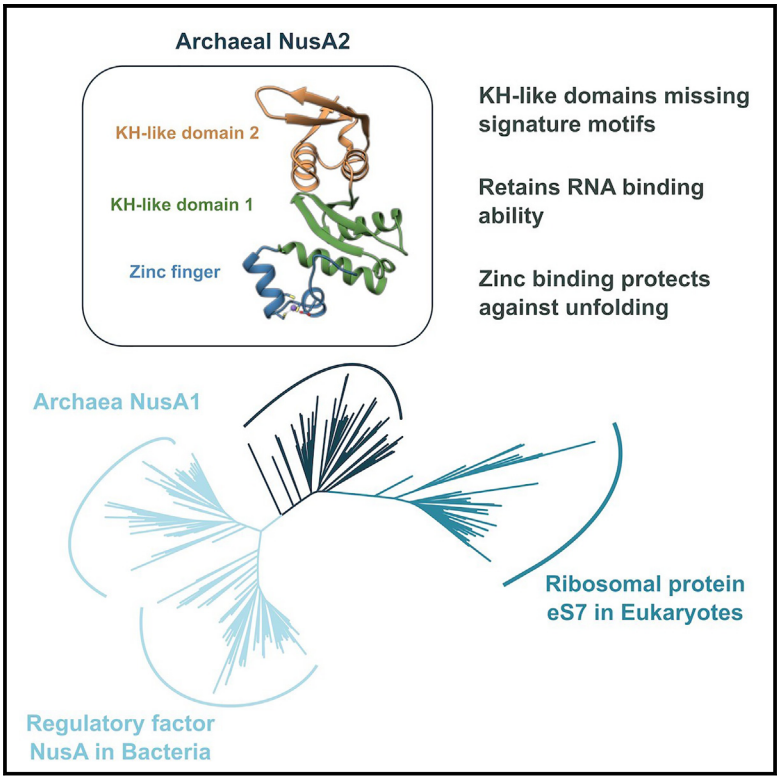
Structure

Archaeal NusA2 is the ancestor of ribosomal protein eS7 in eukaryotes

Graphical abstract



Highlights

- Archaea harbor two paralogues with homology to the KH domains of bacterial NusA
- Archaeal NusA2 KH-like domains lack key features of canonical type II KH domains
- The gene coding for eukaryotic ribosomal protein eS7 emerged from *nusA2*

Authors

Duy Khanh Phung, Simona Pilotto, Dorota Matelska, ..., Ladislav Hovan, Francesco Luigi Gervasio, Finn Werner

Correspondence

f.werner@ucl.ac.uk

In brief

An archaeal clade of proteins that Phung et al. termed NusA2 shows homology to the RNA-binding KH domains of bacterial regulatory factor NusA. The structures of two archaeal NusA2 proteins show two KH-like domains missing key features. The archaeal NusA2 clade appears to harbor the ancestor of eukaryotic ribosomal protein eS7.

Phung et al., 2025, Structure *33*, 1–11
January 2, 2025
https://doi.org/10.1016/j.str.2024.10.019

CellPress

Please cite this article in press as: Phung et al., Archaeal NusA2 is the ancestor of ribosomal protein eS7 in eukaryotes, Structure (2024), https://doi.org/10.1016/j.str.2024.10.019

Structure

CellPress
OPEN ACCESS

Article

Archaeal NusA2 is the ancestor of ribosomal protein eS7 in eukaryotes

Duy Khanh Phung,[1,6] Simona Pilotto,[1,6] Dorota Matelska,[1] Fabian Blombach,[1] Nikos Pinotsis,[2] Ladislav Hovan,[3] Francesco Luigi Gervasio,[3,4,5] and Finn Werner[1,7,*]

[1]RNAP Laboratory, Institute for Structural and Molecular Biology, University College London, London WC1E 6BT, UK
[2]Institute for Structural and Molecular Biology, Birkbeck College, London WC1E 7HX, UK
[3]Pharmaceutical Sciences, University of Geneva, 1206 Genève, Switzerland
[4]Institute of Pharmaceutical Sciences of Western Switzerland (ISPSO), University of Geneva, 1206 Genève, Switzerland
[5]Department of Chemistry, University College London, London WC1E 6BT, UK
[6]These authors contributed equally
[7]Lead contact
*Correspondence: f.werner@ucl.ac.uk
https://doi.org/10.1016/j.str.2024.10.019

