quantified using featureCounts software v.2.0.1 (ref. [69]). Statistical analysis was performed using DESeq2 (ref. [70]), using the likelihood ratio test selecting genes that show an adjusted $P \leq 0.01$. Clusterization of the data was obtained using the DEGreport software package on BioconductoR.

Functional annotation was performed using EGGnog 5.0 using precomputed clusters and phylogenies[71]. The curated list of metabolic genes from Kyoto Encyclopedia of Genes and Genomes[72] was generated by retrieving *Capsaspora* genes using EGGnog 5.0 and GHOSTKoala[73]. This was further complemented by using Orthofinder[74] to identify *Capsaspora* orthologues of the human queries. The absence of orthologues was confirmed by Blast search. The resulting list has further been manually curated and genes involved in multiple pathways have been assigned to their parent metabolic pathway and categorized as anabolic, catabolic, context dependent (anabolic or catabolic), other and unknown.

The relative abundance of proteins was extracted from previously generated proteomics data using raw measurements, and averaging three replicates per condition[31].

Plasmids. For plasmid construction, we used standard cloning techniques. *Capsaspora* macroH2A sequence (National Center for Biotechnology Information (NCBI) sequence ID: XM_004347479.1) was synthesized by Life Technologies Inc. (Thermo Fisher Scientific) and cloned into pcDNA3.1 backbone with an N-terminal GFP tag. *C. owczarzaki* (Gly224Glu, Gln225Glu, Asn316Arg) and murine (Gly224Glu) macroH2A-binding pocket mutants were generated via Stratagene's site-directed mutagenesis QuickChange protocol. Macrodomain

sequences (corresponding to amino acids 182–368 for *Capsaspora* macroH2A and 182–369 for murine macroH2A1.1) were cloned into a pETM-11 backbone with an N-terminal 6 His-tag. *Capsaspora* PARP1 (NCBI sequence ID: XP_004363957.1) N-terminal sequence (corresponding to amino acids 1–350) was amplified from *Capsaspora* complementary DNA and cloned into a pETM-11 backbone with an N-terminal 6 His-tag.

For mammalian expression constructs, the full-length sequences of *Capsaspora* macroH2A, or mouse macroH2A1.1 or macroH2A1.2, were cloned into pLVX-Puro lentiviral backbone (Clontech) adding an N-terminal enhanced GFP tag. Mouse–*C. owczarzaki* chimeras were generated by sequential cloning of fragments. First, histone-fold and linker-domain sequences from mouse macroH2A1.1 were inserted into the backbone, followed by the insertion of either WT or mutant (Gly224Glu) *Capsaspora* macroH2A macrodomain. The pLVX-Puro with enhanced GFP alone was cloned and kindly provided by M. Gamble[75]. All sequences were verified by sequencing.

Protein production and purification. Rosetta (DE3), chemically competent *Escherichia coli* were transformed with bacterial expression vectors and grown in lysogeny broth medium supplemented with $34\,\mu g\,ml^{-1}$ of chloramphenicol and $50\,\mu g\,ml^{-1}$ of kanamycin at 37 °C overnight. The culture was used to inoculate 1 l of this Terrific Broth medium (Sigma) and grown at 37 °C and 200 r.p.m. until reaching an absorbance at 600 nm of 0.4–0.6. The protein expression was then induced with 0.5 mM isopropyl-β-D-1-thiogalactopyranoside for 16 h at 20 °C. The next day, bacteria were pelleted by centrifugation at 10,000*g* for 15 min at 4 °C. The bacterial pellet was lysed in 50 mM Tris, 300 mM NaCl, 10 mM imidazole and 1 mM DTT, pH 7.4, supplemented with $1\,mg\,ml^{-1}$ of lysozyme, $10\,\mu g\,ml^{-1}$ of DNase I and protease inhibitors (Roche cOMPLETE EDTA free). The lysates were cleared by centrifugation at 30,000*g* for 45 min at 4 °C. Subsequently, the cleared lysates were incubated with Ni–NTA beads (QIAGEN) for 1 h at 4 °C and passed over a gravity flow column 3×. After washing the beads with 3 column volumes using 50 mM Tris, 1 M NaCl, 10 mM imidazole and 1 mM DTT, pH 7.4, the proteins of interest were eluted with 50 mM Tris, 300 mM NaCl, 300 mM imidazole and 1 mM DTT, pH 7.4. The eluted proteins were dialyzed overnight into a phosphate buffer (50 mM KH_2PO_4 and 1 mM DTT, pH 6.5), unless stated otherwise. The purified proteins were concentrated using a 3-kDa molecular mass cutoff, centrifugal concentrator (Amicon), and then flash-frozen in liquid nitrogen and stored at −80 °C.

STD-NMR. Saturation transfer difference NMR (STD-NMR) experiments were performed as described elsewhere[76,77]. Briefly, 10 μM His-tagged *Capsaspora* macroH2A macrodomain, or murine macroH2A1.1 WT or mutant macrodomains, was dialyzed and prepared in a deuterated water buffer containing 16.2 mM Na_2HPO_4, 3.8 mM NaH_2PO_4, 1 mM tris(2-carboxyethyl) phosphine, pH 7.4 and 10 μM disuccinimidyl suberate. STD spectra of 1 mM ADP-ribose in the presence of 10 μM macroH2A macrodomains were obtained at 25 °C (298K) on a Bruker Avance 600-MHz spectrometer equipped with a cryoprobe. A pseudo-two-dimensional version of the STD-NMR sequence was used for the interleaved acquisition of on-resonance and off-resonance spectra. The on-resonance frequency was set to 0.0 p.p.m. and the saturation time was 3 s. The STD effect (%) was quantified based on the following equation: $I_{STD} = 100 \times (I_0 - I_{SAT})/I_0$, where I_{SAT} and I_0 are the intensities of a given signal in the on-resonance and the off-resonance spectra, respectively.

Isothermal titration calorimetry. Isothermal titration calorimetry (ITC) was performed as previously described[17]. Before the experiment, proteins were dialyzed overnight against 50 mM KH_2PO_4 and 1 mM DTT, pH 6.5, at 4 °C. The dialyzed proteins were then centrifuged for 20 min at 20,000*g* at 4 °C, and the protein concentration was determined by absorbance measurements at 280-nm wavelength using calculated molar extinction coefficients. The nucleotides and ribose were prepared in the same buffer in the concentration range 1–1.5 mM. The concentration of ADP-ribose was additionally confirmed by absorbance measurements at 260 nm, using a molar extinction coefficient of $13,500\,M^{-1}\,cm^{-1}$. Assays were conducted on the PEAQ-ITC instrument (MicroCal) at 25 °C, and experimental data analysis was performed using MicroCal PEAQ-ITC Analysis Software.

Thermal shift assays. Fluorescence-based thermal shift assays were performed as previously described[18]. Briefly, 5 μM protein solutions supplemented with 8× SYPRO Orange were heated in 50 mM KH_2PO_4, pH 6.5 and 1 mM DTT from 5 °C to 95 °C at a ramp rate of 1%. The assays were conducted in MicroAmp Fast 96-well reaction plates sealed with MicroAmp Optical Adhesive Film (Applied Biosystems). The fluorescence measurements at 554 nm were normalized to the lowest value before the transition and the maximum fluorescence.

Crystallization, data collection and processing. All crystallization experiments were conducted at the Crystallization Facility of the Max Planck Institute of Biochemistry (Martinsried, Germany). Before setting up the crystallization, the proteins were dialyzed overnight against 20 mM Bis–Tris, pH 7.0, at 4 °C. The dialyzed proteins were then centrifuged for 20 min at 20,000*g* and 4 °C, and the protein concentration was determined by absorbance measurements at 280 nm using calculated molar extinction coefficients.

Crystals of *Capsaspora* macroH2A macrodomain in apo-form were obtained in sitting drop vapor diffusion experiments performed at 20 °C by mixing 100 nl of 0.1 M Bis–Tris, pH 5.5 and 25% (w:v) poly(ethylene glycol) 3350 (PEG-3350) with 200 nl of a solution containing the protein at 27 mg ml⁻¹. Crystals of *Capsaspora* macroH2A macrodomain in complex with ADP-ribose were obtained in sitting drop vapor diffusion experiments performed at 20 °C by mixing 100 nl of 0.2 M ammonium tartarate dibasic and 20% (w:v) PEG-3350 with 200 nl of a solution containing the protein at 21 mg ml⁻¹ and 3.8 mM ADP-ribose. Crystals were cryoprotected by soaking in mother liquor supplemented with 30% ethylene glycol and flash cooled in liquid nitrogen. Diffraction data of proteins were collected on the Swiss Light Source or the in-house X-ray source of the Crystallization Facility of the Max Planck Institute of Biochemistry. All datasets were processed using XDS[78].

