SUPPORTING INFORMATIONS

SUPPLEMENTARY TABLE

Table S1: Primers used in this study

Primer	Sequence
OLC5	CCGCTCGAGAGTGCATTGATAAAAAGGGA
OLC3	CGCGGATCCTCACCTCAGCCTTATTGTG
OLΔKHC5	CATATGCTCGAGGCCAGAAAGCCCGAGTACAAG
OLΔC3	CGCGGATCCAGTTGAAAGCCCAAATTT
OL1A261	CACGCTCACTTGGATGCTAGCGGAATGTTACCA
OL2A261	TGGTAACATTCCGCTAGCATCCAAGTGAGCGTG
OL1A594	GATGGATTCTCTGGTGCTGCGGATAGGAGGGAA
OL2A594	TTCCCTCCTATCCGCAGCACCAGAGAATCCATC
OLsR47 MutCG3	AAGGTGAATCAGCACTCAGCAGATCCTCATCACTCCTCAGGACC
OLsR47 MutCG5	CCGGAATTCTAATACGACTCACTATAGATGAAGATGATGAGCTCG GCAGGTCCTGAGGAGTGATGAGG
OLsR47 Mut21U5	CCGGAATTCTAATACGACTCACTATAGATGAAGATGATGAGCTCG GTAGGTCCTGAGGAGTGATGAGG
OLsR47 Mut21U3	AAGGTGAATCAGCACTCAAGATCCTCATCACTCCTCAGGACC

SUPPLEMENTARY FIGURE S1

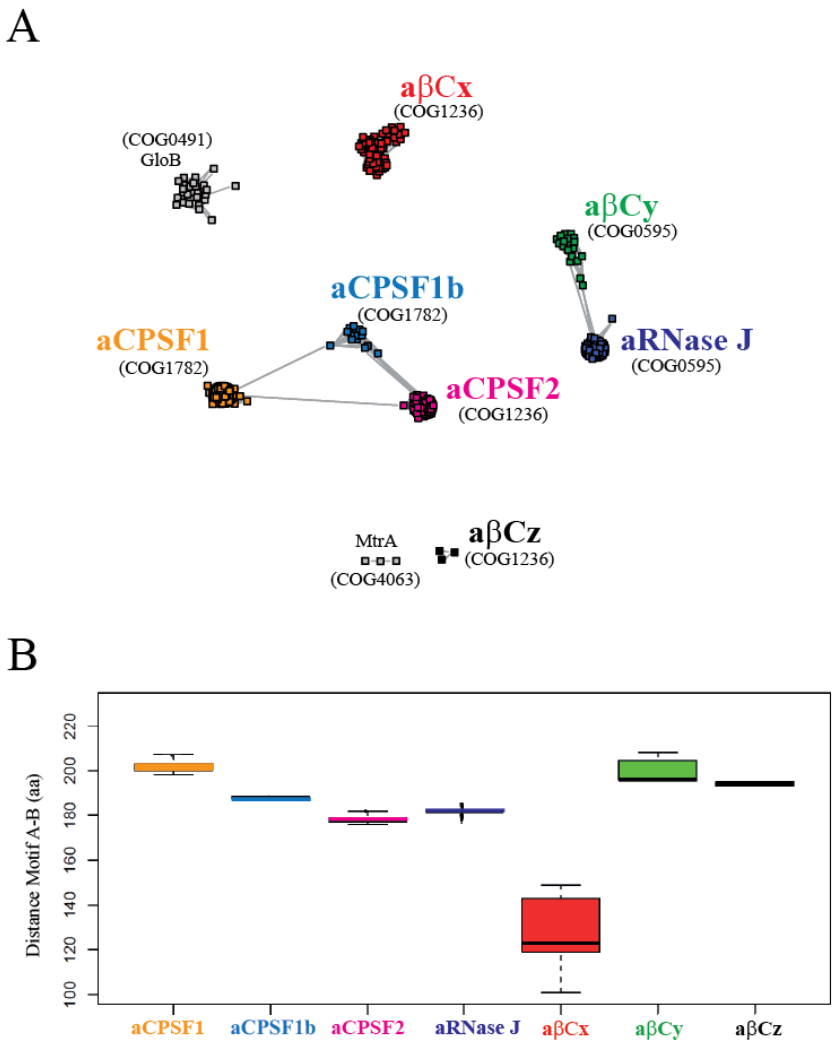

Figure S1

(*A*) Graph of orthologous archaeal β-CASP proteins. The pairs of orthologous links between proteins were retrieved from our database and used to construct the graph with nodes corresponding to proteins (small colored squares) and edges to the orthologous links between proteins (grey lines). The nine clusters were named according to available functional information (aCPSF, aRNase J, GloB, MtrA), or as archaeal β-CASP families (aβC[x:z]). The GloB (COG0491), MtrA (COG4063) clusters appeared to be false positives (see Methods) and were excluded from further analysis. The aβCy and aβCz clusters included highly divergent sequences, scattered in different taxonomic orders and therefore not monophyletic (*B*). Distribution of the length (in amino acid residues) of the β-CASP domains of the seven archaeal β-CASP clusters. The distance in amino acids between the A and B motifs were computed for each protein. The results were summarized as boxplots for each archaeal β-CASP cluster. The distance is in an interval of 160 to 210 amino acids for aCPSF1, aCPSF1b, aCPSF2, aRNase J, αβCy and αβCz (average 202 ± 3.3, 188 ± 2.0, 178 ± 3.2 and 181 ± 1.3, 197±14.6 and 194±1.4 respectively) except for the heterogeneous αβCx, which falls between 100 and 160 amino acids (average 129 ± 13.5).