SUMMARY

N-utilization substance A (NusA) is a regulatory factor with pleiotropic functions in gene expression in bacteria. Archaea encode two conserved small proteins, NusA1 and NusA2, with domains orthologous to the two RNA binding K Homology (KH) domains of NusA. Here, we report the crystal structures of NusA2 from *Sulfolobus acidocaldarius* and *Saccharolobus solfataricus* obtained at 3.1 Å and 1.68 Å, respectively. NusA2 comprises an N-terminal zinc finger followed by two KH-like domains lacking the GXXG signature. Despite the loss of the GXXG motif, NusA2 binds single-stranded RNA. Mutations in the zinc finger domain compromise the structural integrity of NusA2 at high temperatures and molecular dynamics simulations indicate that zinc binding provides an energy barrier preventing the domain from reaching unfolded states. A structure-guided phylogenetic analysis of the KH-like domains supports the notion that the NusA2 clade is ancestral to the ribosomal protein eS7 in eukaryotes, implying a potential role of NusA2 in translation.

INTRODUCTION

Archaea are prokaryotic organisms that use a eukaryotic RNA polymerase (RNAP) II-like transcription machinery to transcribe a bacterial-like genome organized in multicistronic operons with coupled transcription-translation.[1] The promoter elements, transcription initiation, elongation, and termination factors are also homologous to the RNAP II system, with B-recognition elements (BRE), TATA-boxes, TATA-binding protein (TBP), transcription factors TFB, TFE, TFS1 (homologous to TFIIB, TFIIE, and TFIIS), Spt4/5, Elf1, and aCPSF1 (archaeal cleavage and polyadenylation specificity factor 1).[2] The archaeal homologs of bacterial N-utilization substance A (NusA) factors are the "odd one out" as they are conserved between archaea and bacteria without any obvious eukaryotic counterparts. Bacterial NusA is a multidomain protein composed of an N-terminal domain (NTD) essential for the binding to the RNAP; three RNA-binding domains, S1, KH1, and KH2 (hnRNP K-Homology)[3,4]; and in some species, such as *Escherichia coli*, two regulatory acidic rich domains (AR1 and AR2) (Figure 1A). *E. coli* NusA is essential for cell viability, it enhances pausing of RNAP, facilitates antitermination jointly with other Nus factors,[5] and enables transcription-translation coupling[6,7] and rRNA processing.[8]

Archaeal genomes harbor genes annotated as *nusA* based on the protein sequence homology to the two KH domains. We refer to the proteins encoded by these genes hereafter as NusA1. Archaeal NusA1 comprise only the two KH domains and lack the NTD and S1 domain of bacterial NusA.[9] Archaeal RNAP subunit Rpo7, permanently associated with RNAP, comprises an S1 domain that is placed in a similar position as bacterial NusA S1 relative to the transcription elongation complex and might potentially be a hidden ortholog or a compelling case of parallel evolution.[10,11]

Many archaea, including Sulfolobales, encode a second gene that is distantly sequence-related to NusA1, lacking the GXXG motifs and containing an extra NTD (Figure 1A). Both the canonical archaeal *nusA1* and this unusual *nusA2* paralog are essential genes in *Sulfolobus islandicus*.[12] The NusA2 NTD includes four metal-chelating amino acids (three cysteines and one aspartate) predicted to form a short zinc finger. Zinc fingers are usually small domains where the zinc coordination stabilizes a specific fold, making them highly versatile as protein-protein, protein-nucleic acid, protein-lipid, and protein-small molecule interaction domains, and thus, they are highly pervasive in biology.[13–16] KH domains are nucleic acid-binding domains that are highly abundant and often present in multiple copies within a same protein. They are classified in 2-folds, KH type I that is dominant in eukaryotes, and type II mainly present in prokaryotes; both types share a minimal KH module that comprises the conserved GXXG motif.[17] KH domains show a preferential binding for single-stranded DNA or

Please cite this article in press as: Phung et al., Archaeal NusA2 is the ancestor of ribosomal protein eS7 in eukaryotes, Structure (2024), https://doi.org/10.1016/j.str.2024.10.019