The structures were solved by molecular replacement using the human macroH2A1.1 macrodomain (PDB accession no. 3IID (ref. [19])) as a search model. Model building and real space refinement were performed in COOT[79,80] and the structures refined using PHENIX REFINE[81]. Model and restraints for ADP-ribose were prepared using Phenix.Elbow[82]. A summary of the data collection and refinement statistics is shown in Supplementary Table 3. The Ramachandran statistics for the final refined models were 97.86% favored and 2.14% allowed (apo, PDB accession no. 7NY6), and 96.55% favored and 3.45% allowed (ADP-ribose bound, PDB accession no. 7NY7). UCSF Chimera software[83] and the PyMOL Molecular Graphics System (Schrödinger, LLC) have been used for visualization.

The LigPlot diagram for the crystal structure of *Capsaspora* macroH2A macrodomain in complex with ADP-ribose was generated using an online platform LIGPLOT v.4.5.3 (ref. [84]). A Consurf bioinformatic tool[85] was used for the projection of evolutionary conservation scores. Briefly, the conservation analysis of amino acid positions was calculated based on the phylogenetic relationships between sequences of 305 species with 1 macroH2A1.1-like isoform and using *Capsaspora* apo-macroH2A macrodomain structure as a query. UCSF Chimera software was used for visualization[83].

Antibodies. We generated a specific antibody against *Capsaspora* proteins by immunizing rabbits with purified His-tagged *Capsaspora* macroH2A macrodomain or carrier protein-coupled peptides of *Capsaspora* PARP1. Specifically, we used a mix of three different peptides corresponding to amino acids 103–114, 132–142 and 299–310 of *Capsaspora* PARP1 protein. Sera were collected from terminal bleeds after three to four rounds of inoculation. The obtained antibody sera were used at 1:150 dilutions. The animal procedures were carried out by the CID-CSIC Antibody Generation Service (Spain) and UVic-animal care facility (Canada).

The following additional antibodies were used (if not stated otherwise a dilution of 1:1,000 was used for western blotting): anti-histone H3 (RRID: AB_302613, 1:10,000); anti-PAR (RRID: AB_2272987); anti-GFP (RRID: AB_1196614, 1:5,000); anti-macroH2A1.1 (ref. [86]) (1:500); anti-macroH2A1.2 (RRID: AB_1950388); anti-His-tag (RRID: AB_2744546, 1:2,000); anti-SAF-A[87] (1:500); anti-PARP-1 (RRID: AB_659884), and fluorophore-conjugated secondary anti-mouse and anti-rabbit (RRID: AB_621842 and RRID: AB_621843 at 1:20,000).

PARP1 activity assay. In vitro PARP1 activity was measured using an auto-PARylation assay as previously described[11]. Briefly, 0.3× activated DNA (diluted from 10× activated DNA, Trevigen) and 200 μM NAD⁺ were added to the buffer containing 50 mM Tris-HCl, pH 8, 50 mM NaCl and 1 mM MgCl₂. Subsequently, purified macrodomains were added to the reaction mix at the defined concentrations (10 and 50 μM for mouse, and 10, 25 and 50 μM for *Capsaspora* macroH2A macrodomains). Finally, 0.2 units μl⁻¹ of PARP1 high specific activity enzyme (Trevigen) was added to all reactions and the reaction mix was incubated at 25 °C for 20 min. Reactions were stopped by the addition of Laemmli's sample buffer and boiling at 95 °C, separation on sodium dodecylsulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and analysis by western blotting.

Culture of human and *Capsaspora* cells. Unless stated otherwise, the following conditions were used for cell culture. MacroH2A-depleted HepG2 (DKD) cells[32] and HEK293T (American Type Culture Collection (ATCC), catalog no. CRL-3216) were routinely cultured in Dulbecco's modified Eagle's medium (DMEM) containing 4.5 g l⁻¹ of glucose (Gibco) supplemented with 10% v:v fetal bovine serum (Gibco), 2 mM ʟ-glutamine (Gibco), 50 U ml⁻¹ of penicillin (Gibco) and 50 mg ml⁻¹ of streptomycin (Gibco). Cells were authenticated, incubated at 37 °C in 5% CO₂, and periodically checked for *Mycoplasma* contamination. Cells were collected by scraping, washed with phosphate-buffered saline (PBS) and pelleted. *C. owczarzaki* was cultured, and cells from all three different stages were harvested as described[29]. Before analysis, *Capsaspora* cells were washed with PBS and flash-frozen in liquid nitrogen.

Gene transduction and establishment of stable cell lines. HEK293T cells were used as packaging cells to produce viral particles for lentiviral infections. Four million HEK293T cells were seeded in P10 plates and cultured to 60–70% confluency. At that point the cells were transfected with 10 μg of the lentiviral plasmid of interest and 3 μg of the pCMV-VSV-G plasmid and pCMV-dR8.91.

Plasmid DNA was mixed in 1× HBS solution (2× HBS: 272 mM NaCl, 2.8 mM Na₂HPO₄ and 55 mM HEPES, pH 7) containing 125 mM CaCl₂ in a total volume of 800 μl. The mix was added on to the cell culture medium dropwise and left overnight. Transfection efficiency was controlled using a GFP expression vector. The supernatant containing viral particles produced by HEK293T cells was collected for 24 h at 48 h and 72 h after transfection, filtered using a 0.45-μm filter and supplemented with 8 μg ml⁻¹ of polybrene (Sigma-Aldrich). The fresh viral supernatant was added to target cells at 60–70% confluency in 6-well plates that were centrifuged for 45 min at 1,200 r.p.m. and 37 °C, incubated at 37 °C for 45 min and then cultured overnight in fresh medium. The same process was repeated 24 h after the first infection. The transduced cells were selected with 2 μg ml⁻¹ of puromycin. The necessary selection time was determined by using a negative control plasmid without resistance. This procedure was used to generate all HepG2 stable cell lines. The efficiency of the cell infection was validated by live cell fluorescence and flow cytometry.

Cell fractionation, immunoprecipitation and western blotting. For western blotting of total cell material, cell pellets were directly resuspended in Laemmli's sample buffer, sonicated using a Bioruptor Plus (Diagenode) and incubated at 95 °C for 10 min before loading the samples on polyacrylamide gels.

For chromatin lysis, nuclei from mammalian cells were isolated and sonicated to solubilize chromatin, as previously described[11]. Briefly, cells were collected by scraping and lysed in sucrose buffer (0.32 M sucrose, 10 mM Tris-HCl, pH 8.0, 3 mM CaCl₂, 2 mM MgOAc, 0.1% Triton buffer, 1:100 phenylmethylsulfonyl fluoride and 1:200 leupeptin). Samples were passed through a syringe to facilitate membrane disruption and then further incubated for 8 min on ice. Intact nuclei were collected by centrifugation for 3 min at 1,000g and washed once more in sucrose buffer. Washed nuclei were resuspended in lysis buffer (50 mM Tris-HCl, pH 7–8, 135 mM NaCl, 0.1% Triton, 1 mM EDTA, 1 mM DTT, 1:100 phenylmethylsulfonyl fluoride and 1:200 leupeptin), and chromatin was solubilized by progressive sonication with a Bioruptor Plus (Diagenode).

For the coimmunoprecipitation of PARP1, the lysis buffer was complemented with poly(ADP-ribose) glycohydrolase (PARG) and PARP inhibitors (1 μM ADP-HPD from CalBioChem and 1 μM olaparib from SelleckChem, respectively). Insoluble material was removed by centrifugation, and lysates were precleared with Sepharose beads. At this step, 5% of the total lysate was kept as input material, and the rest of the lysate was incubated for 3 h with anti-GFP nanobodies coupled to magnetic beads (ChromoTek), which were previously blocked with 1% bovine serum albumin in lysis buffer. Precipitates were washed 3× with lysis buffer containing 1% Triton X-100. For SDS–PAGE and western blotting analysis, typically 1% input and 20% immunoprecipitated material were loaded.

For cell fractionations, nuclei were prepared as described above and the supernatant was kept as the cytosolic fraction. Nuclei were then incubated with high-salt buffer (20 mM HEPES, pH 7.9, 410 mM KCl, 1.5 mM MgCl₂, 0.2 mM EDTA, 25% glycerol and 0.5% NP-40) for 30 min. Ultracentrifugation at 50,000g was used to separate the chromatin (pellet) and nucleosol (supernatant).

After the transfer of proteins to the nitrocellulose membrane (GE Healthcare), the membranes were blocked with 5% low-fat milk (Nestle) and incubated with primary antibodies overnight at 4 °C. The next day, membranes were washed with Tris-buffered saline and Tween 20 (TBST) and incubated with fluorophore-conjugated secondary antibodies for 1 h at 25 °C in the dark. After washing with TBST, the dried membranes were scanned with an Odyssey CLx Imager and analyzed using ImageStudioLite quantification software (LI-COR Biosciences).

