

# A paralog of Pcc1 is the fifth core subunit of the KEOPS tRNA-modifying complex in Archaea

Received: 5 September 2022

Accepted: 20 January 2023

Published online: 01 February 2023

 Check for updates

Marie-Claire Daugeron<sup>1,5</sup>, Sophia Missouri<sup>1,3,5</sup>, Violette Da Cunha<sup>1,4</sup>,  
Noureddine Lazar<sup>1</sup>, Bruno Collinet<sup>1,2</sup>, Herman van Tilbeurgh<sup>1</sup>✉ &  
Tamara Basta<sup>1</sup>✉

In Archaea and Eukaryotes, the synthesis of a universal tRNA modification, N<sup>6</sup>-threonyl-carbamoyl adenosine (t<sup>6</sup>A), is catalyzed by the KEOPS complex composed of Kae1, Bud32, Cgi121, and Pcc1. A fifth subunit, Gon7, is found only in Fungi and Metazoa. Here, we identify and characterize a fifth KEOPS subunit in Archaea. This protein, dubbed Pcc2, is a paralog of Pcc1 and is widely conserved in Archaea. Pcc1 and Pcc2 form a heterodimer in solution, and show modest sequence conservation but very high structural similarity. The five-subunit archaeal KEOPS does not form dimers but retains robust tRNA binding and t<sup>6</sup>A synthetic activity. Pcc2 can substitute for Pcc1 but the resulting KEOPS complex is inactive, suggesting a distinct function for the two paralogs. Comparative sequence and structure analyses point to a possible evolutionary link between archaeal Pcc2 and eukaryotic Gon7. Our work indicates that Pcc2 regulates the oligomeric state of the KEOPS complex, a feature that seems to be conserved from Archaea to Eukaryotes.

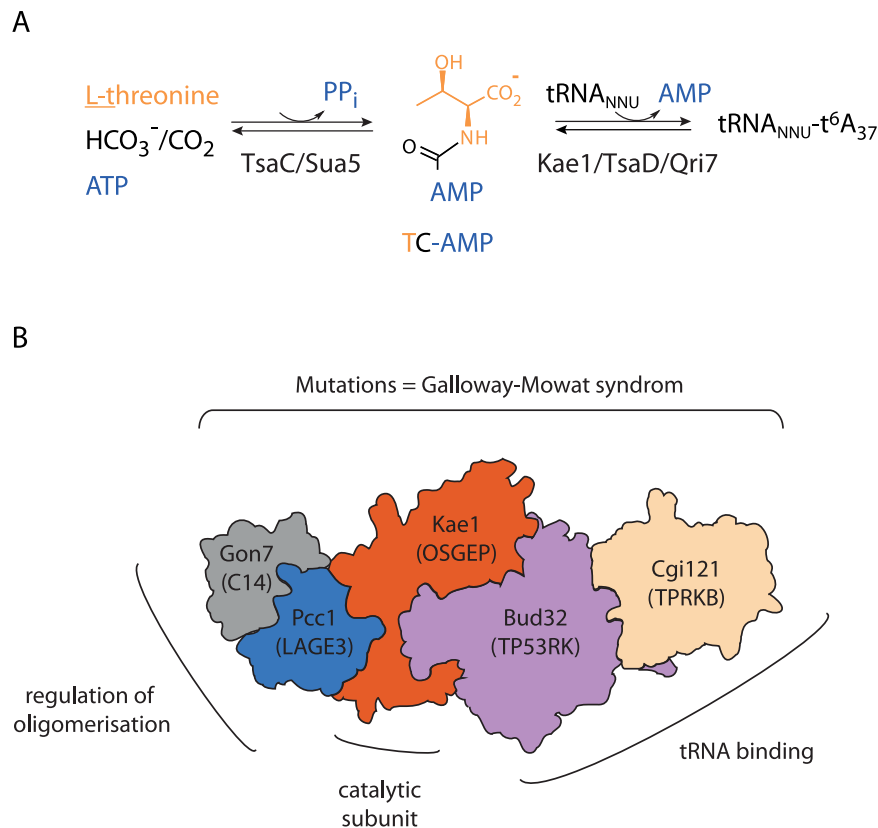
The genes that are shared by all cellular life forms are coding for a very restricted set of about 60 proteins, primarily involved in translation<sup>1,2</sup>. Two protein families in this set, Sua5/TsaC (COG0009) and Kae1/TsaD/Qri7 (COG0533), are tRNA-modifying enzymes involved in the synthesis of a noncanonical nucleotide N<sup>6</sup>-threonyl-carbamoyl adenosine (t<sup>6</sup>A)<sup>3–5</sup>. This modified adenosine derivative is always located at position 37, adjacent to the anticodon, in almost every tRNA reading A-starting codons (ANN, N being one of the four canonical nucleotides A, C, G, or U)<sup>6–8</sup>. t<sup>6</sup>A is critical for accurate decoding of mRNA<sup>5,9–13</sup> and for binding of tRNA to the ribosome and translocation to the P site during protein synthesis<sup>14–16</sup>. Mutations in the t<sup>6</sup>A biosynthetic genes can be lethal in bacteria and archaea<sup>5,17</sup> or result in a wide range of severe phenotypes in eukaryotes such as translation defects<sup>18</sup> mitochondrial dysfunction<sup>19</sup> genomic instability<sup>20,21</sup>, transcription defects<sup>22,23</sup>, telomere shortening<sup>24,25</sup> and body development defects<sup>26</sup>.

In humans, such mutations were linked to a rare genetic disease, Galloway-Mowat syndrome (GAMOS), whereby neurological and renal functions are severely affected and result most often with childhood mortality<sup>27,28</sup>.

The biosynthesis of t<sup>6</sup>A is a two-step process in which Sua5/TsaC proteins first catalyze the formation of threonyl-carbamoyl adenylate (TC-AMP)<sup>29</sup>. This unstable intermediate is used as substrate by the Kae1/TsaD/Qri7 enzyme family which catalyzes the transfer of a threonyl-carbamoyl group to the A<sub>37</sub> of the tRNA substrate<sup>30,31</sup> (Fig. 1A). While the mitochondrial ortholog Qri7 is a standalone enzyme acting as a homodimer, in Bacteria the TsaD protein associates with two bacteria-specific proteins, TsaB (an inactive paralog of TsaD) and TsaE (a P-loop ATPase) into DEZ complex<sup>32</sup>. In archaea and eukaryotes, the Kae1 enzymes are part of a multiprotein complex named KEOPS (Kinase Endopeptidase and Other Proteins of Small size)<sup>24</sup> or EKC

<sup>1</sup>Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198 Gif-sur-Yvette, France. <sup>2</sup>Institut de Minéralogie de Physique des Matériaux et de Cosmochimie (IMPMC), Sorbonne-Université, UMR7590 CNRS, MNHN, Paris, France. <sup>3</sup>Present address: Department of structural biology and chemistry, Institut Pasteur, Paris, France. <sup>4</sup>Present address: Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France. <sup>5</sup>These authors contributed equally: Marie-Claire Daugeron, Sophia Missouri.

✉ e-mail: [herman.van-tilbeurgh@i2bc.paris-saclay.fr](mailto:herman.van-tilbeurgh@i2bc.paris-saclay.fr); [tamara.basta@i2bc.paris-saclay.fr](mailto:tamara.basta@i2bc.paris-saclay.fr)



**Fig. 1 | KEOPS complex structure and function in t<sup>6</sup>A biosynthesis.** **A** Enzymatic synthesis of t<sup>6</sup>A. Sua5/TsaC protein family catalyzes the condensation of L-threonine, bicarbonate or CO<sub>2</sub> and one ATP molecule to form an unstable intermediate threonylcarbamoyl-adenylate (TC-AMP). Kae1/TsaD/Qri7 family catalyzes the transfer of the threonylcarbamoyl moiety of TC-AMP to adenosine 37 within the anticodon loop to form t<sup>6</sup>A modified tRNA. **B** Schematic representation of the KEOPS structure and functional assignment of subunits. The complex is a linear assembly of five subunits Cgi121/Bud32/Kae1/Pcc1/Gon7. The names of the human orthologs are given in the brackets. Gon7/C14 subunit has been found only in fungi and animals. The shape of each subunit roughly recapitulates the corresponding

crystal structures in Gon7-Pcc1 (PDB 4WXA), Pcc1-Kae1 (PDB 3EN0), Kae1-Bud32 (PDB 3EN9), and Bud32-Cgi121 (PDB 4WWA) binary complexes. Genetic and in vitro functional assays established that Bud32 and Cgi121 function in tRNA recognition and binding. Pcc1 forms homodimers and provides the interface for oligomerization of the four subunit (4SU) KEOPS complex into a dimer of two heterotetramers. In fungi and animals, the Gon7/C14 subunit binds to Pcc1 through its dimerization interface and blocks the assembly of the KEOPS dimer. Mutations in all five genes encoding human KEOPS were linked with a severe genetic disease called Galloway-Mowat syndrome.

(Endopeptidase-like Kinase Chromatin-associated)<sup>22</sup>. For simplicity, we will use the name KEOPS thereafter. For a detailed account on KEOPS structure and function the reader is referred to a recent review<sup>33</sup>. KEOPS was initially isolated from yeast by tandem affinity purification, and it contained Kae1, a metalloprotein from the ASHKA superfamily, Bud32 a Rio-like protein kinase and three small proteins Pcc1, Cgi121 and Gon7 of unknown function. The latter protein is specific for fungi, while Bud32, Pcc1 and Cgi121 are conserved both in archaea and eukaryotes<sup>22,24</sup>. The complete structure of the KEOPS complex could be modelled based on crystal structures of subcomplexes<sup>34–36</sup>. The complex exhibits a linear arrangement of subunits in the order Pcc1-Kae1-Bud32-Cgi121. The fifth subunit, Gon7, binds to Pcc1 on the opposite side from Kae1 in human and yeast KEOPS<sup>35,37</sup> (Fig. 1B).

Despite considerable progress made in the last decade, how the auxiliary proteins Pcc1, Bud32, Cgi121, and Gon7 emerged and why they were recruited within the KEOPS complex in Archaea and Eukaryotes remains poorly understood. Using in vitro reconstituted KEOPS complex from the archaeon *Pyrococcus abyssi*, we initially showed that PaKae1, PaPcc1, and PaBud32 are necessary and sufficient for the synthesis of t<sup>6</sup>A while the addition of PaCgi121 stimulated the activity<sup>38</sup>. Using EMSA experiments, we showed that the binary Pcc1-Kae1 complex did bind to tRNA while Cgi121 alone did not. Beenstock and colleagues recently showed that Cgi121 recruits tRNA to KEOPS by binding specifically to its 3' CCA tail. Modelling of the tRNA/KEOPS complex based on low-resolution electron microscopy and biochemical data,

further indicated that the four KEOPS subunits form an extended tRNA-binding surface and are required for correct positioning of the A<sub>37</sub> nucleoside in the catalytic site of Kae1<sup>39</sup>.

Pcc1 is a small, ~10 kDa globular protein composed of two anti-parallel alpha helices and three-stranded beta-sheet. It has some structural resemblance with the K homology (KH) domain, which functions as a single-stranded DNA and RNA binder<sup>34</sup>. The X-ray crystal structure of Pcc1 from *Pyrococcus furiosus* showed that this protein forms homodimers whereby the two protomers are arranged in anti-parallel fashion and form a continuous 6-stranded beta-sheet with two alpha helices packed against one side of the sheet<sup>34</sup>. The Pcc1 homodimer provides two equivalent and non-overlapping surfaces for binding Kae1 and is responsible for the dimerization of the archaeal and human KEOPS complexes in vitro with 2:2:2:2 binding stoichiometry<sup>37,38</sup>. Perturbation of Pcc1-mediated dimerization of archaeal KEOPS by mutation of one Kae1-Pcc1 interface had no effect on t<sup>6</sup>A biosynthesis activity in vitro<sup>40</sup>. However, genetic experiments established that the synthetic Pcc1-Pcc1 fusion protein engineered to form homodimers in cis requires two intact Kae1 binding surfaces to support the growth of yeast thus raising the possibility that KEOPS could act as a dimer in vivo<sup>34</sup>. In fungi, Pcc1 interacts with Gon7 which is a small ~14 kDa intrinsically disordered protein that becomes partially structured upon interaction<sup>35</sup>. The specific function of Gon7 is unknown but this protein was shown to be essential for the t<sup>6</sup>A synthesis and telomere maintenance in yeast<sup>34,41</sup>. Gon7 binds to Pcc1 at

the opposite side of Kae1. It has a different fold from Pcc1 and its binding to Pcc1 precludes the dimerization of the yeast KEOPS<sup>35</sup>.