SUPPLEMENTARY FIGURE S2

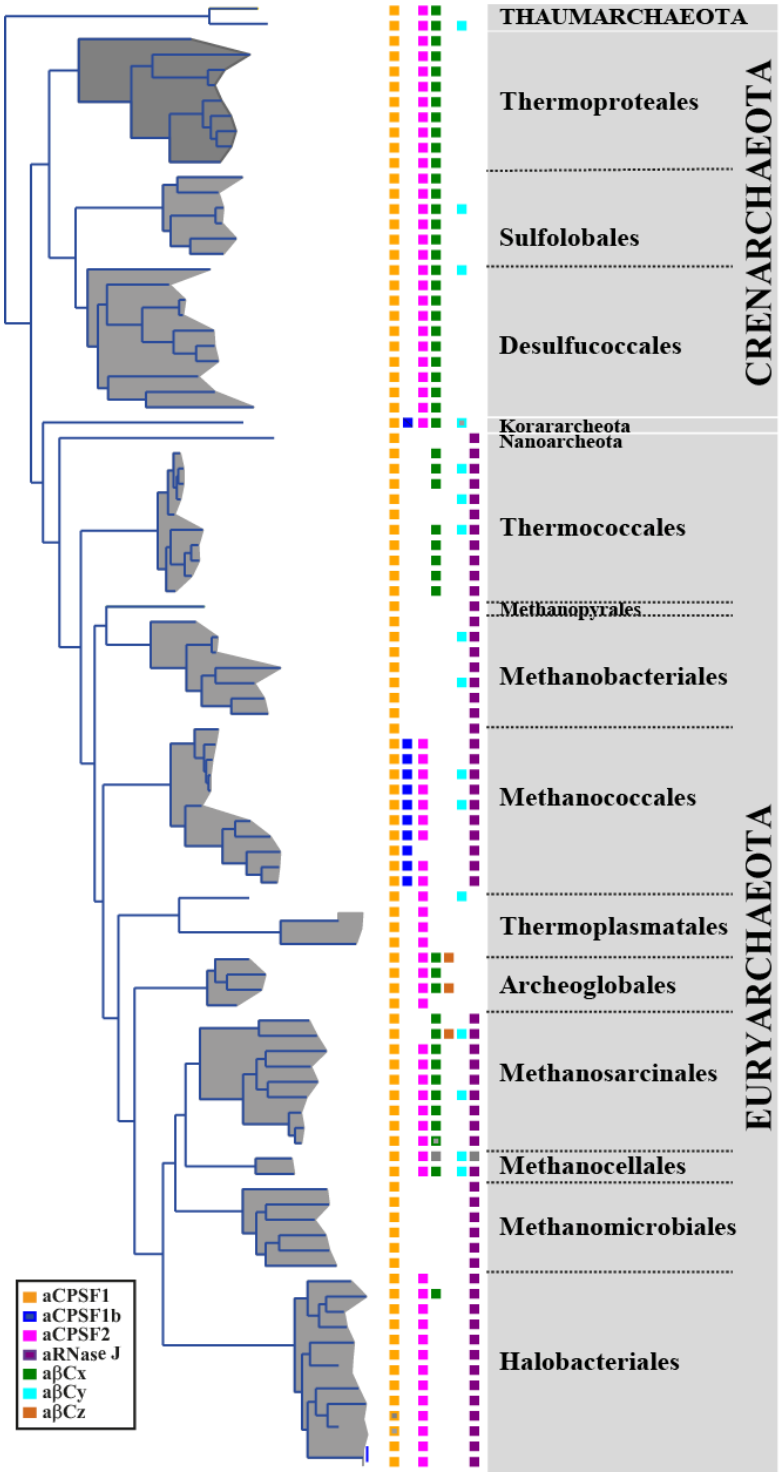

Figure S2 Repartition of the seven β–CASP clusters in archaeal phylogeny. Cluster colors are as in **Fig. S1**