RNA and DNA analysis. Total RNA from mammalian cells was isolated by the Maxwell RSC simplyRNA Cells Kit (Promega) using the Maxwell RSC Instrument (Promega) according to the manufacturer's instructions. Total RNA, 1 μg, was used for cDNA synthesis using the First Strand cDNA synthesis kit (Thermo Fisher Scientific) and oligo(dT) primers according to the manufacturer's instructions. Relative cDNA levels were quantified by reverse-transcription quantitative PCR (RT-qPCR) (LightCycler 480 II instrument, Roche). All samples were analyzed in technical triplicates. Values were normalized to two reference housekeeping genes (*RPLP0* and *GAPDH*) and plotted relative to a reference sample set to 1. To measure mitochondrial and genomic DNA, we extracted the total DNA from all cell lines of interest. Briefly, cells were pelleted and the DNA isolation buffer (10 mM Tris-HCl, pH 8.5, 5 mM EDTA, 0.5% SDS, 200 mM NaCl and 0.1 mg ml⁻¹ of proteinase K) was added directly to the pellets, and the samples were incubated overnight at 37 °C while shaking (Thermoblock). Proteinase K was inactivated by incubating the samples for 10 min at 99 °C. An equal volume of isopropanol was added to the lysates and left incubating for 20 min at 25 °C under constant shaking to precipitate the DNA. The precipitated DNA was pelleted by centrifuging at 10,000 r.p.m. for 10 min at 4 °C. The supernatant was removed and the precipitate washed with ice-cold 70% ethanol and centrifuged at 10,000 r.p.m. for 10 min at 4 °C. After removing the ethanol, the pellet was air-dried and resuspended in an appropriate volume of DNase-free water. The obtained DNA was used to perform qPCR with oligos of mitochondrial (MT-TL1, MT-ND2) and genomic DNA (ACTB, NCOA3). Results were demonstrated as a mitochondrial:genomic DNA ratio. The sequences of all primers used are given in Supplementary Table 7.

Analysis of mitochondrial oxidative phosphorylation. Mitochondrial respiration was monitored with the XFe-96 Cell Bionalyzer (Seahorse Biosciences). Optimal cell density and drug concentrations had been previously determined[11]. A standard MitoStress assay was performed. Briefly, 20,000 cells were plated in an XFe-96-well plate, and cells were kept for 6 h in DMEM–10% fetal bovine serum to allow the cells to attach. Then, the medium was changed to 10 mM glucose, 2 mM glutamine and 1 mM pyruvate XFe DMEM (5 mM HEPES, pH 7.4), and cells were incubated for 1 h at 37 °C without CO_2. Three different modulators of mitochondrial respiration were sequentially injected. After determination of the basal oxygen consumption rate, 1.5 μM oligomycin, which inhibits ATPase, was injected to determine the amount of oxygen dedicated to ATP production by mitochondria. To determine the maximal respiration rate or spare respiratory capacity, 1.5 μM carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone (FCCP) was injected to free the gradient of H^+ from the mitochondrial intermembrane space, and thus activate maximal respiration. Finally, 0.75 μM antimycin A and 0.75 μM rotenone were added to completely inhibit the mitochondrial respiration.

Statistical analysis and figure editing. In all bar plots, the height of the bar corresponds to the mean value and the bars indicate the s.d.. In all box plots, the box signifies the upper (75th) and lower (25th) quartiles, the median is represented by a horizontal line within the box and the mean by a rhombus within the box. The upper whisker extends from the upper hinge to the largest value no further than 1.5× the interquartile range (IQR) from the hinge (that is, the IQR is the distance between the first and third quartiles). The lower whisker extends from the hinge to the smallest value, at most 1.5× the IQR of the hinge. The statistical test and comparison used to calculate P values as well as the significance level are reported in each figure and/or figure legend. If not indicated otherwise, a two-tailed Student's t-test was used to assess statistical significance. The number of technical replicates or independent cell culture experiments is indicated in the relevant figure legend(s). Figures were edited using Inkscape (inkscape.org).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The reported protein structures are deposited in the Protein Data Bank with PDB accession nos. 7NY6 (unliganded *Capsaspora* macroH2A macrodomain) and 7NY7 (ADP-ribose bound *Capsaspora* macroH2A macrodomain). Source data are provided with this paper.

Code availability

We have exclusively used publicly available packages for bioinformatic analysis and provide their references in Methods. If not stated otherwise, we have used default parameters. Specific scripts are available on request.

References

53. Afgan, E. et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44**, W3–W10 (2016).
54. Brown, M. W. et al. Phylogenomics places orphan protistan lineages in a novel eukaryotic super-group. *Genome Biol. Evol.* **10**, 427–433 (2018).
55. de Mendoza, A., Suga, H., Permanyer, J., Irimia, M. & Ruiz-Trillo, I. Complex transcriptional regulation and independent evolution of fungal-like traits in a relative of animals. *eLife* **4**, e08904 (2015).
56. Dudin, O. et al. A unicellular relative of animals generates a layer of polarized cells by actomyosin-dependent cellularization. *eLife* **8**, e49801 (2019).
57. Grau-Bové, X. et al. Dynamics of genomic innovation in the unicellular ancestry of animals. *eLife* **6**, e26036 (2017).
58. Hehenberger, E. et al. Novel predators reshape holozoan phylogeny and reveal the presence of a two-component signaling system in the ancestor of animals. *Curr. Biol.* **27**, 2043–2050.e6 (2017).
59. Richter, D., Berney, C., Strassert, J., Burki, F. & de Vargas, C. EukProt: a database of genome-scale predicted proteins across the diversity of eukaryotic life. Preprint at *bioRxiv* https://doi.org/10.1101/2020.06.30.180687 (2020).
60. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
61. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
62. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188–1190 (2004).
63. Bawono, P. & Heringa, J. PRALINE: a versatile multiple sequence alignment toolkit. *Methods Mol. Biol.* **1079**, 245–262 (2014).
64. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
65. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320 (2008).
66. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**, 2971–2972 (2006).
67. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
68. Suga, H. et al. The Capsaspora genome reveals a complex unicellular prehistory of animals. *Nat. Commun.* **4**, 2325 (2013).
69. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
70. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
71. Huerta-Cepas, J. et al. EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
72. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
73. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).
74. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 236 (2019).
75. Ruiz, P. D. & Gamble, M. J. MacroH2A1 chromatin specification requires its docking domain and acetylation of H2B lysine 20. *Nat. Commun.* **9**, 5143 (2018).
76. Mayer, M. & Meyer, B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chem.* **38**, 1784–1788 (1999).
77. Aretz, J. et al. Allosteric inhibition of a mammalian lectin. *J. Am. Chem. Soc.* **140**, 14915–14925 (2018).
78. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
79. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
80. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
81. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
82. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. Electronic ligand builder and optimization workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074–1080 (2009).
83. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
84. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. Ligplot: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng. Des. Sel.* **8**, 127–134 (1995).
85. Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).
86. Sporn, J. C. et al. Histone macroH2A isoforms predict the risk of lung cancer recurrence. *Oncogene* **28**, 3423–3428 (2009).
87. Fackelmayer, F., Dahm, K., Renz, A., Ramsperger, U. & Richter, A. Nucleic-acid-binding properties of hnRNP-U/SAF-A, a nuclear-matrix protein which binds DNA and RNA in vivo and in vitro. *Eur. J. Biochem.* **221**, 749–757 (1994).

Author contributions

Competing interests

Additional information

Extended data Extended date are available for this paper at https://doi.org/10.1038/s41594-021-00692-5.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-021-00692-5.

Correspondence and requests for materials should be addressed to Marcus Buschbeck.

Peer review information *Nature Structural and Molecular Biology* thanks Matthew Gamble and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Beth Moorefield was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. Peer review reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

a

Protein

Compound of interest

Protein saturation

Transfer of saturation

1D NMR spectra of a compound

I_0

I_{sat}

Difference spectra (STD-NMR)