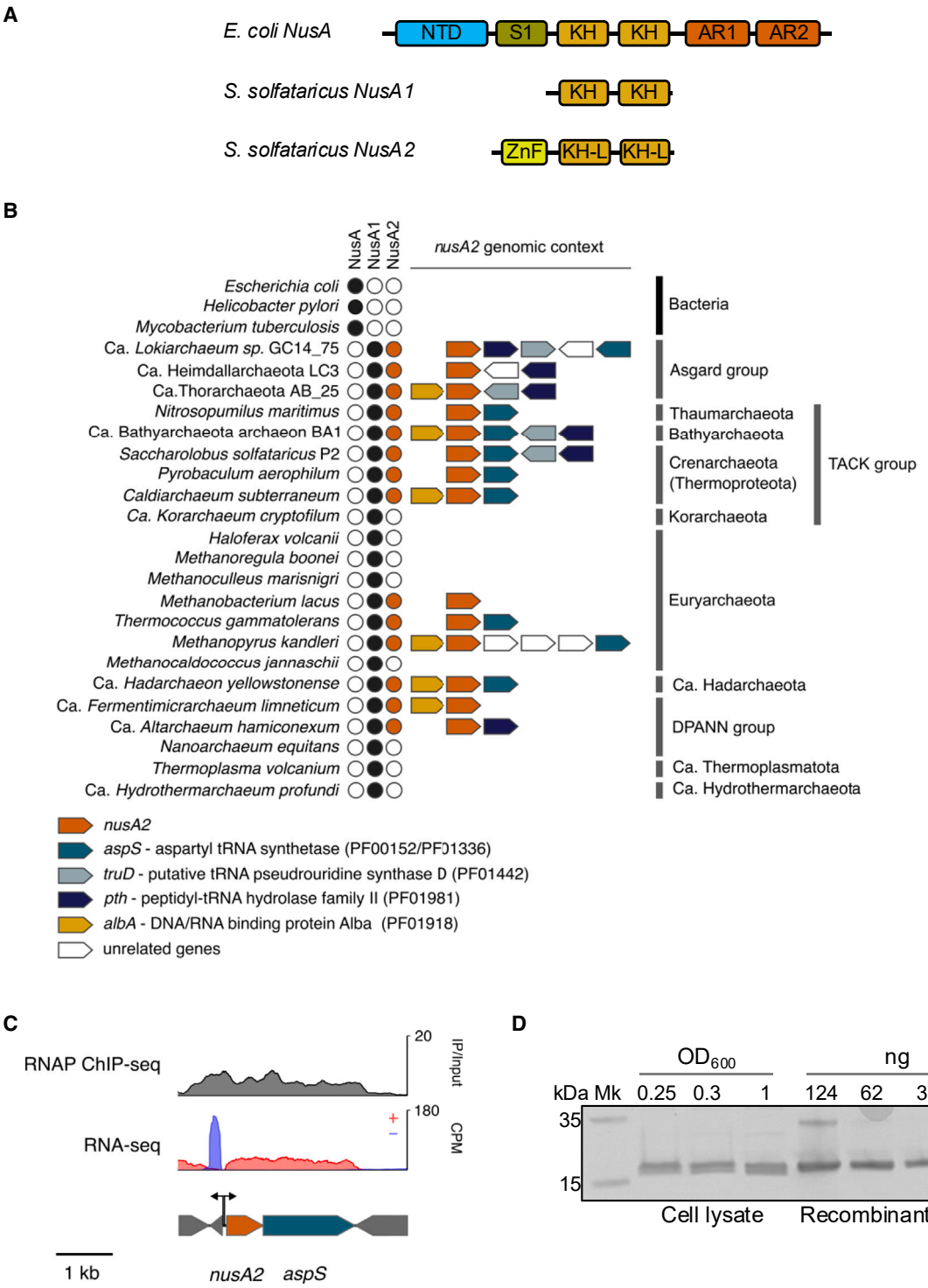


Figure 1. NusA2 is evolutionary conserved and constitutively expressed

(A) Domain composition of bacterial NusA, the archaeal NusA1 and NusA2. NTD, N-terminal domain; ZnF, zinc finger, KH, KH domain; KH-L, KH-like domain; AR, acidic C-terminal repeats.

(B) Schematic of NusAs gene distribution across bacteria, eukaryotes, and representative species from all archaea phyla. For the archaeal species having NusA2, the genomic context is shown color coded according to the legend.

(legend continued on next page)

Please cite this article in press as: Phung et al., Archaeal NusA2 is the ancestor of ribosomal protein eS7 in eukaryotes, Structure (2024), https://doi.org/10.1016/j.str.2024.10.019

RNA[18–20]; however, they can also facilitate protein-protein interactions, e.g., in the *E. coli* RNAP-NusA complex where the KH domains interact with the RNAP ω subunit,[21] or in the *E. coli* coupled RNAP-ribosome complexes where they interact with the ribosomal proteins uS2 and uS5.[6] Here, we show that *nusA2* gene is ancestral in archaea with wide phylogenomic distribution in all major taxonomic groups of archaea and a partially conserved genomic context. We solved the crystal structures of two NusA2 representatives from *Saccharolobus solfataricus* (Sso) and *Sulfolobus acidocaldarius* (Saci), which showed that the KH domains have alternative topologies to an extent where they are strictly speaking not KH domains any longer, but still able to bind single-stranded RNA *in vitro*. Biochemical analyses and molecular dynamics (MD) simulations elucidated how the zinc finger domain in Sulfolobales contributes to the thermostability of NusA2. Finally, phylogenomic analysis showed that the ancestral version of NusA2 evolved into the ribosomal protein eS7 in eukaryotes, firmly placing archaeal NusA factors at the intersection of transcription and translation.