SUPPLEMENTARY FIGURE S3

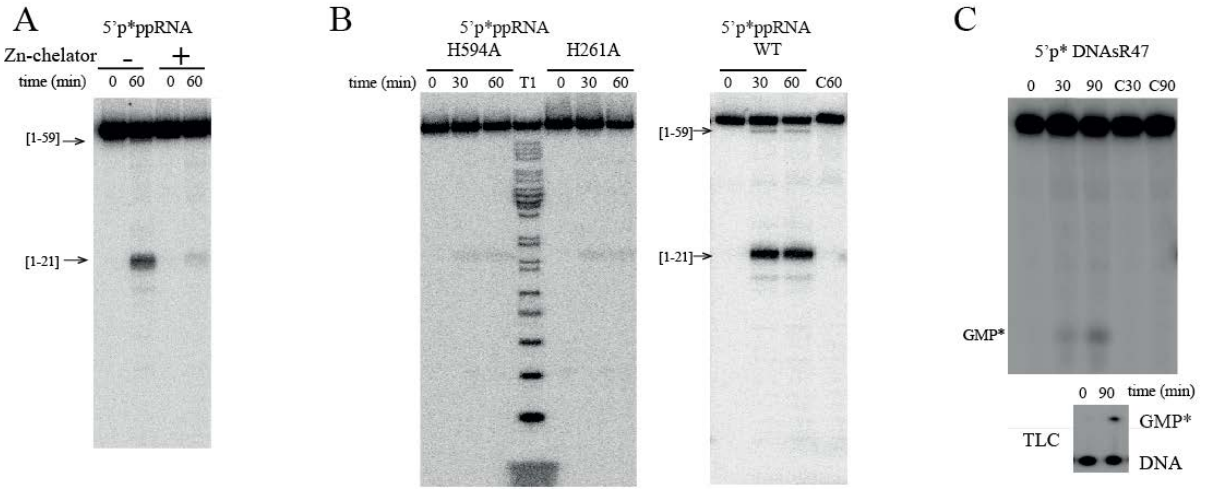

Figure S3. (*A*) Kinetic analysis of 5'p*ppRNA sR47 cleavage in presence (+) or in absence (-) of 10 mM 1,10-phenanthroline, a strong Zn chelator. (*B*) *In vitro* activity of Pab-aCPSF1H261A (H261A) and Pab-aCPSF1H594 (H594) variants. Time course analysis of RNA cleavage on 5'p*ppRNA substrate by protein variants (6 µM) at 65°C as indicated. See legend to Fig. 2 and 3 for symbols. (*C*) Kinetic analysis of of 5' end labeled DNA identical in sequence to sR47 RNA (5'p*DNA sR47). The products were analyzed on 10% PAGE and by thin layer chromatography (TLC) (bottom).

SUPPLEMENTARY FIGURE S4

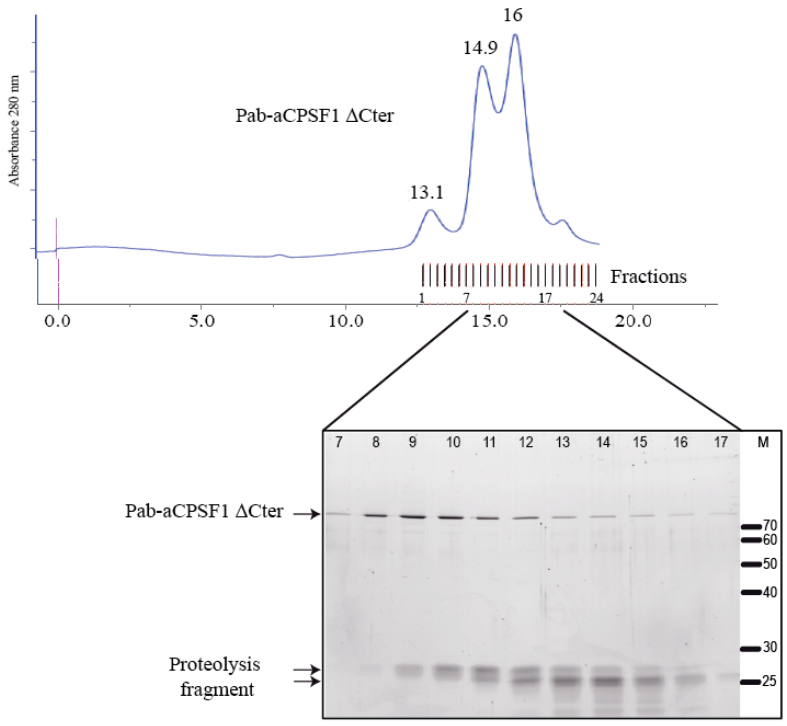

Figure S4. Identification of Pab-ΔCter protein proteolysis products. Fractions (7 to 17) of the size exclusion chromatography were separated on a 12% SDS-PAGE and sypro-orange stained. (M) stands for the protein marker (Page ruler unstained protein ladder, Thermo Scientific). The three discrete species indicated by arrows were excised and then sequenced by mass spectrometry (LC-ESI-MS/MS). The peptide sequences of the upper band overlap the entire amino sequence of Pab-aCPSF1ΔCter whereas the two lower bands match exclusively the N-terminal KH domain.