$I_0 - I_{sat}$

Ligand

Non-binder

b

mH2A

0.6% 3.0% 0.8%

mH2A1.1

1.0% 4.95% 1.1%

mH2A1.1^{G224E}

ADPR

8.5 8.0 7.5 7.0 6.5 6.0

[ppm]

c

mH2A + ADPR

K_d = 1,26 µM

d

mH2A1.1 + ADPR

K_d = 11,3 µM

e

ADPR

f Nucleotide analog binding to *Capsaspora* macroH2A macrodomain

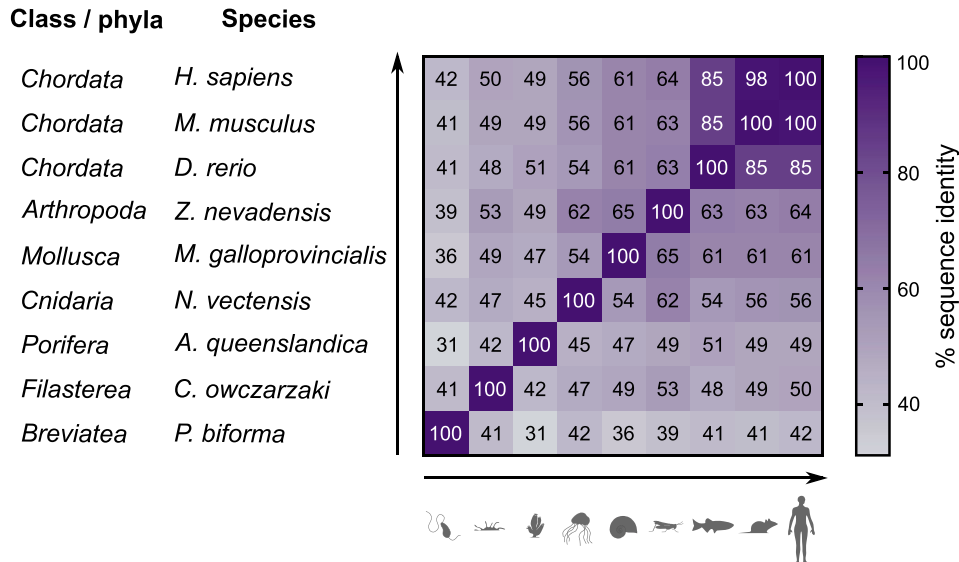| Ligand	K_d (µM)	N	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)
ADP-ribose	1,26 ± 0,19	1,01 ± 0,01	-5,32 ± 0,14	-2,73 ± 0,23
ADP	60,43 ± 35,05	1,10 ± 0,23	-2,94 ± 0,93	-2,89 ± 1,24
ATP	> 100	ND	ND	ND
AMP	> 100	ND	ND	ND
GDP	> 100	ND	ND	ND
Ribose	> 100	ND	ND	ND

Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | The *Capsaspora* macroH2A macrodomain specifically binds ADP-ribose (accompanying Fig. 2). a, Scheme of the saturation transfer difference nuclear magnetic resonance (STD-NMR) experiment, which involves recording two analogous 1D NMR spectra: the reference spectrum of a ligand without saturation of protein signals (irradiating at a position far enough from any signal in the ^{1}H spectrum to avoid saturation), and a spectrum recorded after selective saturation of the protein and bound molecules. The difference spectrum shows only the signals of the binders. **b**, STD-NMR experiments indicate the presence of interaction between *Capsaspora* macroH2A macrodomain (mH2A) and ADP-ribose (ADPR) comparable to the interaction with mouse macroH2A1.1 macrodomain (mH2A1.1) and ADPR. No interaction is observed between ADPR and the loss-of-function mutant mH2A1.1^{G224E} that we included as a control. STD spectrum for ADPR in the absence of protein is shown as a reference. The STD effect (%) is quantified as $I_{STD} = 100 * (I_0 - I_{sat}) / I_0$. **c, d, e**, Representative plots of raw heat evolution in isothermal titration calorimetry (ITC) experiments. ADPR was titrated into a solution with either purified *Capsaspora* mH2A (**c**), mouse mH2A1.1 (**d**) macrodomain, or into the buffer alone (**e**) as a control. The integration of the raw heat evolution results in a curve representing the equilibrium binding isotherm, represented below the raw heat evolution plots (**c, d**). **f**, Table showing the results of ITC binding affinity measurements (K_d, N, ΔH and $-T\Delta S$) of *Capsaspora* mH2A macrodomain and a range of nucleotides (ATP, ADP, AMP, GDP) and ribose. K_d values for ATP, AMP, GDP and ribose are estimated to be higher than 100 μM, as their ITC binding curves show weak or absence of binding. If no binding occurred or binding was below ITC detection limit, reaction stoichiometries and thermodynamics could not be determined (ND).

a *Capsaspora* apo-structure topology diagram
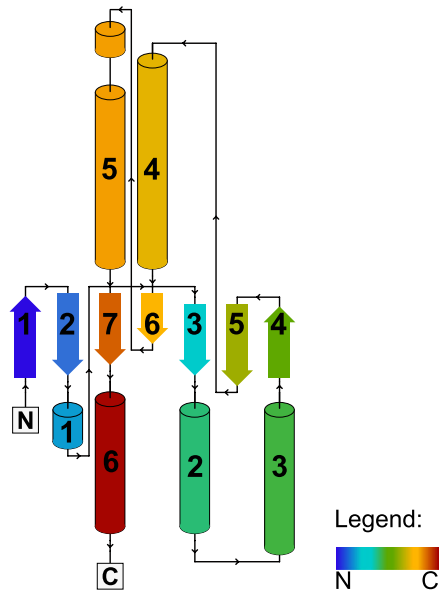
b *Capsaspora* ADPR-bound topology diagram
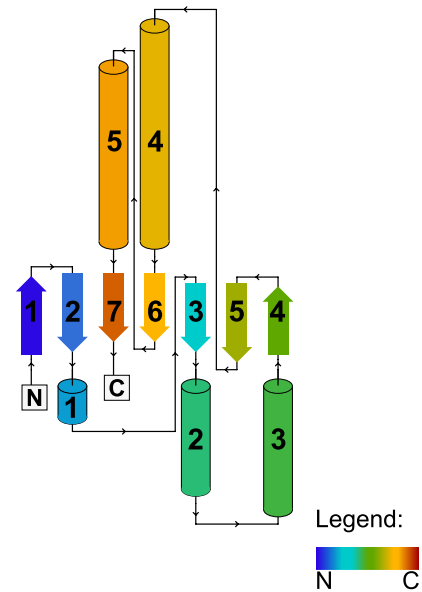
c Ligplot interaction network
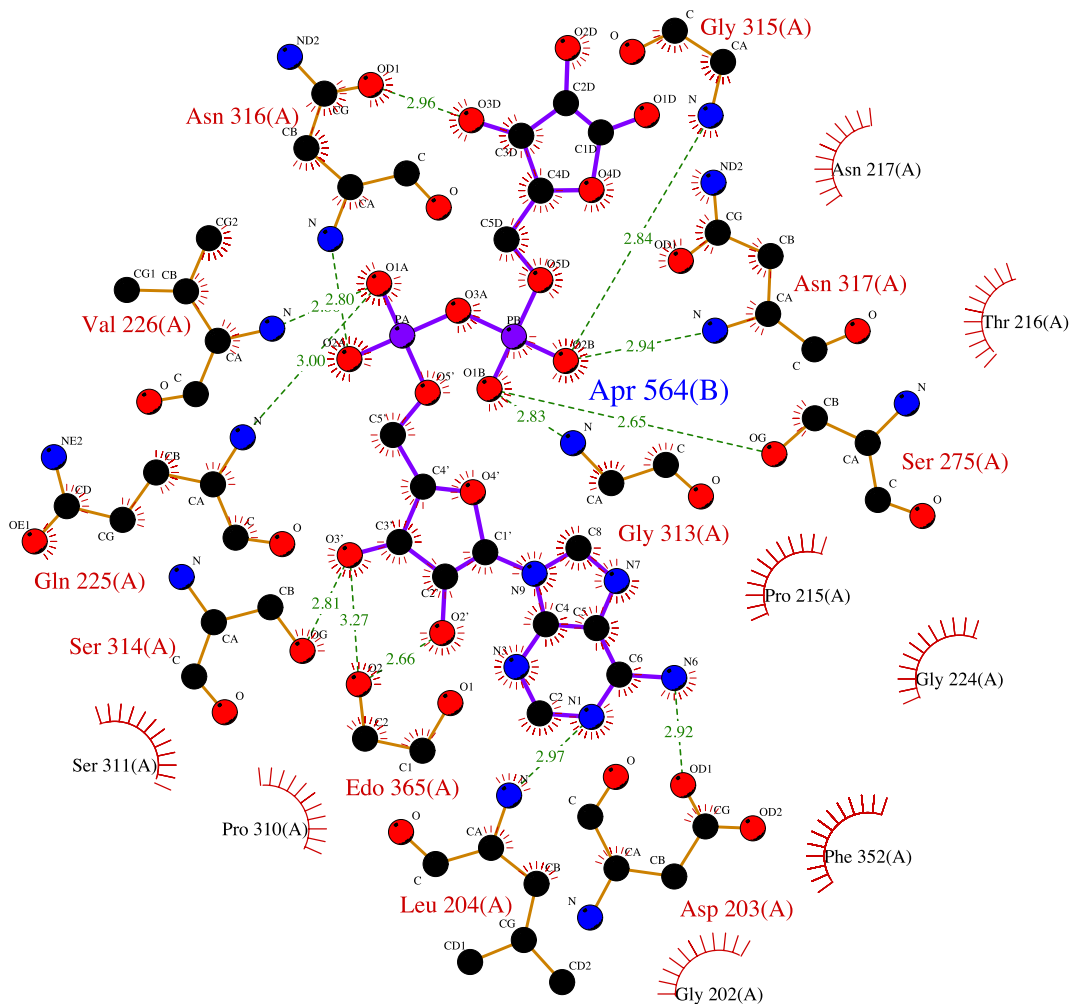
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | The apo and ADPR-bound-structure of the *Capsaspora* macrodomain (accompanying Fig. 2). a, Topology diagram of apo-form of *Capsaspora* mH2A macrodomain. The topology diagram shows β-sheets in a conserved 1276354 order, surrounded by 6 α-helices. **b**, Topology diagram of *Capsaspora* mH2A macrodomain in complex with ADP-ribose (ADPR). The topology diagram shows β-sheets in a conserved 1276354 order, surrounded by 5 α-helices. The C-terminal α-helix observed in the apo-structure (**a**) was not detected in the ADPR-bound structure. This prevented the modelling of the C-terminal 13 amino acids, while the apo structure could be fully refined all the way to the C-terminus of the macrodomain. Whether this is the result of a conformational change upon ligand binding or an artefact from the crystal packing remains to be determined. However, we previously observed that ADP-ribose binding induces a conformational change in the human macroH2A1.1 macrodomain, specifically by a 30° rotation of the C-terminal α-helix away from the globular macrodomain fold (Timinszky et al., 2009, NSMB). **c**, The schematic representation of ADPR interaction with *Capsaspora* macroH2A macrodomain. All non-covalent interactions between ADPR and the macrodomain are summarized in the LigPlot diagram. ADPR ligand is represented in thick purple line and the amino acid residues of *Capsaspora* macroH2A macrodomain in thin orange lines. Hydrogen bonds are represented by the dashed lines between atoms involved, while the circles or semicircles with radiating lines represent atoms or residues involved in hydrophobic contacts between protein and ligand.