RESULTS

NusA2 is evolutionary conserved and constitutively expressed

We identified *nusA2* orthologues in all major archaeal taxonomic groups. However, whereas *nusA1* is conserved in all archaeal species, *nusA2* gene loss occurred independently multiple times (Figure 1B). The wide phylogenetic distribution of *nusA2* is further supported by a partially conserved genomic context across all major taxonomic groups with the exception of Thermoplasmatota and possibly Hydrothermarchaeota (Figure 1B). The genes in the *nusA2* neighborhood include other genes with functions related to gene expression and RNA metabolism, coding for an aspartatyl-tRNA synthetase (*aspS*), *truD* family tRNA pseudouridine synthase, a tRNA peptidyl hydrolase, and a paralog of the RNA/DNA-binding protein Alba. We confirmed that *nusA2* is transcribed and translated in *S. solfataricus* P2. Available transcription start site mapping, RNAP chromatin-immunoprecipitation-sequencing (ChIP-seq) and RNA sequencing (RNA-seq) results showed that *nusA2 and aspS* are co-transcribed as a bicistronic operon in *S. solfataricus* (Figure 1C) with an estimated transcript abundance of ~900 transcripts per million.[2] Western blotting comparing recombinant NusA2 and cell lysate revealed that NusA2 is abundant during both exponential and stationary phase reaching ~0.2% of total protein (Figure 1D). Having ascertained the evolutionary conservation and the constitutively expression in *S. solfataricus*, we commenced a structural and functional analysis of NusA2.

The crystal structure of NusA2

To solve the structure of the NusA2, we successfully expressed and purified both paralogues from *S. solfataricus* and *S. acidocaldarius* recombinantly in *E. coli*, and we obtained crystals in a range of different precipitation conditions at 20° C in vapor diffusion, with Sso NusA2 best crystal diffracting at 1.68 Å of resolution, and Saci NusA2 at 3.1 Å (Figures S1A and S1B). Despite the high resolution obtained for Sso NusA2, we could not determine the structure by molecular replacement or ab initio programs. To resolve the phase problem, the dataset was collected using single-wavelength anomalous diffraction (SAD) to record the anomalous signal of the zinc ion. This was sufficient to determine the correct structure of the whole protein without any further modeling approach using SHELX[22] and ARP/wARP[23] programs (Table 1). The Sso NusA2 structure was subsequently used to resolve Saci NusA2 by molecular replacement[24] (Table 1; Figure S1C).

NusA2 has the expected modular composition of three domains, an N-terminal zinc finger followed by two $\alpha + \beta$ plates that we named KH-like domains as they miss key structural features of KH domains sensu stricto (Figure 2A). The NusA2 zinc finger is located on a loop connecting two short alpha helices and, in both species, it is coordinated by three cysteines and one aspartate (D8, C11, C18, and C21) (Figures 2B and S1C). However, it does not belong to any known zinc finger fold type as classified by Krishna and co-workers.[27] The two NusA2 $\alpha + \beta$ plates resemble the KH domains of *Aeropyrum pernix* (Ape) NusA1[9] (Figure S1D), but a deeper analysis of the secondary structure composition revealed that both domains have lost the KH signatures. KH domains have a well-defined and conserved topology, an $\alpha + \beta$ plate with three beta strands facing three alpha helices. The basic unit of the KH domain is the $\beta\alpha\alpha\beta$ module carrying the GXXG motif within the loop connecting the two central helices, where the position of the additional beta strand and alpha helix determines the classification of KH domains in type I and type II (Figure S1E).[17,28] Ape NusA1 has two type II KH domains; by contrast, in NusA2 the first KH-like domain lacks one helix (between the β 3 and the α 4 in Figure 2A) and the GXXG motif, while the C-terminal KH-like domain is still intact despite lacking the GXXG motif (Figures 1A and S1C–S1E). The Saci NusA2 structure showed little differences compared to Sso NusA2 except for the first two missing residues that caused a rearrangement of the first beta strand to ensure the correct binding of M1 into the hydrophobic pocket formed by L41, L42, L52, and Y57, stabilized by a salt bridge with E45, as seen also in Sso NusA2 (Figure S1C). In summary, NusA1 retained the structural signatures of bacterial NusA KH domains and its RNA-binding function, while NusA2 speciated. This involved the deterioration of both KH domains and the acquisition of the N-terminal zinc finger domain.