Sequence similarity searches failed to detect Gon7 homologs outside fungi but using proteomics approaches the C14ORF142 protein (hereafter C14) was identified as the fifth core subunit of human KEOPS<sup>37,42</sup>. It was further shown that C14 and Gon7 have the same structure and they interact identically with the LAGE3 (the human ortholog of Pcc1) and Pcc1 subunits, respectively. Therefore, C14 and Gon7, despite their weak sequence similarity, are homologs. Like Gon7, C14 binds to the homodimerization interface of LAGE3 and thereby prevents the dimerization of the human KEOPS complex<sup>28</sup>. Mutations in human GON7 lead to a milder form of Galloway-Mowat syndrome. Enzymatic assays using in vitro reconstituted system further established that C14 stimulates the t<sup>o</sup>A synthetic activity up to four-fold<sup>37</sup>.

The presence of a homologous fifth core subunit in the KEOPS complex from two distinct eukaryotic lineages raised the possibility that Gon7/C14 orthologs could also be present in Archaea. In search for the putative fifth KEOPS subunit in archaea, we identified in silico a small uncharacterized protein that showed significant sequence similarity to the Pcc1 KEOPS subunit. We present here the biochemical and structural characterization of this protein, named Pcc2, and establish it as a fifth core subunit of the archaeal KEOPS complex. Pcc2 shares several sequence and structural similarities with the eukaryotic Gon7/C14 proteins suggesting that the 5-subunit complex is the ancestral form of the KEOPS complex. Our work lays foundation for the future comprehensive studies of the biological and biochemical function of Pcc2 and warrants searches for the fifth core subunits in eukaryotic lineages outside animals and fungi.

## Results

### A paralog of Pcc1 is widely distributed among archaea

The recent discovery of a fifth subunit of KEOPS in metazoa raised the possibility that a corresponding subunit could exist in archaea. In search for such a subunit, we noticed in the model archaeon *Thermococcus kodakarensis* KOD1 the existence of two genes (TK1253 and TK0642) both annotated as KEOPS subunit Pcc1 in the NCBI database. TK0642 encoded an 82 aa protein (UniProt Q5JF86) while TK1253 encoded an 85 aa protein (UniProt Q5JGMS) (Supplementary Fig. 1A). The two proteins displayed rather low pairwise sequence identity (22%) but high sequence similarity (55%). The alignment of their sequences with the previously described Pcc1 from *Pyrococcus furiosus* (PfPcc1)<sup>34</sup> showed that TK0642 is more similar to PfPcc1 (55% sequence identity versus 29% sequence identity for TK1253) (Supplementary Fig. 1B). This suggested that TK0642 and TK1253 are paralogs derived from a gene duplication event whereby TK0642 encodes the canonical Pcc1 protein. We named the product of the gene TK1253 Pcc2 to underline the evolutionary link with Pcc1 protein family.

Next, we set out to investigate the distribution of Pcc2 among archaea using HMM-based sequence similarity searches. We detected Pcc2 orthologs in 11% of DPANN, 38% of Asgard, 83% of TACK, and 88% of Euryarchaeota genomes. Except for the DPANN superphylum where Pcc2 was found in a single lineage (Altiarchaeota), the vast majority of archaeal genomes contained both Pcc1 and Pcc2 encoding genes (Fig. 2A and Supplementary Fig. 2). These numbers should probably be regarded as conservative since we may have missed highly divergent orthologs. Moreover, when searching for Pcc2 orthologs in specific lineages we found that they were missing in protein sequence databases but we could detect them by performing manual tblastn searches, thus indicating that some *pcc2* genes (and likely also *pcc1* genes) were left undetected by automated gene annotation tools.

The alignment of representative Pcc1 or Pcc2 sequences from 27 archaeal phyla revealed low amount of global sequence conservation within each gene family, except for a few highly conserved residues grouped in two separate regions (box 1 and box 2) (Supplementary Fig. 3). While the box 1 sequence contains conserved residues in both

protein families, the box 2 is specific to each family. Of note, the conserved residues N<sup>72</sup>, S/T<sup>73</sup> in Pcc1 sequences are replaced by negatively charged residues D/E<sup>72</sup>, D/E<sup>73</sup> in Pcc2 sequences (*Pyrococcus furiosus* numbering, UniProt Q8TZII) (Fig. 2B). Consistent with them being paralogs, the phylogenetic tree of representative Pcc1 and Pcc2 sequences showed bipartite topology with one monophyletic clade corresponding to Pcc1 orthologs and the other to Pcc2 orthologs (Fig. 2C).

Collectively, the data show that Pcc2 and Pcc1 are paralogs emerging after an ancient gene duplication event that occurred before the diversification of Archaea. Following the duplication, the proteins were conserved in the majority of archaeal lineages and they diverged in their C-terminal part which harbors paralog-specific conserved residues. The discovery of Pcc2 in Archaea offered the exciting possibility that this paralog evolved to provide a specific function within the KEOPS complex.

### Pcc2 protein co-purifies with the archaeal KEOPS

To investigate if Pcc2 could bind to the KEOPS complex we produced in *E. coli* subcomplexes containing two or three of the known four KEOPS subunits from *Pyrococcus abyssi* (PaPcc1, PaKae1, PaBud32, PaCgi121) plus PaPcc2 whereby only PaPcc1 or PaPcc2 carried a hexahistidine tag (see material and methods for details). The corresponding cell lysates were mixed, and recombinant complexes were purified using classical two-step procedure consisting of Ni-NTA affinity chromatography followed by size exclusion chromatography.

This protocol yielded a complex containing PaCgi121 (C), PaBud32 (B), PaKae1 (K), PaPcc1 (P1), and PaPcc2 (P2) that migrated slightly below PaPcc1 (Fig. 3). We named this 5-subunit (5SU) complex CBKPIP2. After the first step of purification, the CBKPIP2 complex co-eluted with an excess of PaPcc1 and PaPcc2 subunits (Supplementary Fig. 4) which could be separated by size exclusion chromatography. PaPcc1 and PaPcc2 eluted as a single ~29 kDa peak suggesting they could form heterodimers (Supplementary Fig. 4) and that PaPcc1 could be the direct binding partner of PaPcc2 within the CBKPIP2 complex.

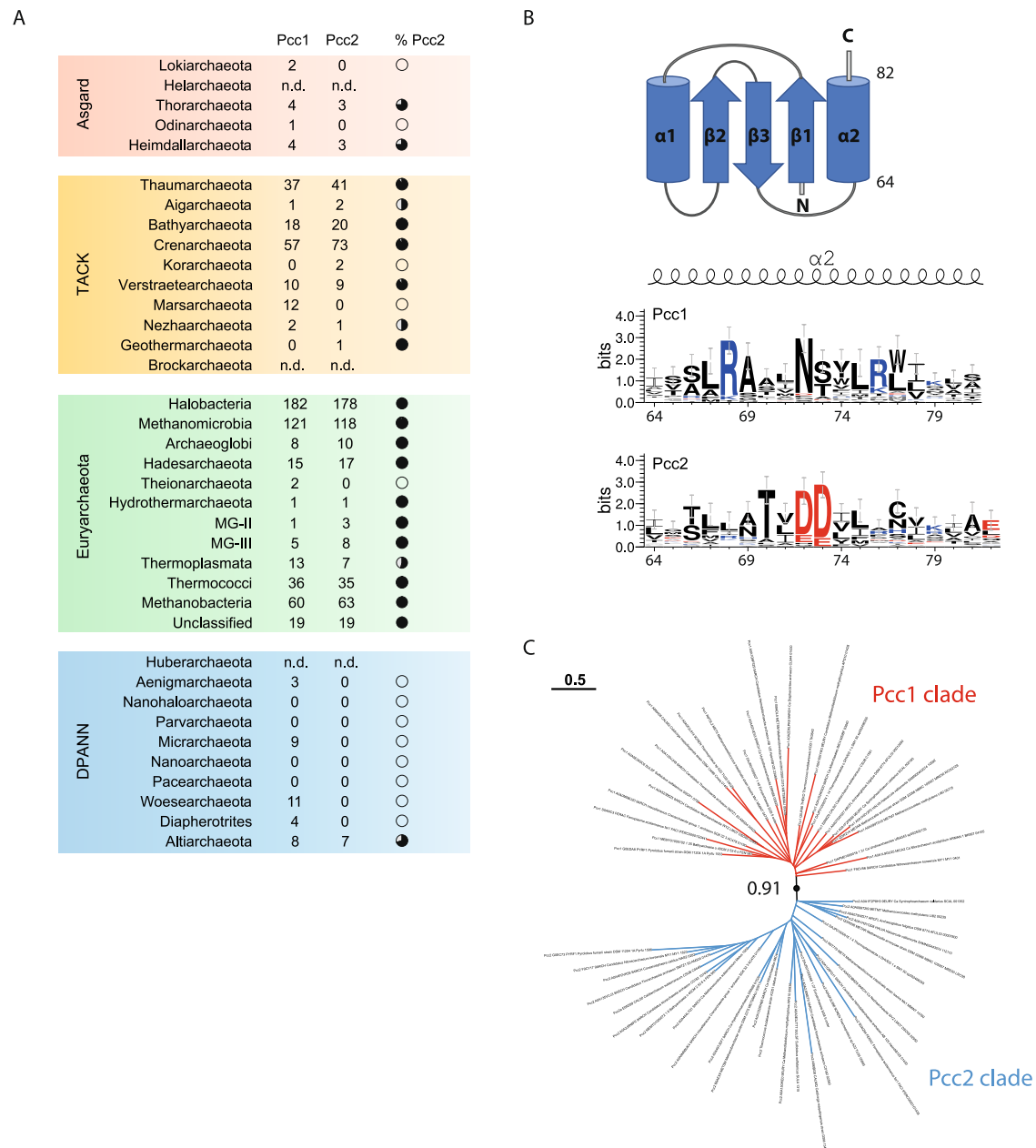
Because of their similarity in sequence and size, we hypothesized that PaPcc2 might mimic PaPcc1. To test this, we co-expressed PaCgi121, PaBud32, PaKae1, and PaPcc2His6 and applied the same two-step purification protocol. The three canonical KEOPS subunits co-purified with PaPcc2 thus showing that PaPcc2 can indeed replace PaPcc1 probably by binding directly to Kae1. We named this complex CBKP2 (Fig. 3, Supplementary Fig. 4). Alongside CBKP2, an excess of PaPcc2 could also be purified as a single peak with an apparent molecular mass of ~26 kDa suggesting that Pcc2 is a homodimer in solution and could, in analogy to Pcc1 homodimer, potentially provide the interface for the dimerization of the CBKP2 complex.

To investigate the oligomeric state of CBKPIP2 and CBKP2 we started by comparing their size exclusion profile with that of CBKPI1, known to build dimers in solution. CBKPI1 eluted as two separate peaks (66.3 mL and 73.6 mL) both containing the four subunits suggesting that the complex exists as a mixture of dimers and monomers in our experimental conditions (Fig. 3, Supplementary Fig. 4). The CBKPIP2 and CBKP2 each eluted as a single peak but their elution volumes (77.3 mL and 68.9 mL, respectively) were substantially different from each other and from that of the CBKPI1 complex suggesting that the three complexes adopt different quaternary structures.