a

Class / phyla	Species										
Chordata	H. sapiens	42	50	49	56	61	64	85	98	100	
Chordata	M. musculus	41	49	49	56	61	63	85	100	100	
Chordata	D. rerio	41	48	51	54	61	63	100	85	85	
Arthropoda	Z. nevadensis	39	53	49	62	65	100	63	63	64	
Mollusca	M. galloprovincialis	36	49	47	54	100	65	61	61	61	
Cnidaria	N. vectensis	42	47	45	100	54	62	54	56	56	
Porifera	A. queenslandica	31	42	100	45	47	49	51	49	49	
Filasterea	C. owczarzaki	41	100	42	47	49	53	48	49	50	
Breviatea	P. biforma	100	41	31	42	36	39	41	41	42	

% sequence identity (100, 80, 60, 40)

b Number and type of bonds established between macrodomains and ADP-ribose

	Capsaspora macroH2A	Human macroH2A1.1
Hydrogen bonds	11	17
Other bonds	8	6

List of residues establishing other bonds than hydrogen bonds

Capsaspora macroH2A	Human macroH2A1.1
Gly202	Pro215
Pro215	Thr216
Thr216	Asn217
Asn217	Ser310
Gly224	Ile311
Pro310	Phe351 (pi-stacking)
Ser311	
Phe352 (pi-stacking)	

c

N316

Q225

y = 90º

y = 180º

y = 180º

y = 90º
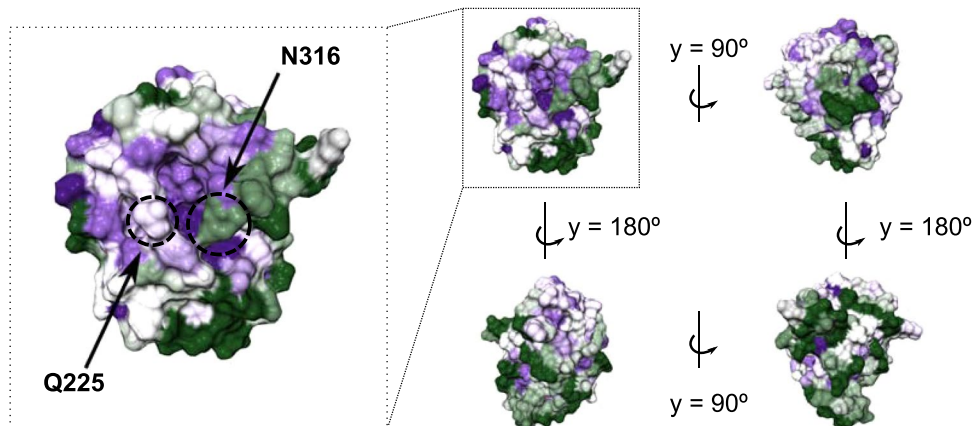
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Conservation of the macrodomain and bonds established with ADP-ribose (accompanying Fig. 3). a, Pairwise sequence identity (%) of macrodomains from representative macroH2A sequences of 9 different phyla. For the vertebrates, amino acid sequences corresponding to macroH2A1.1 have been used. **b**, Table indicating the amino acids establishing H-bond or other bonds with ADP-ribose for *Capsaspora* macroH2A (see also Extended Data Fig. 2c) and human macroH2A1.1 (Timinszky et al., 2009). **c**, Snapshots of the apo structure of *Capsaspora* macroH2A macrodomain starting from the front perspective of the binding pocket and subsequently rotated by 90°, 180° and 270° around y-axis. The level of conservation is calculated by Consurf server and is projected on the protein surface. See also the movie provided as Supplementary Data File S5. Please note that in the movie conservation degree is color-coded from low (turquoise), intermediate (white) to high (magenta).
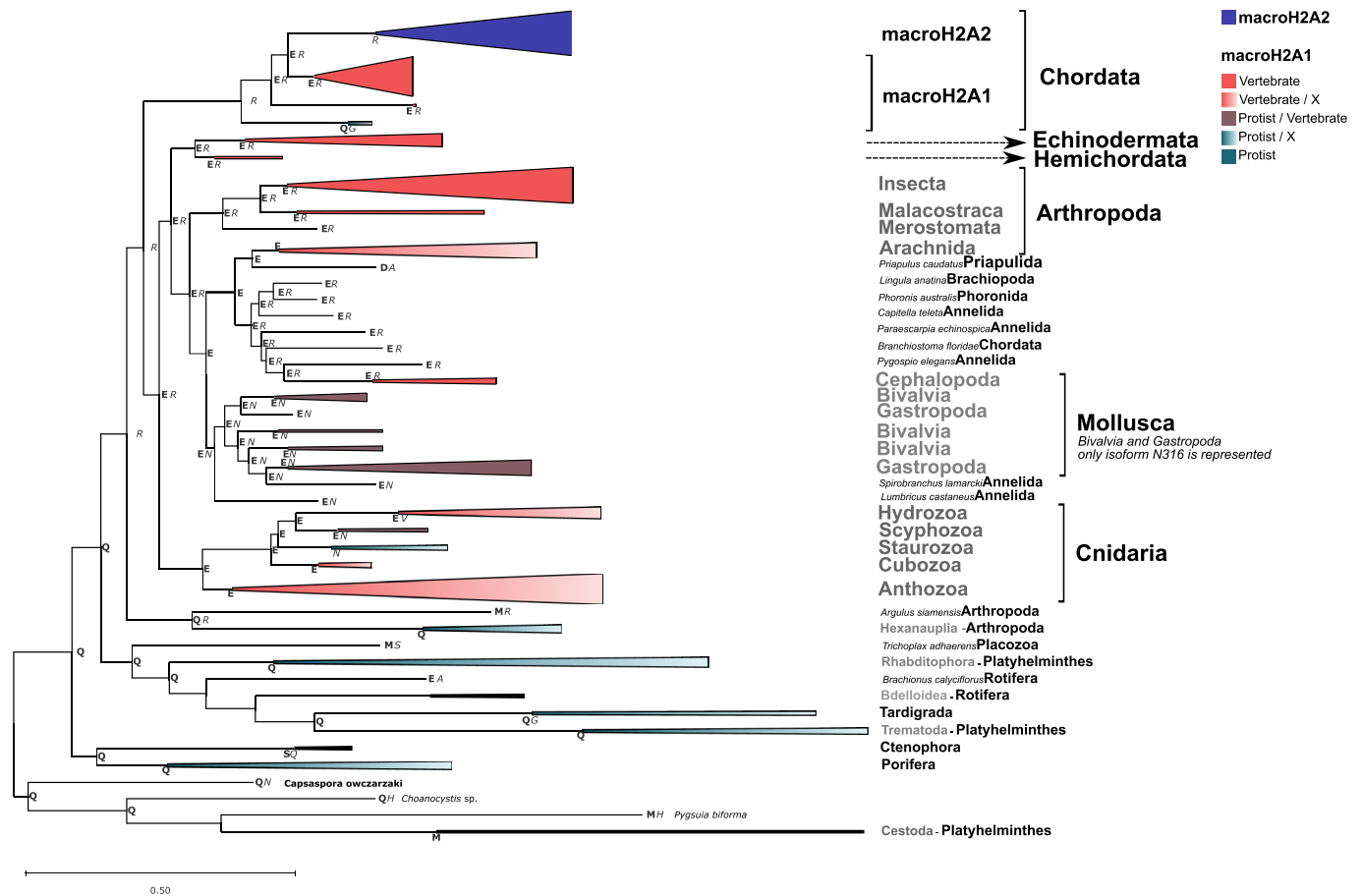
Extended Data Fig. 4 | Inferred ancestral amino acid states at sites 225 and 316 (accompanying Fig. 4). Simplified version of the maximum likelihood tree shown in Supplementary Data File S3. The tree shows a set of possible amino acid (states) at each ancestral node based on their inferred likelihood at sites 225 (in bold) and 316 (in italics). Protein numeration is based on the full-length *Capsaspora* macroH2A. Probability cutoff is 0.95 and the ambiguous states are not shown. Collapsed branches in choral represent the vertebrate-characteristic pair E225-R316 (diffused when only one was inferred), in turquoise the *Capsaspora*-characteristic pair Q225-N316 (diffused when only one was inferred), and in brown the species with E225 (vertebrate-like) and N316 (*Capsaspora*-like) residues.