NusA2 is an RNA-binding protein

One of the key features of bacterial NusA proteins is their capacity to bind single-stranded (ss)RNA.[21,29] To test the nucleic acid binding properties of NusA2 *in vitro*, we used recombinant Sso NusA2 and carried out electrophoretic mobility shift analyses (EMSA) using a 26-nt ^{32}P-labeled ssRNA or ssDNA oligonucleotide probes. NusA2 formed a single complex with ssRNA in a dose-response

(C) RNAP chromatin immuno-precipitation sequencing (RNAP ChIP-seq) and RNA-sequencing data are shown for the *nusA2-aspS* bicistronic operon in *S. solfataricus* P2 strain. RNAP ChIP-seq data are normalized against the chromatin input (IP/input). RNA-seq signal was scaled to counts per million (CPM), showing in red the transcripts from the DNA template strand (indicated with the red plus symbol) and the non-template strand (the blue minus symbol).
(D) Western blot against Sso NusA2 showing the NusA2 expression levels in *S. solfataricus* at three growth points, two in the exponential phase and one (OD of 1) for stationary phase. The expression level is compared to the recombinantly expressed Sso NusA2 for a quantitative evaluation.

Please cite this article in press as: Phung et al., Archaeal NusA2 is the ancestor of ribosomal protein eS7 in eukaryotes, Structure (2024), https://doi.org/10.1016/j.str.2024.10.019

Table 1. Data collection and refinement statistics

Protein name PDB code	Sso NusA2 9F9U	Saci NusA2 9FJL
Data collection		
Space group	P 1 21 1	P 32 2 1
Wavelength	1.2826	0.9762
Unit cell parameters		
a, b, c (Å)	35.47, 72.81, 41.75	118.69, 118.69, 77.95
α, β, γ (°)	90.00, 111.95, 90.00	90.00, 90.00, 120.00
Resolution (Å)	72.88–1.68 (1.74–1.68)[a]	62.11–3.10 (3.31–3.10)
Unique reflections	21422 (1683)	11814 (2124)
CC1/2	1.00 (0.46)	1.00 (0.82)
R_{merge}	0.06 (1.086)	0.119 (0.267)
I/ δ (I)	16.6 (1.3)	16.2 (1.2)
Completeness (%)	95.3 (75.3)	100 (100)
Redundancy	6.7 (5.6)	11.6 (12.0)
Refinement		
Resolution (Å)	1.68	3.10
Reflections	20507	11799
R_{work}/R_{free}	0.2164/0.2356	0.2200/0.2445
No. atoms		
Protein (No. of residues)	170	168
Ligand/ion	1 Zn	1 Zn
Water	141	–
B-factors (Å)		
Protein	33.97	94.39
Ligand/ion	25.59	98.91
Water	40.11	–
R.m.s. deviations		
Bond lengths (Å)	0.008	0.011
Bond angles (°)	1.012	1.351
Ramachandran Plot		
Favored (%)	96.43	98.80
Outliers (%)	0.00	0.00
Rotamer outliers (%)	0.62	0.00
C-beta outliers (%)	0.00	0.00
Clashscore	6.95	11.91

[a]Values in parentheses are for highest-resolution shell.

fashion (Figure 2D). Unfortunately, the precipitation of Sso NusA2/RNA complexes in the electrophoresis gel well at higher protein concentrations made unfeasible to measure the affinity. NusA2 did not form any additional shifted bands in the EMSAs with ssDNA even at 5 μM protein concentrations. In conclusion, the archaeal NusA2 interacts with single-stranded RNA *in vitro* despite its highly divergent KH domains.