Collectively, these experiments demonstrated that PaPcc2 binds both to CBKPI1 and to CBK complexes. The capacity of PaPcc2 to form heterodimers with PaPcc1, suggests direct interaction with PaPcc1 or, in absence of Pcc1, with the catalytic subunit Kae1. Pcc1 and Pcc2, alone or in combination, seem to dictate the oligomeric state of PaKEOPS.

### The Pcc2 crystal structure closely resembles that of Pcc1

Using the pure PaPcc2 fractions we collected during purification of CBKP2 complex we solved the crystal structure of PaPcc2 by

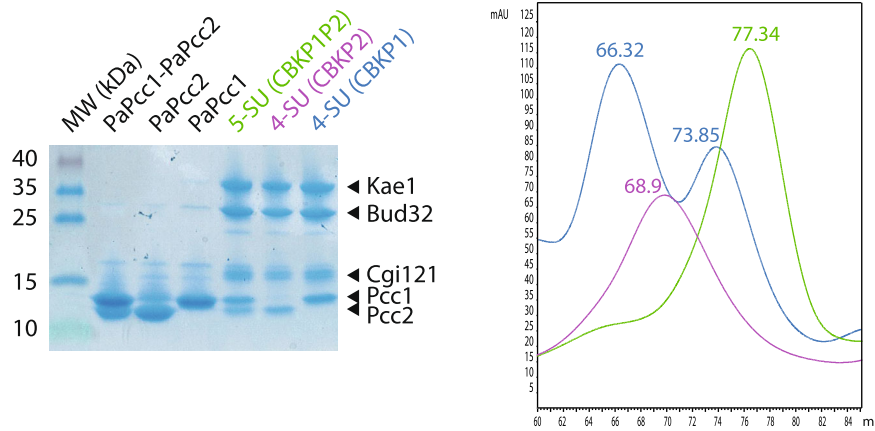


**Fig. 2 | Identification of Pcc2 proteins in archaea. A** Prevalence of Pcc1 and Pcc2 encoding genes among archaeal phyla. Number of identified Pcc1 or Pcc2 proteins is indicated in the corresponding columns. Full black circle corresponds to 100% and white circle to 0% of genomes containing both Pcc1 and Pcc2 encoding genes. n.d. not determined. Source data are provided in Supplementary Data 1. **B** Pcc1 and Pcc2 proteins exhibit divergent C-terminal sequences. The alignment of 27 representative Pcc1 and Pcc2 proteins was used to generate the sequence logo (see Supplementary Fig. 3 for complete logo). Only the part of the sequence corresponding to the  $\alpha 2$  helix of Pcc1 proteins is shown. The complete topology of Pcc1 proteins is shown on the top. The size of the letters is proportional to the

conservation score. The amino acid numbering is from Pcc1 ortholog of *Pyrococcus furiosus* (UniProt Q8TZII). **C** Pcc1 or Pcc2 sequences segregate into two distinct clades. Maximum likelihood phylogenetic tree was inferred from an alignment of 27 Pcc1 or Pcc2 sequences representative of the archaeal diversity. The bootstrap value for the branch separating Pcc1 and Pcc2 clades was evaluated from 500 boot trees using transfer bootstrap expectation. Tree scale gives the correspondence between the branch length and sequence evolution in number of substitutions per site. The phylogenetic signal contained in the sequence dataset was insufficient to resolve robustly the topology of the tree within each clade.

molecular replacement at a resolution of 1.84 Å (Supplementary Table 1). The asymmetric unit contained three nearly identical (RMSD around 1.3 Å) PaPcc2 copies. Two copies form a tight dimer whereas the third copy is more loosely associated with the two others. The latter copy forms a tight dimer with a symmetry related partner. PaPcc2 forms a three stranded anti-parallel  $\beta$ -sheet with three  $\alpha$  helices packed against one face of the sheet ( $\beta\alpha\alpha'\beta\beta\alpha$  topology) (Fig. 4A). Interestingly, despite moderate pairwise sequence identity (29%), PaPcc2 and Pcc1 from *Pyrococcus furiosus*

(PfPcc1) superpose very well (Z score of 7.2 and RMSD of 1.48 Å for 76 equivalent C $\alpha$  positions) (Fig. 4B). The main structural difference between PfPcc1 and PaPcc2 resides in the connection between strands  $\beta 1$  and  $\beta 2$  that consists of a single  $\alpha$  helix ( $\alpha 1$ ) for PfPcc1 and two perpendicularly oriented  $\alpha$  helices ( $\alpha 1$  and  $\alpha 1'$ ) for PaPcc2. Also the PaPcc2 homodimer structure is very similar to that of PfPcc1. In both dimers, the two monomers associate via their  $\beta 1$  strands to form a continuous six-stranded anti-parallel  $\beta$ -sheet and the two main helices are involved in hydrophobic packing. The



**Fig. 3 | Purification of recombinant PaKEOPS complexes.** SDS-PAGE analysis (left panel) and size-exclusion chromatography profiles (right panel) of purified recombinant KEOPS complexes produced in *E. coli* cells. See Supplementary Fig. 4

for complete chromatograms. The purification experiments were repeated at least four times with reproducible results.

superposition of the two homodimers reveals about 10 Å longitudinal sliding of PaPcc2 protomers alongside  $\beta$ 1 strand (Fig. 4B).

In the X-ray crystal structure of archaeal Kae1-Pcc1 complex, the  $\alpha$ 1 helix of Pcc1 is part of a four-helix bundle building the Kae1-Pcc1 interaction interface<sup>34</sup>. To explore how the observed differences in  $\alpha$ 1 helix conformation could impact the interaction of PaPcc2 with Kae1 protein we superposed the structures of the PaPcc2 homodimer with the PfPcc1 homodimer in complex with one Kae1 protein from the archaeon *Thermoplasma acidophilum* (Fig. 4C). This revealed a significant displacement of both  $\alpha$ 1 $\alpha$ 1' and  $\alpha$ 2 helices with respect to their PfPcc1 counterparts suggesting that PaPcc2 likely binds differently to Kae1. This may be one of the reasons why there are considerable differences in the quaternary structures of the CBKP1, CBKP1P2, and CBKP2 complexes (see further below).

### Structure of the PaPcc1-PaPcc2 heterodimer

During the gel-filtration purification step of PaCBKP1P2, we collected an excess of PaPcc1 and PaPcc2 as a single peak suggesting that these two proteins form a heterodimer. However, we could not exclude this peak contained a mixture of homodimers and/or heterodimers and we, therefore, decided to crystallize the protein compound in these fractions. We solved its structure at 3.18 Å resolution, establishing that PaPcc1-PaPcc2 heterodimer was present in the crystal (Supplementary Table 1, Fig. 5A). The two proteins align their respective  $\beta$ 1-strands and create a six stranded anti-parallel  $\beta$ -sheet similar to one observed in the PaPcc2 and PfPcc1 homodimers. The heterodimer is further stabilized by the hydrophobic packing of the two C-terminal helices. Overall, the Pcc1 and Pcc2 homodimers and the Pcc1-Pcc2 heterodimer have all very similar structures (Fig. 5B). The structures of PaPcc2 as present in the homodimer and in the PaPcc2-PaPcc1 heterodimer are almost identical (RMSD 0.98 Å) suggesting that binding of PaPcc2 to PaPcc1 does not induce a significant conformational change in the PaPcc2 structure.

PfPcc1 binds to Kae1 by engaging its two helices in a four helical bundle. The opposite surface in principle should remain available to form dimers and hence induce the formation of a (super)dimer of two heterotetramers. Our data show that Pcc2 mimics a Pcc1 subunit upon formation of the heterodimer. Although we do not dispose of the CBKP1P2 structure, in this complex Pcc2 binds very likely to Pcc1 as it does in the heterodimer. The subunit interface (1125.1 Å<sup>2</sup> vs. 771.9 Å<sup>2</sup>) and the number of potential hydrogen bonds across interface (18 vs. 8) is much higher in the PaPcc1-PaPcc2 heterodimer than for the PaPcc2 homodimer, suggesting the heterodimer might be the preferred arrangement.

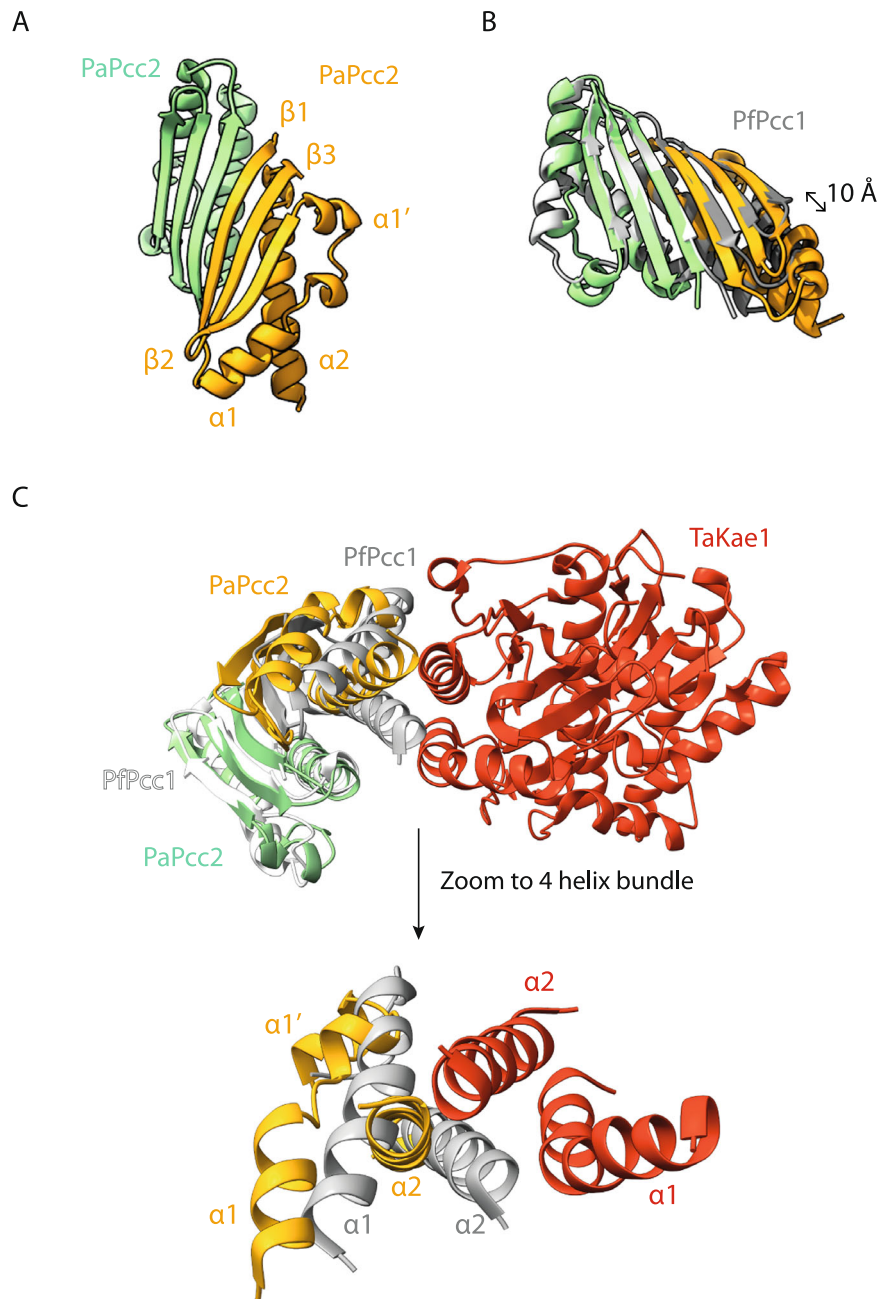
### SAXS analysis of PaKEOPS complexes in solution

To learn more about the oligomeric status of the CBKP1, CBKP1P2, and CBKP2 complexes in solution we performed SAXS experiments. The Guinier plots for the three complexes were linear, testifying that the samples were monodisperse (Supplementary Fig. 5). From these plots, we calculated the radius of gyration, a measure of the spread of molecular mass, which was almost identical for CBKP1 and CBKP2, but was smaller for CBKP1P2 suggesting that the latter had a lower mass. (Supplementary Fig. 5).