a Identity of other residues at positions 225 and 316

n (species with other at 225) = 57   n (species with other at 316) = 108

d Thermal shift assay

Residues: positive | negative | polar | non-polar
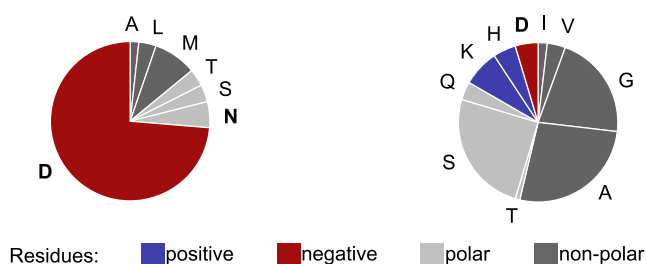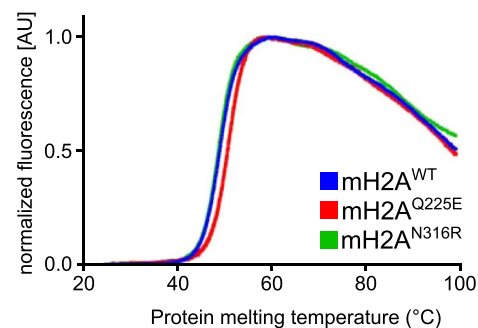
b Residue pair identity at the position 316

Residue 225: Q225, 32 species   Residue 225: E225, 216 species
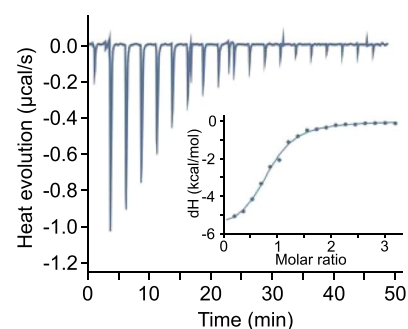
Residues: positive | negative | polar | non-polar

c Residue pair identity at the position 225

Residue 316: N316, 41 species   Residue 316: R316, 156 species

Residues: positive | negative | polar | non-polar

e mH2A^{Q225E} + ADPR

mH2A^{N316R} + ADPR
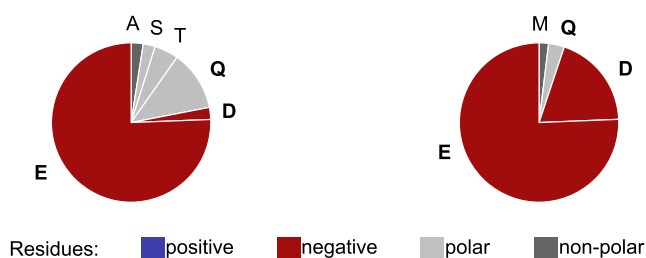
f Mouse and *Capsaspora* wild type and mutant macrodomains affinity to ADPR

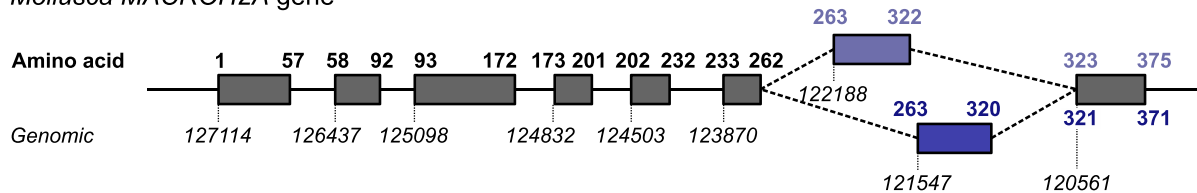| macroH2A_MD	K_d (μM)	ΔG (kcal/mol)	ΔH (kcal/mol)	-TΔS (kcal/mol)
*Co*_WT	1,26 ± 0,19	-8,05 ± 0,09	-5,32 ± 0,14	-2,73± 0,23
*Co*_Q225E	8,64 ± 0,79	-6,91± 0,06	-6,12 ± 0,16	-0,79± 0,21
*Co*_N316R	> 100	ND	ND	ND
*Mm*_mH2A1.1	11,30 ± 3,04	-6,77 ± 0,19	-10,58 ± 1,64	3,82 ± 1,84

Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | The identity and biophysical consequences of evolution of residues 225 and 316 closing the binding pocket (accompanying Fig. 4). a, The pie charts showing the identity of other residues at positions 225 and 316. Residue colors are based on their biochemical properties: positively charged in blue, negatively charged in red, polar in gray, and non-polar in black. **b**, The pie charts showing the identity of the residues at the position 316 which pair with either protist-characteristic Q225, or vertebrate-characteristic E225. The numbering of the residues is based on the *Capsaspora* mH2A. In particular, the protist-characteristic residue Q225 was paired with varying residues at 316. **c**, The pie charts showing the identity of the residues at the position 225 which pair with either protist-characteristic N316, or vertebrate-characteristic R316. The numbering of the residues is based on the *Capsaspora* mH2A. Strikingly, R316 was paired with E225 or the biochemically similar D225 in more than 95% of the species, while its pairing with Q225 was rare. **d**, Thermal shift assay profile of melting temperatures of wild type, Q225E and N316R mutants of *Capsaspora* macroH2A macrodomains (mH2A^{WT}, mH2A^{Q225E} and mH2A^{N316R}, respectively) show that the folding of the mutant proteins is unaffected. The subtle shift in the melting temperature of Q225E mutant macrodomain to higher temperatures might be indicative of an increase in stability. **e**, Representative plots of raw heat evolution in isothermal titration calorimetry (ITC) experiments. ADP-ribose was titrated into a solution with either purified *Capsaspora* Q225E (left panel) or N316R (right panel) mutant macroH2A macrodomain. The integration of the raw heat evolution resulted in a curve representing the equilibrium binding isotherm, represented in the inset. **f**, Table showing the thermodynamics of the binding interaction calculated on the basis of ITC data shown in Fig. 4b.
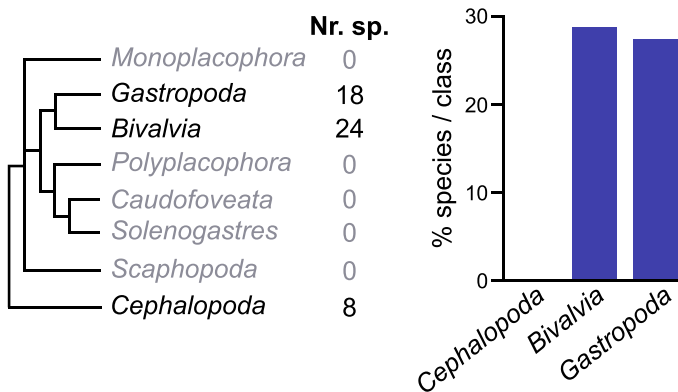
a Vertebrate *MACROH2A1* gene
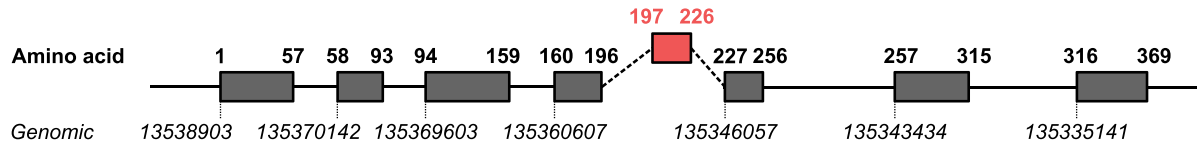


b *Mollusca* classes with alternative exon 7

**c** Focused analysis of the residue identity at position 316 in *Mollusca* with alternative splicing
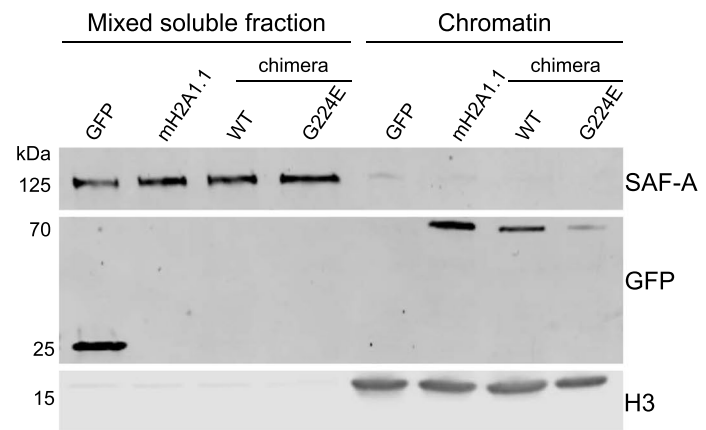

Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Alternative splicing of exon 7 in molluscs (accompanying Fig. 4). a, The organization of the human *MACROH2A1* gene and the corresponding gene from the mollusc *Crassostrea gigas* (Gene ID: 105347164) are represented on an arbitrary scale for comparison purposes. For simplification only the alternative exon 5 defining macroH2A1.1 is represented for human *MACROH2A1*. Both alternative exons 7 of the *Mollusca* macroH2A gene are represented. **b**, Transcriptomic data was available for 50 *Mollusca* species from three of eight different classes. The number of species per class is indicated (left panel). The percentage of species per *Mollusca* class containing alternatively spliced macroH2A isoforms are shown (right panel). **c**, Alignment of C-terminal alternative exons from the mollusc species in which we detected the 2 alternative splice isoforms. Almost all species have both R316 and N316-containing isoforms.
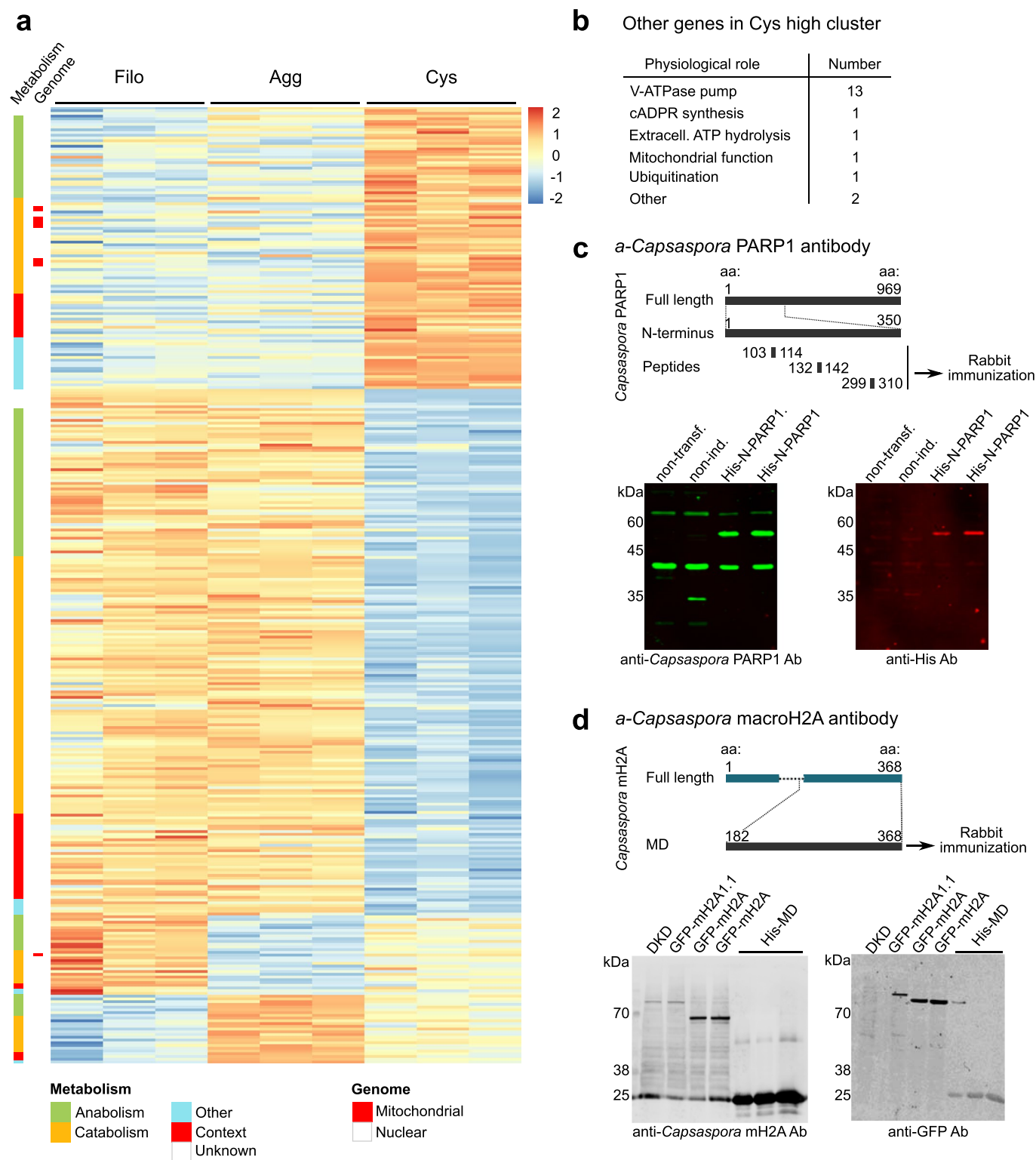
Extended Data Fig. 7 | Metabolic dynamics of *Capsaspora* life stages (accompanying Fig. 5). a, Heatmap of 349 metabolic genes differentially expressed across the three life stages of *Capsaspora*. **b**, Table showing the 'other' genes in the Cys high cluster (Fig. 5c) grouped according to their physiological role. **c**, Antibody specific for *Capsaspora* PARP1. Rabbits were immunized with three different peptides corresponding to the N-terminal part of PARP1. The antibody specificity was corroborated by immunoblotting bacterial lysates: non-transformed (non-transf.), transformed with a plasmid encoding for N-terminal tagged fragment of PARP1 corresponding to amino acids 1-350 but not induced (non-ind.), and bacteria producing His-tagged PARP1 (His-N-PARP1). Representative immunoblot is shown (n = 3). **d**, Antibody specific for *Capsaspora* macroH2A. Rabbits were immunized with purified His-tagged macroH2A macrodomain (AA: 182-368). The specificity was corroborated by immunoblotting using stable cell lines expressing GFP-tagged constructs of mouse macroH2A1.1 (mH2A1.1) and *Capsaspora* macroH2A (mH2A) and purified His-tagged macrodomain of *Capsaspora* mH2A. Representative immunoblot is shown (n = 3).

a GFP expression in established stable cell lines
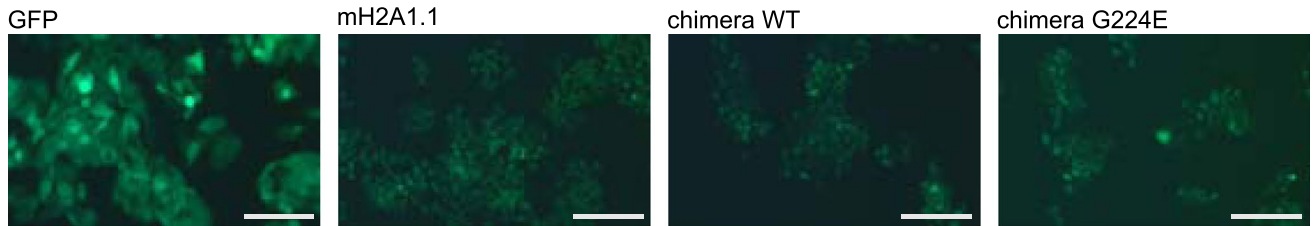
b Quantification of GFP expression in established stable cell lines
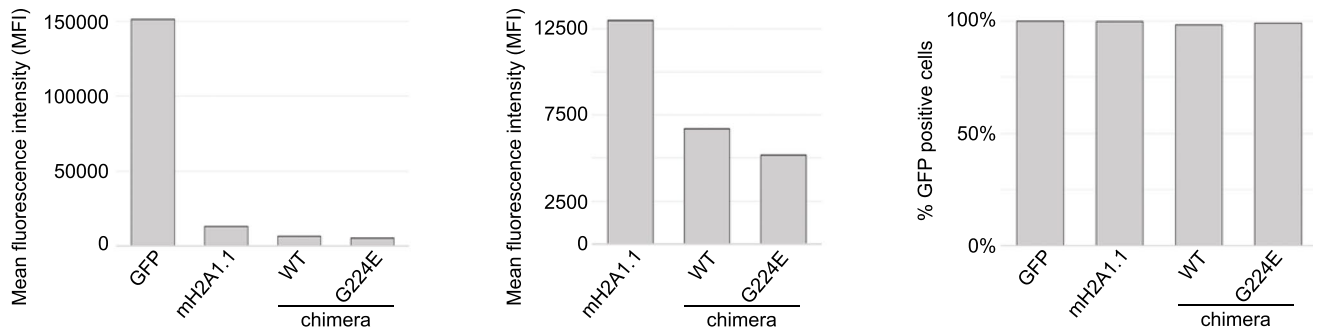
c
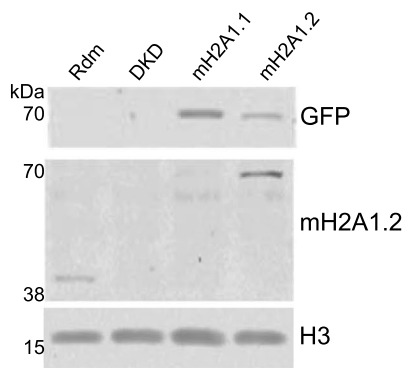
d