To examine the conservation of KH domain nucleic acid binding modes, we compared all KH domain structures available in the Protein DataBank (https://www.rcsb.org) in their free state and bound to their DNA or RNA targets. The search led to a final set of 33 structures from 24 different proteins including 26 type I KH domains, of which 4 bound to DNA and 22 to RNA, and 7 type II bound to RNA. Analyzing the structure-based sequence alignments, a conserved pattern of sequence similarity across all KH domains emerged, although with low sequence identity (Figure S2A). Despite the low sequence conservation, all KH domain structures are highly conserved, in particular around the nucleic acid binding site, and no conformational changes were observed upon DNA or RNA binding (Figure S2B). This makes it especially noteworthy that the NusA2 KH-like domain differ so much from the topology of both type I and II KH domains (Figure 2E), distorting the predicted nucleic acid binding site. In addition, both surface shape and electrostatic surface charge distribution, which contributes to the RNA binding, differs significantly from the *Mycobacterium tuberculosis* NusA KH domains (Figure 2F). The NusA2 paralog originated either by gene duplication and speciation of the ancestral *nusA* gene, or alternatively *nusA2* was introduced by lateral gene transfer. In contrast to NusA1 that preserved the KH type II topology and canonical RNA binding mode, NusA2 has evolved at an accelerated rate compared to NusA1.

The zinc finger ensures heat stability of NusA2

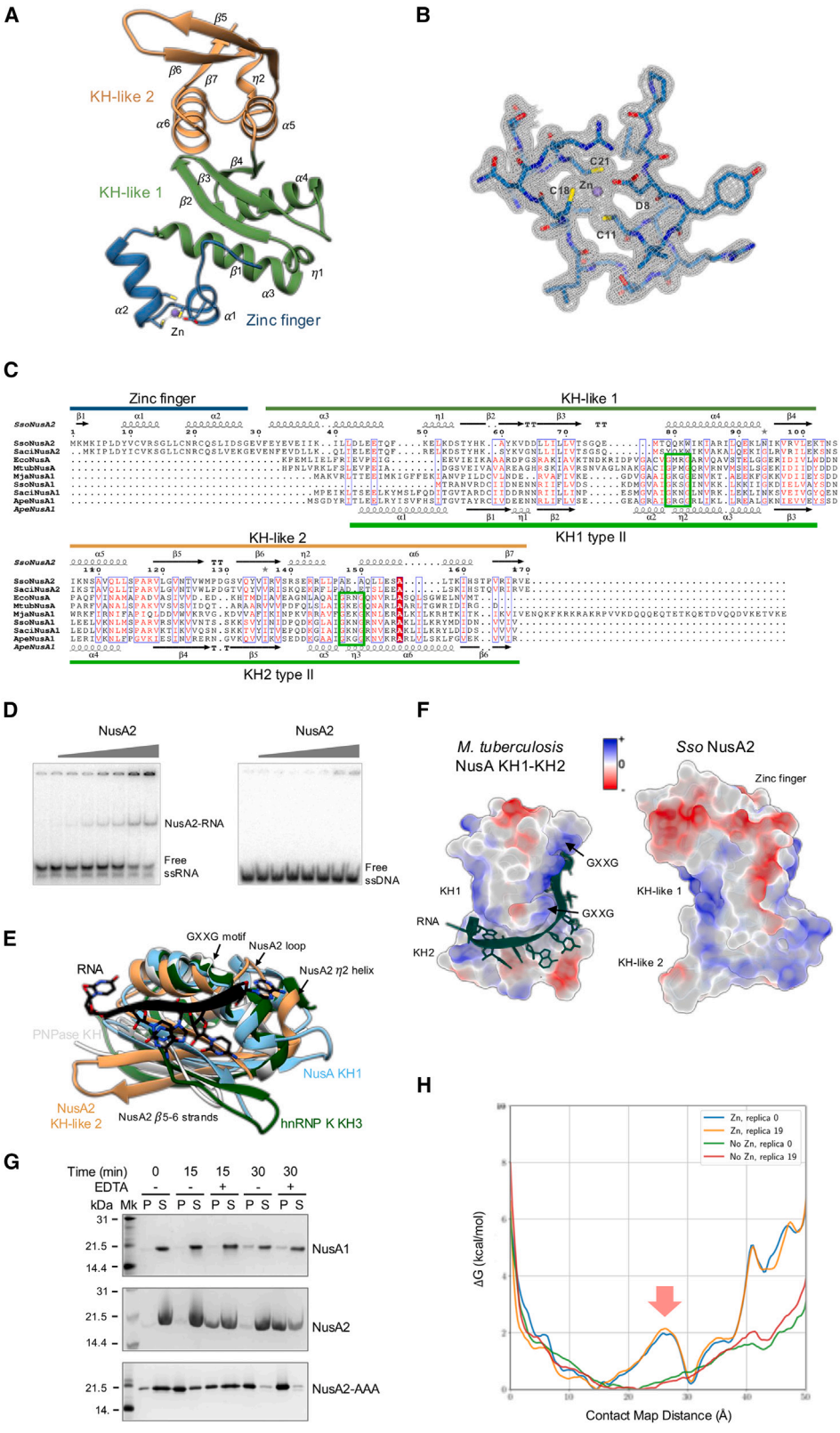
Given the extensive interface between the zinc finger and the KH-like 1 domains, we carried out thermal denaturation experiments to test the influence of the zinc finger on the structural integrity and solubility of Sso NusA2. NusA2 samples were incubated at 80°C for 15 or 30 min with or without the divalent metal chelator EDTA, and Sso NusA1, which has no zinc finger, was used as negative control (Figure 2G). Heat-stable proteins, expected in the soluble (Figure 2G, "S") fraction, were separated from the denaturated proteins in the precipitated ("P") fraction by centrifugation. As predicted, NusA1 was exclusively found in the supernatant even after extended incubation at 80°C and in good agreement with the high temperatures of the Sulfolobales habitat. Similarly, NusA2 was soluble at high temperatures, but sensitive to EDTA, which led to increasing precipitation after 15 and 30 min at 80°C. This suggests that the zinc coordination is important for the thermostability of NusA2. To validate this hypothesis, the three cysteines (C11, C18, and C21) were mutated to alanine residues creating the NusA2-AAA triple mutant. Already at room temperature, NusA2-AAA was partially precipitated, indicating that the lack of zinc binding compromised the structural integrity of NusA2 leading to aggregation and precipitation. Heat treatment of NusA2-AAA led to further precipitation but without any synergistic effects of EDTA. We furthermore prepared an N-terminal truncation variant by deleting residues 1–29 removing the zinc finger domain entirely. The resulting NusA2ΔZF variant could be expressed in *E. coli* but was unstable and prone to aggregation and precipitation even at ambient temperatures (data not shown). In summary, the NusA2 mutants corroborate the role of the zinc finger in stabilizing the NusA2 structure in Sulfolobales.