We previously established using gel filtration that PaCBKP1 forms superdimers<sup>38</sup>. Our present SAXS data for this complex yield a Mw of approximately 186 kDa, consistent with a 2:2:2:2 superdimer stoichiometry (Supplementary Table 2). Similarly, for CBKP2 we obtained a Mw of approximately 180 kDa indicating that CBKP2 also forms a superdimer with 2:2:2:2 stoichiometry (Supplementary Table 2). In contrast, for CBKP1P2 we measured a Mw of 90 kDa close to the theoretical Mw for a heteropentamer with 1:1:1:1 stoichiometry (101 kDa) (Supplementary Table 2).

For the fitting of our SAXS data (Fig. 6), we constructed 3D models of the various complexes starting from high resolution structures of subcomplexes (Kae1-Pcc1, PDB 3ENC; Kae1-Bud32, PDB 3EN9, Bud32-Cgi121, PDB 4WW9) and our present PaPcc1-Pcc2 and PaPcc2-Pcc2 structures. The initial model of the octameric CBKP1<sup>40</sup> resulted in a few slight steric clashes which could be relieved using the FoXsDock software, a method for protein docking against SAXS profiles. The theoretical SAXS curve of the adjusted CBKP1 octamer model fitted very well to the experimental data with a  $\chi^2$  value of 1.39 (Supplementary Fig. 6A). The CBKP1 octamer adopts a V-like shape, in line with the experimental maximal extension value of about 188 Å (Supplementary Fig. 6A, middle panel). The Kae1-Pcc1 interface in the model octamer is slightly twisted compared to that in the Kae1-Pcc1 structure (PDB code 3ENO) but we modelled it as the interface of OSGEP-LAGE3 (PDB code 6GWJ).

For modelling the CBKP2 structure, we replaced Pcc1 in the CBKP1 model with PaPcc2 and imposed the Pcc2 homodimer as the central dimerization unit to construct the octamer. This model fitted the experimental SAXS curve satisfactorily with a  $\chi^2$  value of 1.64 (Supplementary Fig. 6A, right panel). In stark contrast to the V-shaped CBKP1 model, CBKP2 octamer model adopted an open conformation whereby the two heterotetramers were pointing into opposite directions (Supplementary Fig. 6B). The model of the CBKP1P2 pentamer fits very well the experimental SAXS curve ( $\chi^2$  value of 0.619) and is in good agreement with the maximal calculated extension value of 125 Å (Supplementary Fig. 6A, left panel).



**Fig. 4 | Crystal structure of PaPcc2 homodimer.** **A** The subunits are in green and yellow, the third copy in the asymmetric unit is not shown. The  $\beta 1$  strands of the green and the orange copies associate in antiparallel to create a continuous  $\beta$ -sheet. The third copy (gray) is loosely associated with the orange copy and this interface is probably purely crystallographic. This copy forms with a crystal symmetry related mate the same type of dimer as the green/orange. **B** Superposition of the PaPcc2

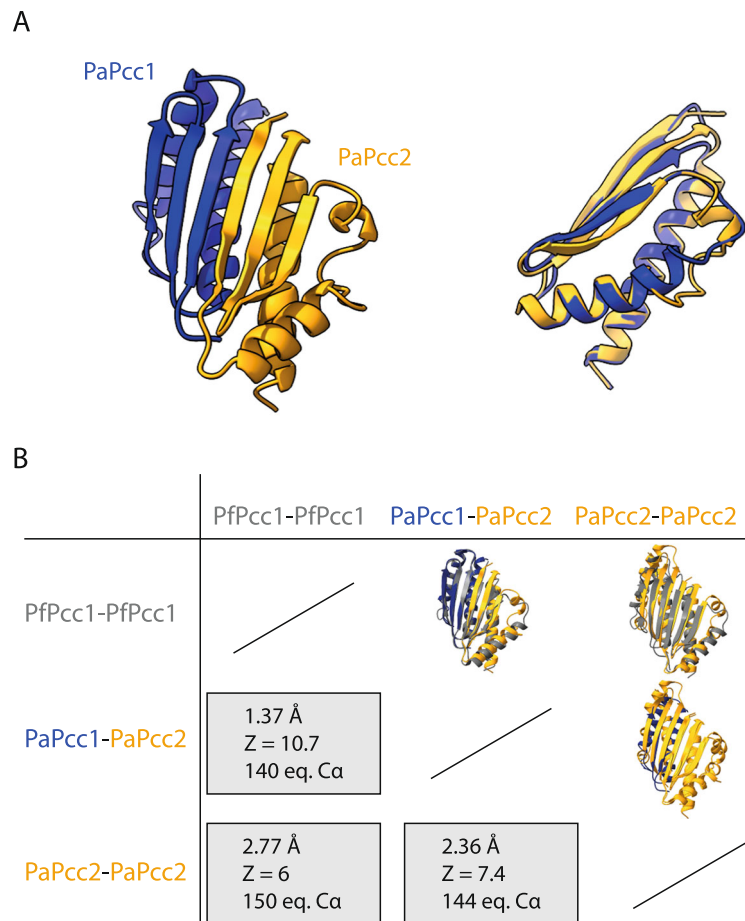
homodimer (green and orange) with PfPcc1 homodimer (grey). The 10 Å longitudinal shift of one PaPcc2 monomer relative to PfPcc1 monomer is indicated with double arrow. **C** Superposition of PaPcc2 homodimer (green and orange) with the crystal structure of PfPcc1 homodimer (grey and white) bound to one copy of TaKae1 (red) from *T. acidophilum* (PDB 3ENO). A zoom to the four-helix bundle composing the PfPcc1-TaKae1 interface is shown below.

Together, the SAXS data confirmed that PaPcc2 prevents the dimerization of CBKPIP2 similar to the effect of Gon7/C14 proteins within the yeast and human KEOPS complexes. Furthermore, the data indicate that in absence of PaPcc1, PaPcc2 induces dimerization of CBKIP2 but its quaternary structure is significantly different from that of CBKIP1.

#### PaPcc2 and PaPcc1 are not isofunctional

We next examined the impact of PaPcc2 on the activity of the pentameric and octameric complexes. To this end, we measured the maximal velocity of  $t^6A$  synthesis and tRNA binding properties for the three complexes.

The kinetic data showed that CBKIP1 and CBKIP2 were active while CBKIP2 complex was inactive (Fig. 7A). Notably, we could fully restore the activity of the CBKIP2 complex when adding five-fold molar excess of PaPcc1 to the reaction mixture. The reaction could also be partially restored by equimolar concentrations of PaPcc1 suggesting that PaPcc1 has the capacity to displace PaPcc2 from the preformed CBKIP2 complex (Supplementary Fig. 7A). For the two active complexes, we compared the maximal reaction velocity as a function of complex concentration. This revealed that the CBKIP1 complex exhibited a two-fold higher initial velocity as compared to the CBKPIP2 complex (Supplementary Fig. 7B). Assuming that CBKIP1 complex has two catalytic sites per octamer and can process two



**Fig. 5 | Crystal structure of PaPcc1-PaPcc2 heterodimer. A** Cartoon presentation of PaPcc1-PaPcc2 structure (in blue and orange, respectively) is shown on the left and the superposition of PaPcc2 onto PaPcc1 is on the right. **B** Pairwise comparison

of PaPcc1 and PaPcc2 homo and heterodimers. The structures were superposed using ChimeraX 1.4. The RMSD of atomic coordinates and Z score are given for the indicated number of equivalent alpha carbons.

tRNA molecules simultaneously this would indicate that CBKPIP2 and CBKPI complexes are equally efficient for the catalysis of t<sup>6</sup>A formation.

We next measured the binding affinity of the three complexes toward substrate tRNA using fluorescence anisotropy. We first established that our experimental conditions allowed the binding equilibrium to be achieved (see materials and methods). Titration of the CBKPI, CBKPIP2 and CBKPI2 yielded binding equilibrium constants of 13, 100 nM, and ~1800 nM respectively (Fig. 7B, Supplementary Fig. 7C). This suggested that the CBKPI2 complex is affected in tRNA binding, which could be one reason explaining why this complex is inactive.

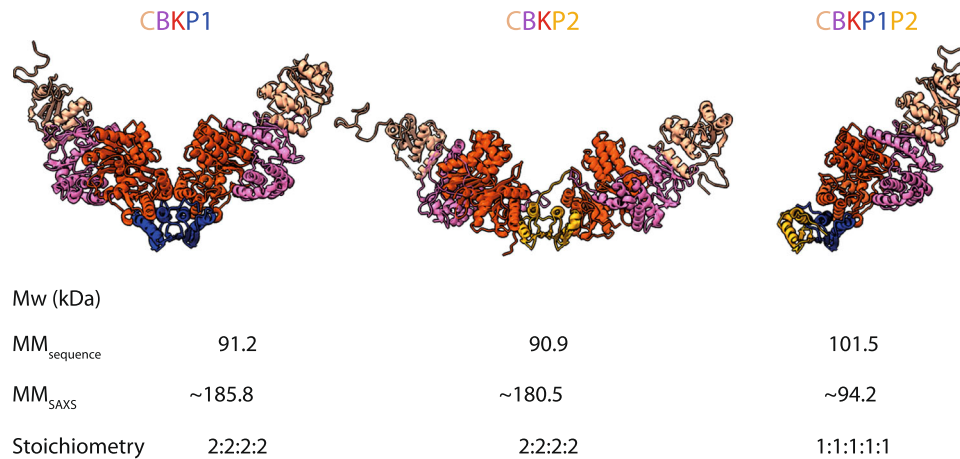
Collectively, the data show that the CBKPI and CBKPIP2 complexes exhibit robust t<sup>6</sup>A biosynthetic activity and tRNA binding affinities in the nanomolar range. However, CBKPI2 has substantially less affinity for tRNA and is inactive indicating that, despite striking structural resemblance, Pcc1 and Pcc2 are not isofunctional. This further suggests that Pcc2 paralogs evolved to provide a specific function as a fifth subunit within the KEOPS complex.

### Search for Pcc2 orthologs in eukaryotes

The evolutionary proximity between archaea and eukaryotes and wide conservation of Pcc2 in archaea raised the possibility that Pcc2 orthologs could also be present in eukaryotes. To test this hypothesis, we performed sensitive sequence similarity searches against eukaryotic protein sequences using the sequence conservation profile of archaeal Pcc2 proteins. This search retrieved only seven eukaryotic sequences passing the significance threshold. However, structure

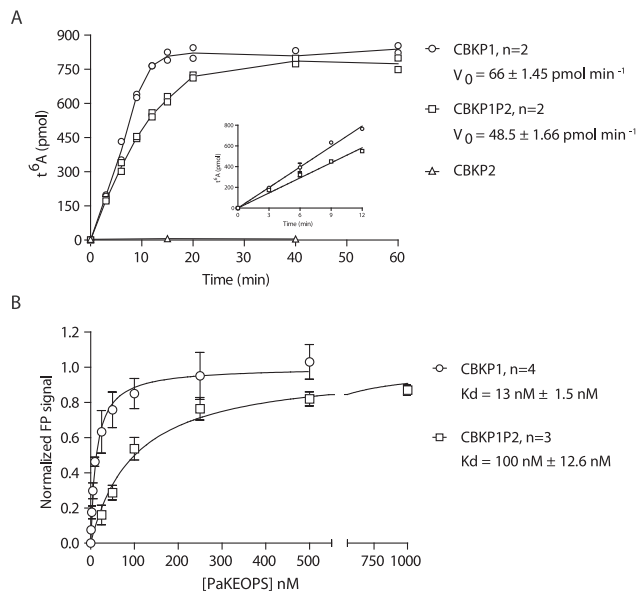
supported sequence alignment showed that the characteristic pair of acidic residues (D/E<sup>72</sup>, D/E<sup>73</sup>) was missing in those sequences indicating that these are likely Pcc1 orthologs (Supplementary Fig. 8). The data thus suggested that Pcc2 orthologs are specific for Archaea or that eukaryotic Pcc2 homologs are too divergent to be detected by sequence similarity searches.