e
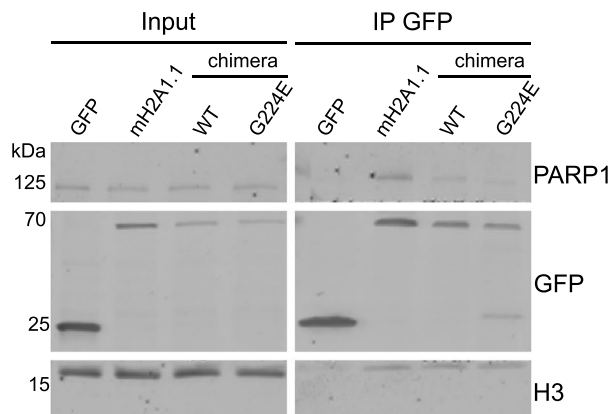
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Characterization of stable DKD cell lines (accompanying Fig. 6). a, Fluorescence intensity of stable DKD cell lines stably expression either GFP or GFP-tagged constructs of mouse macroH2A1.1 (mH2A1.1) or chimeric wild type or G224E proteins containing the *Capsaspora* macrodomain as illustrated in Fig. 6c. Scale represents 50 µm. Cells were checked daily for GFP expression, and the representative images are shown. **b**, Flow cytometric detection of GFP in the same cells lines as **A**. Bar plot to the left compares mean fluorescence intensities of all four stable cell lines. Bar plot in the center shows GFP fluorescence in cell lines stably expressing macroH2A constructs. The bar plot to the right shows high percentage ($\geq$ 99%) of GFP positive cells. **c**, Stable cell lines expressing the GFP-tagged mouse mH2A1.1 were compared to a similar cell line expressing GFP-tagged mouse macroH2A1.2 (mH2A1.2), control DKD cells and the parental HepG2 cell line. Immunoblotting using anti-GFP and anti-mH2A1.2 specific antibodies allows evaluating the expression level of the exogenous mH2A1.1 relative to endogenous mH2A1.2. Comparison to Extended Data Fig. 7b indicates that the chimeric proteins are expressed at a similar range or slightly higher than endogenous macroH2A proteins. Representative immunoblot is shown (n = 3). **d**, Immunoblot analysis after cell fractionation shows that all macroH2A constructs but not GFP alone are incorporated in chromatin. Representative immunoblot is shown (n = 3). **e**, The chimeric *Capsaspora*-mouse macroH2A construct interacts with PARP1 as detected by immunoblotting after co-immunoprecipitation. Representative immunoblot is shown (n = 4).

Corresponding author(s): Marcus Buschbeck

Last updated by author(s): October 21, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	A local DNA sequence database was created from GenBank database by Blast search using human sequences as a query.
Data analysis	The following software were used in the study for data analysis: 1. Trinity software (version 2.2.0) for de novo transcriptome assembly, Galaxy web platform (https://usegalaxy.org/) 2. FastQC (2020) for quality control (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) 3. MAFFT (version 7) for MSA (https://mafft.cbrc.jp/alignment/server/) 4. Jalview (version 2) for MSA (https://www.jalview.org/) 5. BioEdit (version 7.2) for potential error editing (https://bioedit.software.informer.com/7.2/) 6. WebLogo3 (version 3) for logo plot generation (http://weblogo.threeplusone.com/) 7. Morpheus (version 07.01.2021) for homology matrix visualization (https://software.broadinstitute.org/morpheus/) 8. MEGA X (version 10.1.7) for phylogenetic and molecular evolutionary analyses (https://www.megasoftware.net/) 9. TimeTree database (version 07.01.2021) for estimating divergence time between pairs of taxa (http://www.timetree.org/) 10. DnaSP (version 6) for calculating the amount of codon usage bias (http://www.ub.edu/dnasp/) 11. featureCounts software (v.2.0.1) for quantification of RNAseq reads (http://subread.sourceforge.net/) 12. DESeq2 package (version 1.32.00) for differential gene expression analysis (https://bioconductor.org/packages/release/bioc/html/DESeq2.html) 13. DEGreport package (version 1.28.0) for data clusterization plot (bioconductor.org/packages/release/bioc/html/DEGreport.html) 14. EggNOG (version 5.0) for functional annotation of genes (http://eggnog5.embl.de/#/app/home) 15. GHOSTKoala (version as of 01.03.2021) for functional annotation of genes based on K number (https://www.kegg.jp/ghostkoala/) 16. Orthofinder (version 2.5.4) for identification of gene orthologs (https://github.com/davidemms/OrthoFinder) 17. MicroCal PEAQ-ITC Analysis software (provided by Malvern Panalytical 2018) was used for ITC data analysis. 18. COOT (version 0.9.6), phenix.refine and phenix.elbow (versions as of 01.06.2019) were used for resolving the protein crystal structures (https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/; https://phenix-online.org/documentation/reference/refinement.html; https://

phenix-online.org/documentation/reference/elbow.html)
19. Consurf tool (version as of 01.03.2021) for projection of evolutionary conservation scores (https://consurf.tau.ac.il/)
20. UCSF Chimera (01.06.2019) and PyMOL (version 2.4) for protein visualization (https://www.cgl.ucsf.edu/chimera/ , https://pymol.org/2/)
21. ImageStudioLite (version 5.x) for the analysis of immunoblots (https://www.licor.com/bio/image-studio-lite/)
22. Inkscape vector graphics on XQuartz (version 2.7.11) editor for creating and editing figures.
23. LIGPLOT (version 4.5.3) for interaction and bond analysis.
24. PRALINE (version as of 01.03.2019) for multiple sequence alignment.
25. STAR (version 2.7.3a) realignment of RNA sequencing reads and inclusion of the transcriptome data and featureCounts software (version 2.0.1) for quantification.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files. A data availability statement has been added to the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We did not perform sample size calculation prior to experimentation but have chosen similar sample sizes to those published in similar experimental studies elsewhere. Details on used samples sizes are given in the Methods section and Figure legends.
Data exclusions	No data was excluded.
Replication	All experiments were replicated in several replicates (n stated in the figure legend of each figure).
Randomization	The design of our study did not require any formal sample randomization.
Blinding	Our study includes biophyiscal types of analysis that does not allow for user-induced biases. As such we decided to omit blinding. Some experimnets were delegated to technicians that were not aware of sample identity.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used
- anti-Capsaspora macroH2A and anti-Capsaspora PARP1 antibodies were generated in this study as described in the methods section and validated by WB as shown in Extended Data Figure 7.
- anti-histone H3 (AbCam; ab-1791) was validated by ChIP, IP and WB accodring to informatio provided by the manufacturer.
- anti-PAR (Trevigen, 4336-APC-050) has been validated for ELISA and WB by the manufacturer.
- anti-GFP (Santa Cruz Biotechnology; sc-9996) has been validatd by Western blotting and other methods according to the provider's information.
- anti-macroH2A1.1 (Sporn et al., 2009) was validated by Western blot in the cited paper and by us.
- anti-macroH2A1.2 (Cell Signaling; 4827S) was validated by Western blot by the manufacturer and according to our experience has a weak but tolerable cross-reactivity with macroH2A1.1.
- anti-His-tag (ImmunoPrecise antibodies; IPA-2C6.1) has been validated for Western blotting, immunofluorescence and other methods by the manufacturer.
- anti-SAF-A has been provided Frank Fackelmayer and has been validated by Western blotting (Kipp et al., 2000, MCB).
- anti-PARP-1 (Abcam, Ab6079; Trevigen, 4338-MC-50) have been validated by Western blotting by both providers.
- fluorophore-conjugated secondary anti-mouse and anti-rabbit (LI-COR Biosciences IRDye 680RD and IRDye 800CW) have been validated by Western blotting and other methods by the manufacturer.

Validation
The validation of the antibodies generated in this study was done by western blot as shown in the supplementary figure S6C and S6D.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)
HepG2 and HEK293T cell lines was obtained from ATCC.

Authentication
Cells were authenticated by STR profiling.

Mycoplasma contamination
Cells are mycoplasma negative. The mycoplasma test was performed by PCR every 2 weeks.

Commonly misidentified lines
(See ICLAC register)
Used cell lines are not registered with ICLAC.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals
The protist capsaspora owkzarzaki has been grown as described in literature cited in materials and methods.

Wild animals
N.A

Field-collected samples
NA

Ethics oversight
NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.