MD simulations unravel the stabilizing mechanism of the zinc finger domain

To shed light on the mechanism by which the NusA2 zinc finger enables the high thermostability of the protein, we performed enhanced sampling MD simulations and free energy calculations

Please cite this article in press as: Phung et al., Archaeal NusA2 is the ancestor of ribosomal protein eS7 in eukaryotes, Structure (2024), https://doi.org/10.1016/j.str.2024.10.019

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Phung et al., Archaeal NusA2 is the ancestor of ribosomal protein eS7 in eukaryotes, Structure (2024), https://doi.org/10.1016/j.str.2024.10.019

with parallel tempering metadynamics[30] starting from our Sso NusA2 crystal structure. The NusA2 zinc finger interacts closely with the KH-like 1 domain via polar/ionic interactions (Y9, E45, and the peptide backbone carbonyl group of L52) and nonpolar contacts (L41, L52, and Y57) (Figure S2C). As collective variables for the free energy calculations, we used contact maps for the zinc finger and KH1 domains (Figure 2H), and the distance of two key residues, M1 and E45 (Figure S2D). We used the DES-Amber v1.0 forcefield[31] for the protein and water molecules, and *de novo* parametrization using MCPB.py[32] and Gaussian[33] for the zinc atom. With respect to the modeling of the zinc interactions, due to the atomistic force-field formulation, we are mimicking the coordination bonds with simple atom-centered electrostatics, thus the stabilization provided by the zinc might be lower than that provided by the real coordination bonds. Still, the computed free energy landscapes showed a clear difference between the MD simulations of NusA2 in the presence and absence of zinc, and the fully unfolded states are higher in energy. The contact map distance of zero in Figure 2H corresponds to the starting structure and the maximum value to the fully unfolded domain (see also Videos S1 and S2). The most important contribution of the zinc ion is the significant barrier that arises in the main minimum (Figure 2H, highlighted with a red block arrow) that would prevent the domain from easily reaching misfolded states in the presence but not in the absence of zinc. States with the same free energy have the same probability of occurring, meaning that in a large flat minimum like that shown in the absence of zinc, NusA2 is likely to be unfolded as there does not appear to be a barrier to prevent this process from happening. Although unfolded states with favorable free energy still exist in the presence of zinc at approximately 30 Å in contact map distance, the barrier makes the protein unlikely to reach it from the native structure. The combination of the observed barrier and stabilization of the folded basin could justify the observed higher thermal stability in the presence of zinc.