To explore a putative evolutionary link between Pcc2 and Gon7/C14, we superposed the 3D structures of the human LAGE3/C14 (LAGE3 is the human ortholog of Pcc1) and the archaeal Pcc1/Pcc2 complexes (Fig. 8A). Despite the fact that Pcc2 and C14 have seemingly a different fold, the similarities of their structures and their interaction with Pcc1/LAGE3 are striking. The two complexes are characterized by a continuous  $\beta$ -sheet covered by long helices that are the main interacting elements. C14 consists of a N-terminal  $\beta$ -hairpin followed by about 40 disordered residues that precede the long  $\alpha$ -helix. Pcc2 forms a three stranded sheet with  $\beta 1\beta 3\beta 2$  topology. Superposition of the C14/LAGE3 and Pcc1/Pcc2 complexes perfectly aligns  $\beta 2$  of C14 with  $\beta 3$  from Pcc2. A deletion event in *S. cerevisiae* Gon7 of 39 residues (between position 12 and 51) would yield a Pcc2-like structure. To further explore the possibility that Gon7/C14 proteins could be highly diverged Pcc2 orthologs, we performed structure-guided alignment which showed that Pcc2 sequences can actually be fairly well aligned with full length eukaryotic Gon7 and C14 sequences (Fig. 8B, Supplementary Fig. 9). The majority of the positions with high similarity scores in Pcc1 and Pcc2 sequences are also found in Gon7 and C14 proteins. Notably, Gon7 and C14 proteins contain a D/E<sup>73</sup> signature residue (which is never found in Pcc1 proteins) suggesting that they may indeed derive from a Pcc2-like ancestor.



**Fig. 6 | Model structures of PaKEOPS complexes.** The structures of the three complexes modelled against the scattering data are shown. Calculated molecular mass from sequence and from scattering data and the deduced stoichiometry are indicated at the bottom. The subunits are colored as follows: Cgi121 (tan), Bud32 (violet), Kae1 (red), Pcc1 (blue), and Pcc2 (orange) CBKP1 and CBKP2 are dimers of

two heterotetramers whereby Pcc1 and Pcc2 serve as dimerization platform, respectively. CBKP2 superdimer adopts much more open conformation with the two heterotetramers pointing into opposite directions. CBKP1P2 is a heteropentamer with Pcc2 protein engaging the Pcc1 via its dimerization interface.



**Fig. 7 | Enzymatic and tRNA binding activity of PaKEOPS complexes.** **A** The  $t^6A$  formation was measured for the three complexes at different time points and the linear part of the curve (inset) was used to calculate the steady-state velocity ( $V_0$ ) of the reaction. All complexes were present at  $1 \mu\text{M}$  concentration.  $n = 2$  biologically independent experiments. **B** tRNA binding measurements using fluorescence polarisation. Titration experiments were performed with fixed concentration of fluorescently labelled tRNA<sup>Lys</sup> (UUU) from *Pyrococcus abyssi* and increasing concentrations of PaKEOPS complexes. The fluorescence signal was normalized using  $B_{\text{max}}$  values 108.5 and 193.6 for CBKP1P2 and CBKP1, respectively.  $n = 4$  (CBKP1) or  $n = 3$  (CBKP1P2) biologically independent experiments. Data are presented as mean values  $\pm$  standard deviation. Source data are provided as a Source Data file.

If Pcc2 and eukaryotic Gon7/C14 share common ancestry, then the most parsimonious evolutionary scenario postulates that all eukaryotic lineages should be equipped with Pcc2 or Gon7/C14 orthologs. Since we failed to identify Pcc2-like orthologs by sequence similarity searches, we tested if Gon7/C14-like proteins can be found outside opisthokonta (which regroup animals and fungi). Using HMM profile searches derived from representative alignment of Gon7 or C14 orthologs we only identified Gon7 sequences in two closely related

plants *Carpinus fangiana* (hornbeam) and *Quercus suber* (cork oak) belonging to the Fagales order (Supplementary Fig. 10). The phylogenetic analysis placed these sequences within the fungal Gon7 orthologs suggesting that these are not genuine plant Gon7 orthologs but rather genes that were acquired by horizontal transfer from fungi or contamination.

Overall, the data indicate that Pcc2 proteins are specific for Archaea and that Gon7/C14 are either highly divergent orthologs of Pcc2 or a result of convergent evolution whereby opisthokonta had recruited a distinct gene to fulfill a function in regulating the oligomeric state of KEOPS complex.

## Discussion

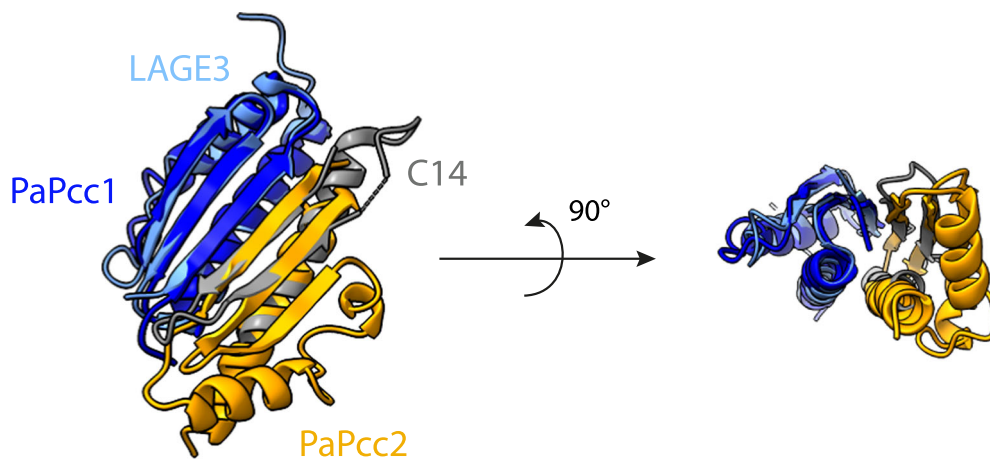
$t^6A$  biosynthetic pathway is a biological process that appeared before the divergence of life into the three distinct domains. The driving forces behind the evolution of this machinery and the resulting impact on physiology and evolution of cells are not well understood. This is particularly true for the KEOPS complex that evolved in eukaryotes and archaea by acquisition of several accessory proteins to assist the catalytic subunit Kae1. In the present study, we identified an additional accessory protein, which we named Pcc2, as a previously unrecognized core subunit of KEOPS in Archaea.

## Gene duplication is a universal mechanism for the evolution of the $t^6A$ synthetic pathway

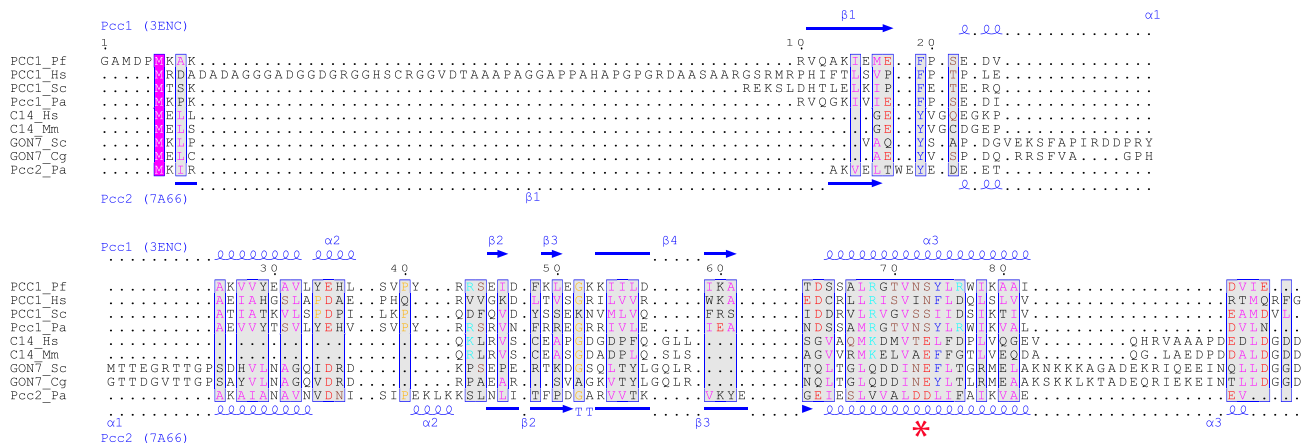
We show here that Pcc2 is a paralog of Pcc1 that emerged before the diversification of archaea through an ancient duplication event of the common ancestral gene. This was also concomitantly reported by Wu and colleagues, who, similar to us, noticed a presence of two genes (SiRe\_1278 and SiRe\_1701) encoding Pcc1-like proteins in the crenarchaeon *Saccharolobus islandicus* REY15A<sup>43</sup>. Occurrence of paralogs within  $t^6A$  synthetic machineries was also described in bacteria and eukaryotes. One of the two known accessory subunits in the bacterial system (TsaB) is a truncated version of the catalytic subunit TsaD which lost its catalytic activity<sup>5</sup>. In addition, a few distantly related bacterial lineages harbor YciO, an uncharacterized paralog of the TsaC protein<sup>32,44</sup>. Interestingly, in addition to Pcc1 (dubbed LAGE3), humans encode two additional members of the Pcc1 protein family (CTAG1 and CTAG2) which are expressed in several human tumors and in normal testes and ovaries<sup>22</sup>. It appears therefore that gene duplication followed by mutation is a common mechanism employed by the members of the three domains of life for evolution of the  $t^6A$  pathway.



A



B



**Fig. 8 | Sequence and structural comparison of Pcc2 and Gon7/C14 proteins.**

**A** Structural comparison of PaPcc1-PaPcc2 and human LAGE3-C14 heterodimers. Superimposition of the crystallographic PaPcc1-PaPcc2 structures (in blue and orange, respectively) on LAGE3-C14 (in light blue and grey, respectively). The left part is a top view showing the continuous 5-strand and 6-strand  $\beta$  antiparallel sheets stabilising the two interfaces in LAGE3-C14 and PaPcc1-PaPcc2 complexes. The right part is a side view showing the  $\alpha$  helices that also contribute to the building of the interfaces. **B** Multiple sequence alignment of Pcc1, Pcc2, C14 and Gon7 proteins from representative metazoan (*H. sapiens*, *M. musculus*), fungal (*S. cerevisiae*,

*C. glabrata*) and archaeal (*P. abyssi*, *P. furiosus*) species. Red star denotes the N<sup>72</sup>, S/T<sup>73</sup> and D/E<sup>73</sup>, D/E<sup>73</sup> motifs characteristic for Pcc1 and Pcc2 paralogs, respectively. Secondary structure distribution of Pcc1 from *Pyrococcus furiosus* (3ENC) and of Pcc2 from *Pyrococcus abyssi* (7A66) are shown above and below the alignment, respectively. The residues are colored according to their physico-chemical properties. The positions with global similarity score higher than the threshold value of 0.7 are framed in blue. Multiple alignment with an extended set of representative sequences is shown in Supplementary Fig. 9.

### Pcc2 evolved to prevent the formation of KEOPS superdimers

Following the gene duplication event, the two paralogs mainly diverged by accumulating mutations in their C-terminal parts, specifically the  $\alpha 2$  helix, which contributes to the Pcc1-Pcc2 interface. We found that the interface statistics (the surface and the number of interactions) of the heterodimer is in favor over Pcc1-Pcc1 or Pcc2-Pcc2 homodimers thus suggesting that Pcc2 evolved to prevent the homodimerization of Pcc1. In line with this, Wu and colleagues reported that adding recombinant Pcc2 from the archaeon *S. islandicus* to the preformed binary Kae1-Pcc1 complex (which has 2:2 stoichiometry) abolished its dimerization<sup>43</sup> indicating, as suggested by us, that the Pcc1-Pcc2 interaction is preferred over the Pcc1-Pcc1 interaction.

Second major difference between Pcc1 and Pcc2 orthologs lies in the  $\alpha 1$  helix which in Pcc2 orthologs is split in two perpendicularly oriented smaller helices. This, as well as a significant shift in the position of the  $\alpha 2$  helix, suggests a substantially different Pcc2-Kae1 interface compared to the Pcc1-Kae1 interface. Accordingly, the CBKP2

and CBKP1 dimers (as modelled from our SAXS data) adopt two very different conformations. These putative differences at the Pcc2-Kae1 interface could explain why the CBKIP2 complex does not form a superdimer.

### Pcc2 is essential in some archaea

Using genome-wide transposon mutagenesis Zhang and colleagues identified 441 essential genes in the crenarchaeon *Sulfolobus islandicus* among which figured both Pcc1 and Pcc2<sup>45</sup>. A similar systematic mutagenesis approach in an euryarchaeon *Methanococcus maripaludis* found however that Pcc1 was essential but not Pcc2<sup>46</sup>. A targeted genetic investigation of KEOPS complex in another euryarchaeal model organism *Haloflex volcanii* (which also encodes Pcc2, HVO\_1146), reported that the Pcc1-encoding gene (HVO\_0652) could not be deleted using markerless pop-in pop out strategy, and the mutants obtained by marker replacement were very slow-growing<sup>17</sup>. These in vivo data indicate, in line with our findings, that Pcc2 cannot functionally complement Pcc1 and therefore has

evolved to provide a distinct function. Moreover, this function is, at least in some archaea, essential most likely because it is required for t<sup>6</sup>A synthesis in vivo. Whether Pcc2 associates with the KEOPS complex in archaea in vivo was not addressed in this work, however, this was investigated by Wu and colleagues in *S. islandicus*. The authors expressed the His-tagged versions of Pcc1, Kae1 or Bud32 in *S. islandicus* and performed affinity purification from cell lysates coupled with mass spectrometry. In the three assays Kae1, Pcc1, and Pcc2 co-purified suggesting that the five subunit complex is the natural version of KEOPS in archaea<sup>43</sup>.

The enzymatic assays conducted in this work show that Pcc2 does not potentiate the t<sup>6</sup>A activity in vitro suggesting that it has no direct influence upon the catalysis. The Sicheri group showed that the C14 subunit increased the t<sup>6</sup>A activity of the human KEOPS complex by 3–4 fold in vitro but it was suggested that this may be attributed to an indirect effect through stabilization of the complex<sup>37</sup>. Indeed, the experiments in human cell lines established that mutations in C14-encoding gene (but not in Kae1 ortholog OSGEF) lead to the decrease in protein levels of the other four KEOPS subunits<sup>28</sup>. Incidentally, we noticed during the purification of the KEOPS complexes that the purified pentameric complex is more stable in solution than the octameric CBKP1 complex which has a tendency to form aggregates. Pcc2 may therefore indirectly impact the intracellular t<sup>6</sup>A levels by stabilizing the KEOPS complex.

### Pcc2 and Gon7/C14 proteins may be homologs

Several observations suggest that Pcc2 and eukaryotic Gon7/C14 may be homologs. Sequence alignments show that Pcc1 and Pcc2 stretches with high sequence similarity scores can be almost systematically aligned with Gon7/C14 sequences. This is particularly true for the  $\alpha$ 2 helix of Pcc1/Pcc2 which (together with  $\beta$ 1 strand) provides the main dimerization interface. The Gon7/C14 proteins harbor a Pcc2-type signature residue D/E<sup>73</sup> which is never present in Pcc1 sequences pointing to a possible common ancestry with the Pcc2 proteins. Consistent with this hypothesis, structural alignment of Pcc1/Pcc2 with human LAGE3/C14 complex shows a remarkable conservation of the interaction interface despite the fact that the  $\beta$ -sheets of C14/Gon7 and Pcc2 proteins have a different topology ( $\beta$ 1 $\beta$ 3 $\beta$ 2 for Pcc2 and  $\beta$ 1 $\beta$ 2 for C14/Gon7) suggesting that this interface is a functionally important requirement that has been conserved from archaea to eukaryotes. Beyond sequence and structural similarities, Pcc2 has a shared function with eukaryotic Gon7/C14 in preventing the formation of KEOPS superdimers.

### Concluding remarks

Understanding the purpose behind the recruitment of several accessory proteins to support the Kae1 catalytic function is one of the most exciting and still poorly understood mysteries in the KEOPS field. The discovery of Pcc2 now suggests that the 5-subunit KEOPS is the ancestral form of the complex that emerged before the diversification of archaea and was vertically transmitted to eukaryotes from an archaeal ancestor. The possibility to prevent the dimerization of the KEOPS complex thus seems to be a common principle in both archaea and eukaryotes. It is therefore plausible to think that highly divergent (in sequence but not in structure) orthologs of Gon7/C14 exist in many if not all eukaryotic lineages. Understanding how Pcc2 regulates the KEOPS function in vivo in archaea is one of the exciting future prospects of this work with possible implications for the eukaryotic system.

### Methods

#### Cloning procedures

The genes encoding KEOPS complex of the archaeon *Pyrococcus abyssi* (PaKEOPS) PAB\_RS09585 (PaKae1), PAB\_RS08575 (PaBud32), PAB\_RS06930 (PaCgi121) and PAB\_RS01470 (PaPcc1) were amplified

by polymerase chain reaction (PCR) using pKEOPS500 plasmid (Perrochia et al., 2013a) as DNA template. The gene PAB\_RS00380 encoding PaPcc2 was amplified by polymerase chain reaction (PCR) using genomic DNA of *P. abyssi* GE5 strain as template. Different combinations of bicistronic or polycistronic sequences were cloned into pET26b (Novagen) via *Nde*I-*Xho*I restriction sites either using T4 ligase (NEB) or by Gibson assembly (NEB) following the protocol supplied by the manufacturer.

pKPIHis6 and pKP2His6 contain each a bicistronic sequence with PaKae1 plus PaPcc1 or PaKae1 plus PaPcc2 encoding genes, respectively. The genes encoding PaPcc1 and PaPcc2 are fused at their 3' end to a sequence encoding a hexa-histidine tag.

pKP2Strep-PIHis6 contains a tricistronic sequence with PaKae1, PaPcc2 and PaPcc1 encoding genes. The gene encoding PaPcc2 is fused at its 3' end to a sequence encoding a Strep-tag whereas the gene encoding PaPcc1 is fused at 3' end to a sequence encoding a hexa-histidine tag.

pCBK and pCB contain PaKae1, PaBud32 and PaCgi121 encoding genes or PaBud32 and PaCgi121 encoding genes respectively. These genes are not fused to an epitope tag encoding sequence.

### Heterologous expression and purification of recombinant proteins

Recombinant proteins were expressed in *E. coli* Rosetta2 (DE3) pLysS strain (Novagen). The cells were transformed either with pCBK, pCB, pKPIHis6, pKP2His6, or with pKP2strepPI His6 plasmids and were grown at 37 °C in LB medium with kanamycin (50  $\mu$ g/ml) and chloramphenicol (25  $\mu$ g/ml) until the optical density (600 nm) reached 0.6–0.7. Gene expression was induced by addition of IPTG (1 mM), and incubation was continued at 37 °C for 2 hours. Cells were collected by centrifugation (5000 g, 15 min) and resuspended in cold lysis buffer (20 mM Tris-HCl pH 8, 500 mM NaCl, 5 mM  $\beta$ -mercaptoethanol, 10% Glycerol) supplemented with complete EDTA-free protease inhibitors cocktail (Roche). Cell pellets containing different subcomplexes were mixed in defined ratio (see below) and lysed at 4 °C with a one-shot cell disruptor (Constant Systems). The lysate was clarified by centrifugation at 30,000 g for 30 min at 18 °C. The clarified lysate was incubated 10 to 15 min at 65 °C to denature the majority of *E. coli* proteins. The precipitates were removed by centrifugation at 30,000 g for 20 min at 18 °C. The soluble fraction was supplemented with 10 mM imidazole and His-tagged proteins were purified by affinity chromatography on NiNTA column (Qiagen) at room temperature. Equilibration, washes, and elution steps were done with lysis buffer containing respectively 10, 40, and 400 mM imidazole. Fractions of interest were pooled, concentrated at 18 °C and submitted to size exclusion chromatography on a HiLoadR 16/600 SuperdexR 200 pg or a HiLoadR 16/600 SuperdexR 75 pg (GE Healthcare) equilibrated with 20 mM Tris-HCl pH 8, 500 mM KCl, 10% Glycerol and 5 mM  $\beta$  Mercaptoethanol.

In order to reconstitute and purify different PaKEOPS complexes, cells transformed with different recombinant plasmids (see cloning procedures) were mixed in defined cell pellet mass ratio and lysed together.

PaPcc1-PaPcc2 and PaPcc2 proteins used for SAXS analysis and crystallization were purified from pCBK transformed cells mixed with pKPIP2 transformed cells (cell mass ratio 2:1) and from pCBK transformed cells mixed with pKP2 transformed cells (cell mass ratio 2:1), respectively. In these conditions, significant excess of PaPcc1-PaPcc2 complex and PaPcc2 homodimer over the complete PaKEOPS complexes is obtained when applying the two-step purification protocol described above.

The PaKEOPS complexes used for SAXS analysis and enzymatic assays were purified from pCBK transformed cells mixed with pCB transformed cells and either pKPIHis6 transformed cells (cell mass ratio 2:1:1) or pKP2His6 transformed cells (cell mass ratio 2:1:1) or pKPIHis6P2Strep- transformed cells (cell mass ratio 2:1:1).

### Sequence similarity searches for identification of Pcc1 and Pcc2

To search for Pcc1 and Pcc2 orthologs, we performed iterative searches against UniProtKB database using the JackHMMR program (<https://www.ebi.ac.uk/Tools/hmmer/search/jackhmmer>) with default settings. We manually selected 21 representative archaeal species each encoding both Pcc1 and Pcc2. We then generated separate alignments for Pcc1 or Pcc2 sequences using MAFFT<sup>47</sup> and used those as input. After two iterations, we retrieved 1264 significant hits that could be unambiguously assigned as archaeal Pcc1 (646 proteins) or Pcc2 (618 proteins) based on e-values. The table containing the raw data and the summary of the search results is available as supplementary file.

To generate the sequence logos specific for Pcc1 or Pcc2 proteins we chose from the above dataset 27 sequences representative of the archaeal diversity. When possible, we chose species with complete genomes, otherwise, metagenome-assembled genomes (MAGs) were chosen for which the Pcc1 and Pcc2 encoding genes were found within the same MAG. The sequences are available on the figshare website under <https://doi.org/10.6084/m9.figshare.21640394.v1>

### Phylogenetic analysis

For the phylogenetic analysis, we chose couples of Pcc2 and Pcc1 paralogs from one representative of 27 phyla of Archaea covering the entire archaeal diversity. Protein alignment was done with MAFFT V7 with the amino acid matrix Blosum 30 and iterative refinement methods L-INS-I or G-INS-I<sup>47</sup>. The trimming was performed with BMGE with a BLOSUM30 matrix and the  $-b$  1 parameter<sup>48</sup>. The Maximum likelihood trees were constructed using the IQ-TREE v1.7 software (<http://www.iqtree.org/>)<sup>49</sup> with the best substitution model as suggested by the ModelFinder option -MFP. The branch robustness was estimated by the SH-aLRT approximate likelihood ratio test<sup>50</sup>, the ultrafast bootstrap approximation<sup>51</sup> (10000 replicates for each), and with the transfer bootstrap expectation approaches<sup>52</sup> based on 500 boot trees.

Sequences of Gon7 orthologs from Metazoa (ENOG503BUEY, 60 sequences) and Fungi (ENOG503P560, 32 sequences) were retrieved from EggNOG database 5.0.0<sup>53</sup> which harbours sequences selected for diversity and filtered by genome quality. Sequences were aligned with the T-coffee webserver<sup>54</sup> and the tree was inferred using IQ-TREE v1.7 software using automatic selection of sequence evolution model.

### Crystallization of the PaPcc1-PaPcc2 heterodimer

PaPcc1-PaPcc2 complex was purified by affinity chromatography on NiNTA column (Qiagen) at room temperature followed by size exclusion chromatography as described above except that the lysis buffer contained 300 mM NaCl. Fractions containing the heterodimeric PaPcc1-PaPcc2 complex were pooled and concentrated to 7.9 mg mL<sup>-1</sup> for crystallization trials. Crystals were obtained by mixing 100 nL of the protein solution with 100 nL of 15% PEG 6K and 5% glycerol using the sitting-drop vapor diffusion method. The crystallization drop was equilibrated against a 70  $\mu$ L reservoir solution of 15% PEG 6K and 5% glycerol. Drops were incubated at 18 °C and crystals appeared after about 21 days. Crystals were cryo-protected by quick-soaking in 20% PEG 6K and 20% glycerol prior to flash freezing in liquid nitrogen.

### Crystallization of the PaPcc2 homodimer

PaPcc2 homodimer was purified using the same procedure as for PaPcc1-PaPcc2 heterodimer. Fractions containing pure PaPcc2 were concentrated to 22 mg mL<sup>-1</sup> for crystallization trials. Crystals of PaPcc2 were obtained by mixing 100 nL of the protein solution with 100 nL of 30% PEG 3 K 0.2 M lithium sulfate and 0.1 M Tris HCl pH 8.5 using the sitting drop vapor diffusion (crystals appeared after 17 days of incubation at 18 °C). The crystallization drop was equilibrated against a 70  $\mu$ L reservoir solution of same the precipitant. Crystals were

cryo-protected by quick-soaking in 50% paraffin and 50% paratone prior to flash freezing in liquid nitrogen.

### Structure solution and refinement

X-ray diffraction data collection was carried out at Synchrotron SOLEIL on beamline Proxima 2 (Saint Aubin, France) at 100 K. Data were processed, integrated and scaled with the XDS software package<sup>55</sup>. The PaPcc2 crystals belong to space group C12<sub>1</sub> with unit cell  $a = 76.80 \text{ \AA}$   $b = 81.31 \text{ \AA}$   $c = 66.48 \text{ \AA}$  and  $\alpha = 90^\circ$   $\beta = 121.59^\circ$   $\gamma = 90^\circ$ . The structure of PaPcc2 was solved by molecular replacement using PHASER<sup>56</sup> implemented in the CCP4 suite. A search model for PaPcc2 was obtained by MODELLER<sup>57</sup> using the known structure of PfPcc1 (PDB 3ENC). 3 copies of PaPcc1 were found in the asymmetric unit of the crystal. The initial structure was refined with BUSTER and interactively manually adjusted with COOT<sup>58</sup>. The PaPcc1-PaPcc2 crystals belong to space group P3<sub>2</sub> 21 with parameters of unit cell of  $a = 69.898 \text{ \AA}$   $b = 69.898 \text{ \AA}$   $c = 92.751 \text{ \AA}$  and  $\alpha = 90^\circ$   $\beta = 90^\circ$   $\gamma = 120^\circ$ . One copy of the heterodimer was found in the asymmetric unit. The initial structures were refined with BUSTER and interactively manually adjusted with COOT<sup>58</sup>. Data reduction and refinement statistics are gathered in the Supplementary Table 1.

### SEC-SAXS analysis

SAXS experiments were carried out on the SWING beamline at SOLEIL synchrotron (Saint-Aubin, France). The sample to detector (Eiger 4 M Dectris) distance was set to 1500 mm, allowing reliable data collection over the momentum transfer range  $0.005 \text{ \AA}^{-1} < q < 0.5 \text{ \AA}^{-1}$  with  $q = 4\pi \sin \theta / \lambda$ , where  $2\theta$  is the scattering angle and  $\lambda$  is the wavelength of the X-rays ( $\lambda = 1.0 \text{ \AA}$ ). To collect data on homogenous protein samples, SAXS data were collected at the exit of a size exclusion high-performance liquid chromatography (SEC HPLC Bio-SEC3 Agilent) column directly connected to the SAXS measuring cell. 65  $\mu$ L of PaCgi121/PaBud32/PaKae1/PaPcc1/PaPcc2 (CBKP1P2), CBKP1, and CBKP2 samples concentrated at 2.55, 0.72, and 3 mg/mL, respectively, were loaded into the column equilibrated with 20 mM Tris pH 8.5, 300 mM NaCl, and 5 mM 2-mercaptoethanol and 5% glycerol. Flow rate was set at 300  $\mu$ L/min, frame duration was 1.0 s, and the dead time between frames was 0.01 s. The protein concentration was estimated by UV absorption measurement at 280 and 295 nm using a spectrophotometer located immediately upstream of the SAXS measuring cell. A large number of frames were collected before the void volume and averaged to account for buffer scattering. SAXS data were normalized to the intensity of the incident beam and background (i.e., the running buffer) was subtracted using the program FoxTrot50, the Swing in-house software. The scattered intensities were displayed on an absolute scale using the scattering by water. Identical frames under the main elution peak were selected and averaged for further analysis. Radii of gyration, maximum particle dimensions, and molecular masses were determined using PrimusQT<sup>59</sup>.

### Modelling of the complete KEOPS complex

The PaKEOPS complexes were modelled using a combination of structures of archaeal KEOPS subcomplexes and of the present PaPcc1-PaPcc2 structure by the Modeller software<sup>60</sup>. The Dadimodo<sup>61</sup> and FoxsDock<sup>62</sup> programs were then used to adjust the models of CBKP1P2, CBKP1, and CBKP2 to optimally fit the experimental curve I(q). The final adjustment was performed using the program CRY SOL<sup>63</sup>.

### Comparative sequence and structure analysis

Multiple alignment of the Pcc1 and Pcc2 sequences was built using the Blosum62 matrix with MAFFT vs. 7<sup>47</sup> (Katoh et al., 2019). The sequence conservation was rendered using WebLogo 3 (<http://weblogo.threeplusone.com/>)<sup>64</sup>. Structure-based alignment of a subset of these members was performed using the ESPRIPT web-server (<https://escript.ibcp.fr/ESPript/ESPript/>)<sup>65</sup>. All figures of structures were generated with PyMol (The PyMOL Molecular Graphics System, Version

2.0 Schrödinger, LLC). Structure similarity comparisons were performed with the PDBeFold web server (<https://www.ebi.ac.uk/msd-srv/ssf/>). Surface of protein-protein interface and the number of potential hydrogen bonds was calculated using PDBePISA web server (<https://www.ebi.ac.uk/pdbe/pisa/>).

### tRNA production and purification

tRNA<sup>Lys</sup> (UUU) from *P. abyssi* was produced by run-off transcription from a pUC18 vector containing a tRNA-encoding gene under the control of the T7 promoter. The transcription unit (TU) consists from 5' to 3' end of: hammerhead ribozyme carrying T7 promoter, tRNA gene, glmS ribozyme and a double MS2 tag<sup>66</sup>. The complete sequence of the TU is shown in Supplementary Fig. 11. This construct allows to produce, after transcription and ribozyme cleavage, tRNA molecules with precisely defined 5' and 3' extremities. The transcription reaction was carried out in the final volume of 10 mL and it contained 0.02 µg of *Bpil*-linearized template plasmid, 40 mM Tris-HCl pH 8, 1 mM spermidine, 0.01% Triton X-100, 5 mM DTT, 2 mM of each NTP, 10 mM of MgCl<sub>2</sub>, 80 U of RNasin (Promega) and 0.03 µg/µL of T7 RNA polymerase. The mixture was incubated for at least 4 hours at 37 °C and ribozymes were removed by addition of glucosamine-6-phosphate (1 mM final concentration) and MgCl<sub>2</sub> solution (5 mM final concentration) and incubation for 20 min at room temperature. The reaction was stopped by heating at 80 °C for 10 min. The reaction mixture was concentrated by ultrafiltration (10 kDa cut-off filter) to a final volume of 1 mL and the tRNA was purified on a 10% acrylamide gel (19:1) containing 8 M urea. The band containing tRNA was excised and the tRNA was extracted by crush and soak method with 5 mL of extraction solution (0.3 M sodium acetate pH 5.3, 0.1% SDS (w/v), 1 mM EDTA pH 8.0). The tRNA-containing solution was concentrated by ultrafiltration using 3 kDa cut-off filter to about 500 µL and precipitated by addition of two volumes of cold absolute ethanol. After centrifugation, the tRNA pellet was washed twice with 70% ethanol (v/v), resuspended in water, and stored at -80 °C.

### In vitro assay for the synthesis of t<sup>6</sup>A-modified tRNA

The assays were performed as previously described<sup>30</sup> with minor modifications: briefly, the reaction mixture (final volume of 25 µL) contained 70 µM C<sup>14</sup> L-threonine (55 mCi/mmol, Isobio), 100 µM MnCl<sub>2</sub>, 5 mM Na<sub>2</sub>CO<sub>3</sub> and 10 µM of Pa\_tRNA<sup>Lys</sup> (UUU) in 50 mM Tris-HCl pH 8, 200 mM KCl, 5 mM MgCl<sub>2</sub>, 5 mM DTT. PaKEOPS complexes and PaSua5 were typically added to 2 µM and 0.5 µM final concentration, respectively, unless stated otherwise. For the initial velocity measurements, the CBKPI and CBKPI2 complexes were added at 1 µM final concentration. The reaction mixture was incubated for indicated time at 55 °C and the reaction was stopped by addition of 1 mL of 15% (w/v) of trichloroacetic acid (TCA) and incubated on ice for 1 h. Precipitated material was retained on pre-wet glass microfibre GF/F filters (Whatmann) using vacuum apparatus (Millipore). Filters were washed with 2 mL of 5% TCA and 3 mL of 95% EtOH and dried. Liquid scintillation counting was used to quantify the amount of t<sup>6</sup>A-modified tRNA using the standard curve to convert CPM to pmol of tRNA (1 pmol of tRNA = 97.8 CPM).

### Fluorescence anisotropy measurements

Fluorescence anisotropy was used to determine the equilibrium binding constant for tRNA-PaKEOPS complexes. The binding mixture contained, in a final volume of 100 µL, 20 mM Tris, HCl pH 8, 150 mM KCl, 3 mM MgCl<sub>2</sub>, 1% glycerol, increasing concentration of PaKEOPS complexes, and 1 nM of Pa\_tRNA<sup>Lys</sup> (UUU) labelled at its 3' end with fluorescein (Horizon). We measured the polarisation signal between 1 and 50 min at 20 °C and concluded that the binding equilibrium was reached within 1 minute. For the binding assays, samples were incubated at 20 °C for 15 min and the fluorescence polarisation was

measured in 96-well half area flat bottom non-treated black polystyrene microplates (Costar) using Infinite M1000 Pro (Tecan) microplate reader set to 470 nm excitation wavelength and 530 nm emission wavelength. All measurements were done at least in triplicates. After subtracting the polarization values obtained for Pa\_tRNA<sup>Lys</sup> alone, fluorescence polarisation in arbitrary units was plotted against protein concentration and the resulting saturation curve was fitted to one-site specific binding equation (GraphPad Prism)  $Y = B_{max} * X / (K_d + X)$ , where Y is the difference between the anisotropy of bound and free tRNA, B<sub>max</sub> is the maximal polarisation signal, X is KEOPS concentration and K<sub>d</sub> is the equilibrium dissociation constant.

### Reporting summary

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

### Data availability

The manually curated set of Pcc1 and Pcc2 sequences used in this study are available at <https://doi.org/10.6084/m9.figshare.21640394.v1>. The raw data from HMM-based searches are available as Supplementary Data 1. The crystal structures of Pcc1 and Pcc2 proteins were deposited to the PDB database, accession numbers 7A66 and 7A67. Source data are provided with this paper.

### References

- Koonin, E. V. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* **1**, 127–136 (2003).
- Galperin, M. Y. 'Conserved hypothetical' proteins: prioritization of targets for experimental study. *Nucleic Acids Res.* **32**, 5452–5463 (2004).
- El Yacoubi, B. et al. The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenosine in tRNA. *Nucleic Acids Res.* **37**, 2894–2909 (2009).
- Srinivasan, M. et al. The highly conserved KEOPS/EKC complex is essential for a universal tRNA modification, t<sup>6</sup>A. *EMBO J.* **30**, 873–881 (2011).
- El Yacoubi, B. et al. A role for the universal Kae1/Qri7/YgjD (COG0533) family in tRNA modification: t<sup>6</sup>A biosynthesis. *EMBO J.* **30**, 882–893 (2011).
- Machnicka, M. A., Olchowik, A., Grosjean, H. & Bujnicki, J. M. Distribution and frequencies of post-transcriptional modifications in tRNAs. *RNA Biol.* **11**, 1619–1629 (2014).
- Cantara, W. A. et al. The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.* **39**, D195–D201 (2011).
- Schweizer, M. P., Chheda, G. B., Baczynskyj, L. & Hall, R. H. Aminoacyl nucleosides. VII. N-(Purin-6-ylcarbamoyl)threonine. A new component of transfer ribonucleic acid. *Biochemistry* **8**, 3283–3289 (1969).
- Murphy, F. V., Ramakrishnan, V., Malkiewicz, A. & Agris, P. F. The role of modifications in codon discrimination by tRNA<sup>Lys</sup>UUU. *Nat. Struct. Mol. Biol.* **11**, 1186–1191 (2004).
- Stuart, J. W. et al. Functional Anticodon Architecture of Human tRNA<sup>Lys3</sup> Includes Disruption of Intraloop Hydrogen Bonding by the Naturally Occurring Amino Acid Modification, t<sup>6</sup>A<sup>+</sup>. *Biochemistry* **39**, 13396–13404 (2000).
- Durant, P. C., Bajji, A. C., Sundaram, M., Kumar, R. K. & Davis, D. R. Structural effects of hypermodified nucleosides in the *Escherichia coli* and human tRNA<sup>Lys</sup> Anticodon Loop: The effect of Nucleosides s<sup>2</sup>U, mcm<sup>5</sup>U, mcm<sup>5</sup>s<sup>2</sup>U, mnm<sup>5</sup>s<sup>2</sup>U, t<sup>6</sup>A, and ms<sup>2</sup>t<sup>6</sup>A<sup>+</sup>. *Biochemistry* **44**, 8078–8089 (2005).
- Lescrier, E. The naturally occurring N6-threonyl adenine in anticodon loop of *Schizosaccharomyces pombe* tRNA<sup>I</sup> causes formation of a unique U-turn motif. *Nucleic Acids Res.* **34**, 2878–2886 (2006).

13. Lin, C. A., Ellis, S. R. & True, H. L. The Sua5 protein is essential for normal translational regulation in yeast. *Mol. Cell. Biol.* **30**, 354–363 (2010).
14. Yarian, C. et al. Accurate translation of the genetic code depends on tRNA modified nucleosides. *J. Biol. Chem.* **277**, 16391–16395 (2002).
15. Phelps, S. S., Malkiewicz, A., Agris, P. F. & Joseph, S. Modified Nucleotides in tRNALys and tRNAVal are Important for Translocation. *J. Mol. Biol.* **338**, 439–444 (2004).
16. Weissenbach, J. & Grosjean, H. Effect of Threonylcarbamoyl Modification (t<sup>6</sup>A) in Yeast tRNAArgIII on Codon-Anticodon and Anticodon-Anticodon Interactions. A thermodynamic and kinetic evaluation. *Eur. J. Biochem* **116**, 207–213 (1981).
17. Naor, A. et al. A genetic investigation of the KEOPS complex in halophilic Archaea. *PLoS ONE* **7**, e43013 (2012).
18. Thiaville, P. C. et al. Global translational impacts of the loss of the tRNA modification t(6)A in yeast. *Micro. Cell* **3**, 29–45 (2016).
19. Lin, H. et al. CO<sub>2</sub>-sensitive tRNA modification associated with human mitochondrial disease. *Nat. Commun.* **9**, 1875 (2018).
20. Oberto, J. et al. Qri7/OSGEPL, the mitochondrial version of the universal Kae1/YgjD protein, is essential for mitochondrial genome maintenance. *Nucleic Acids Res.* **37**, 5343–5352 (2009).
21. He, M.-H. et al. KEOPS complex promotes homologous recombination via DNA resection. *Nucleic Acids Res.* **47**, 5684–5697 (2019).
22. Kisseleva-Romanova, E. et al. Yeast homolog of a cancer-testis antigen defines a new transcription complex. *EMBO J.* **25**, 3576–3585 (2006).
23. Daugeron, M.-C. et al. Gcn4 misregulation reveals a direct role for the evolutionary conserved EKC/KEOPS in the t<sup>6</sup>A modification of tRNAs. *Nucleic Acids Res.* **39**, 6148–6160 (2011).
24. Downey, M. et al. A genome-wide screen identifies the evolutionarily conserved KEOPS complex as a telomere regulator. *Cell* **124**, 1155–1168 (2006).
25. Meng, F.-L. et al. Sua5p a single-stranded telomeric DNA-binding protein facilitates telomere replication. *EMBO J.* **28**, 1466–1478 (2009).
26. Lin, C.-J. et al. An extensive allelic series of *Drosophila kae1* mutants reveals diverse and tissue-specific requirements for t<sup>6</sup>A biogenesis. *RNA* **21**, 2103–2118 (2015).
27. Braun, D. A. et al. Mutations in KEOPS-complex genes cause nephrotic syndrome with primary microcephaly. *Nat. Genet* **49**, 1529–1538 (2017).
28. Arrondel, C. et al. Defects in t(6)A tRNA modification due to GON7 and YRDC mutations lead to Galloway-Mowat syndrome. *Nat. Commun.* **10**, 3967–3967 (2019).
29. Lauhon, C. T. Mechanism of N<sup>6</sup>-Threonylcarbamoyladenonsine (t<sup>6</sup> A) Biosynthesis: isolation and characterization of the intermediate Threonylcarbamoyl-AMP. *Biochemistry* **51**, 8950–8963 (2012).
30. Perrochia, L. et al. In vitro biosynthesis of a universal t<sup>6</sup>A tRNA modification in Archaea and Eukarya. *Nucleic Acids Res.* **41**, 1953–1964 (2013).
31. Deutsch, C., El Yacoubi, B., de Crécy-Lagard, V. & Iwata-Reuyl, D. Biosynthesis of Threonylcarbamoyl Adenosine (t<sup>6</sup> A), a Universal tRNA Nucleoside. *J. Biol. Chem.* **287**, 13666–13673 (2012).
32. Thiaville, P. C., Iwata-Reuyl, D. & de Crécy-Lagard, V. Diversity of the biosynthesis pathway for threonylcarbamoyladenonsine (t(6)A), a universal modification of tRNA. *RNA Biol.* **11**, 1529–1539 (2014).
33. Beenstock, J. & Sicheri, F. The structural and functional workings of KEOPS. *Nucleic Acids Res.* **49**, 10818–10834 (2021).
34. Mao, D. Y. L. et al. Atomic structure of the KEOPS Complex: An ancient protein kinase-containing molecular machine. *Mol. Cell* **32**, 259–275 (2008).
35. Zhang, W. et al. Crystal structures of the Gon7/Pcc1 and Bud32/Cgi121 complexes provide a model for the complete yeast KEOPS complex. *Nucleic Acids Res* **43**, 3358–3372 (2015).
36. Hecker, A. et al. Structure of the archaeal Kae1/Bud32 fusion protein MJ1130: a model for the eukaryotic EKC/KEOPS subcomplex. *EMBO J.* **27**, 2340–2351 (2008).
37. Wan, L. C. K. et al. Proteomic analysis of the human KEOPS complex identifies C14ORF142 as a core subunit homologous to yeast Gon7. *Nucleic Acids Res* **45**, 805–817 (2017).
38. Perrochia, L., Guetta, D., Hecker, A., Forterre, P. & Basta, T. Functional assignment of KEOPS/EKC complex subunits in the biosynthesis of the universal t(6)A tRNA modification. *Nucleic Acids Res.* **41**, 9484–9499 (2013).
39. Beenstock, J. et al. A substrate binding model for the KEOPS tRNA modifying complex. *Nat. Commun.* **11**, 6233 (2020).
40. Wan, L. C. K. et al. Structural and functional characterization of KEOPS dimerization by Pcc1 and its role in t<sup>6</sup> A biosynthesis. *Nucleic Acids Res* **44**, 6971–6980 (2016).
41. Wan, L. C. K. et al. Reconstitution and characterization of eukaryotic N<sup>6</sup>-threonylcarbamoylation of tRNA using a minimal enzyme system. *Nucleic Acids Res.* **41**, 6332–6346 (2013).
42. Costessi, A. et al. The human EKC/KEOPS complex is recruited to Cullin2 Ubiquitin Ligases by the human tumour antigen PRAME. *PLoS ONE* **7**, e42822 (2012).
43. Wu, P. et al. The archaeal KEOPS complex possesses a functional-Gon7 homolog and has an essential function independent of the cellular t<sup>6</sup>A modification level. *mLife* 1–17. <https://doi.org/10.1002/mlf2.12051> (2023).
44. Gerdes, S. et al. Synergistic use of plant-prokaryote comparative genomics for functional annotations. *BMC Genom.* **12**(Suppl 1), S2 (2011).
45. Zhang, C., Phillips, A. P. R., Wipfler, R. L., Olsen, G. J. & Whitaker, R. J. The essential genome of the crenarchaeal model *Sulfolobus islandicus*. *Nat. Commun.* **9**, 4908 (2018).
46. Sarmiento, F., Mrázek, J. & Whitman, W. B. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proc. Natl Acad. Sci.* **110**, 4726–4731 (2013).
47. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinforma.* **20**, 1160–1166 (2019).
48. Criscuolo, A. & Grihaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).
49. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
50. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
51. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
52. Lemoine, F. et al. Renewing Felsenstein’s phylogenetic bootstrap in the era of big data. *Nature* **556**, 452–456 (2018).
53. 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).
54. Di Tommaso, P. et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res* **39**, W13–W17 (2011).
55. Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr.* **66**, 125–132 (2010).

56. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
57. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
58. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
59. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
60. Sali, A., Potterton, L., Yuan, F., van Vlijmen, H. & Karplus, M. Evaluation of comparative protein modeling by MODELLER. *Proteins* **23**, 318–326 (1995).
61. Evrard, G., Mareuil, F., Bontems, F., Sizun, C. & Perez, J. DADIMODO: a program for refining the structure of multidomain proteins and complexes against small-angle scattering data and NMR-derived restraints. *J. Appl. Cryst.* **44**, 1264–1271 (2011).
62. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **44**, W424–W429 (2016).
63. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOLE – a Program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
64. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
65. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
66. Batey, R. T. & Kieft, J. S. Improved native affinity purification of RNA. *RNA* **13**, 1384–1389 (2007).

## Acknowledgements

We acknowledge the support of the SOLEIL (St Aubin) and ESRF (Grenoble) synchrotron beamline staffs. We thank S. Planquel of the I2BC crystallization platform for technical assistance. This work was funded by Agence Nationale de la Recherche, grant number ANR-18-CE11-0018 to T.B. and by the French Infrastructure for Integrated Structural Biology (FRISBI) (ANR-10-INSB-05-01) to H.v.T.

## Author contributions

S.M., B.C., and H.v.T. designed SAXS and protein crystallization assays. S.M. performed SAXS and protein crystallization assays and solved

protein crystal structures. N.L. provided technical help. V.D.C. and T.B. designed and performed bioinformatic and phylogenetic analyses. M-C. D. and T.B. designed protein purification procedures, enzymatic activity assays, and fluorescence anisotropy assays. M-C. D. performed protein purification, enzymatic activity assays, and fluorescence anisotropy assays. M-C. D., S.M., V.D.C., B.C., H.v.T., and T.B. wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-36210-y>.

**Correspondence** and requests for materials should be addressed to Herman van Tilbeurgh or Tamara Basta.

**Peer review information** *Nature Communications* thanks Kira Makarova and the other, anonymous, reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023