Computational prediction of NusA2 interaction binding sites identifies two hotspots

The molecular structure of NusA2 allowed us to exploit computational methods to predict surface exposed regions as potential binding sites for protein/protein interactions. We combined the results of seven different webserver, most of which have been developed to determine the residues exposed on the surface, derived directly from the PDB coordinates, as well as their chemical-physical properties and, using tailored algorithms or docking approaches, provide a per-residue output with scores based on the probability of them being part of an interface. The idea to use more predictors relies not only on the likelihood of increasing the precision of the prediction,[34] but also taking into account that NusA2 has low sequence identity and unique structural features compared to NusA1. In fact, some of these programs use sequence and structural homologies to increase the accuracy that may bias the outcome in our specific case. All programs, although identifying different residues as best candidates (Table S1), predicted two patches on NusA2 interface for an intermolecular interaction at opposite ends of the structure with almost equal probability (Figure 3A). The first patch identifies the zinc finger domain as likely interaction site as expected, and the second patch is located on the KH-like domain 2. The highest scoring residues (N123, T124, V125, W126, M127, and P128) are all located on the same most exposed beta strand. In summary, NusA2 has two likely interaction hotspots at different ends of the molecule, which could enable the recruitment of two different proteins or stabilize a higher-order complex.

NusA2 is ancestral to ribosomal protein eS7

The results obtained from the interfaces prediction prompted us to carry out a structure-based search using DALI[37] to identify homologous proteins that could inform us about NusA2 evolution and potential binding partners. From the PDB50 output, we kept the

Figure 2. The NusA2 structure reveals highly divergent KH-like domains

(A) Crystal structure of *S. solfataricus* NusA2 in cartoon representation. The zinc finger is shown in blue, the KH-like 1 and 2 domains in green and light orange, respectively, the zinc ion is shown in medium purple. All alpha helices and beta strands are numbered from the N- to the C-terminus in sequential order; the 3_{10}-helix type is indicated with the η symbol.

(B) Close up of the electron density map of the zinc finger domain with the coordinating residues shown in stick representation, color coded as in panel A. The density map was prepared with ccp4mg in gray mesh style and a contour level of 0.4090 electrons/A3, clipped with a radius of 1.5 A around the zinc finger domain (excluding water molecules).

(C) Multiple sequence alignment of NusA factors: *S. solfataricus*, and *S. acidocaldarius* NusA2, *E. coli*, *Mycobacterium tuberculosis* NusA, *S. solfataricus*, *S. acidocaldarius*, *Methanocaldococcus jannaschiii*, and *A. pernix* NusA1. On the top and bottom of the MSA is a schematic representation of the secondary structure elements of Sso NusA2 and Ape NusA1[9] with helices shown as springs and strands as arrows, as well as the corresponding domain organization, for NusA2 colored according to the structure in panel A.

(D) EMSAs showing the differential migration of a 26-nt radiolabeled ssRNA or ssDNA probe after incubation with increasing concentrations of Sso NusA2 (0.1, 0.25, 0.5, 0.75, 1, 2.5, and 5 μM).

(E) Superposition of three representative KH domains, one for each sub-type, with the KH-like 2 domain of NusA2, in light orange. hnRNP K KH3[25] is a DNA-binding type I KH domain, PNPase KH[26] is an RNA-binding type I, and NusA KH1[9] is an RNA-binding type II. All structures are shown in cartoon representation; the nucleic acids are shown in black ribbon with the bases in stick. The main outliers in NusA2 KH-like 2 folding are highlighted as well as the GXXG motif in the KH-containing proteins. The complete list of protein names, PDB codes, and species is reported in the STAR Methods.

(F) Electrostatic surface potential of the KH1 and 2 domains from *M. tuberculosis* NusA-RNA complex (PDB code 2atw) and structurally aligned corresponding KH-like domains in *S. solfataricus* NusA2. The surface is shown with 20% transparency with its structure in cartoon representation. The range depicted is −10 to +10 kcal*mol^{-1}*e^{-1}.

(G) SDS-PAGE showing the heat stability of Sso NusA1, as control, Sso NusA2, and NusA2 C11A-C18A-C21A mutant (NusA2-AAA). Both precipitated (P) and soluble (S) fractions were loaded on the gel, after incubation at 80°C for 0, 15, and 30 min, and addition of 100 mM EDTA.

(H) Comparison of the 1D free energy profiles obtained from the PT-metadynamics simulations of 20 replicas reweighted on the contact map distance from the native state of the zinc finger and KH-like 1 domains. The red arrow indicates the position of a free energy barrier when the zinc atom was present in the simulation. Contact map distance of zero corresponds to the starting structure, maximum value to fully unfolded domain.

Please cite this article in press as: Phung et al., Archaeal NusA2 is the ancestor of ribosomal protein eS7 in eukaryotes, Structure (2024), https://doi.org/10.1016/j.str.2024.10.019

CellPress
OPEN ACCESS

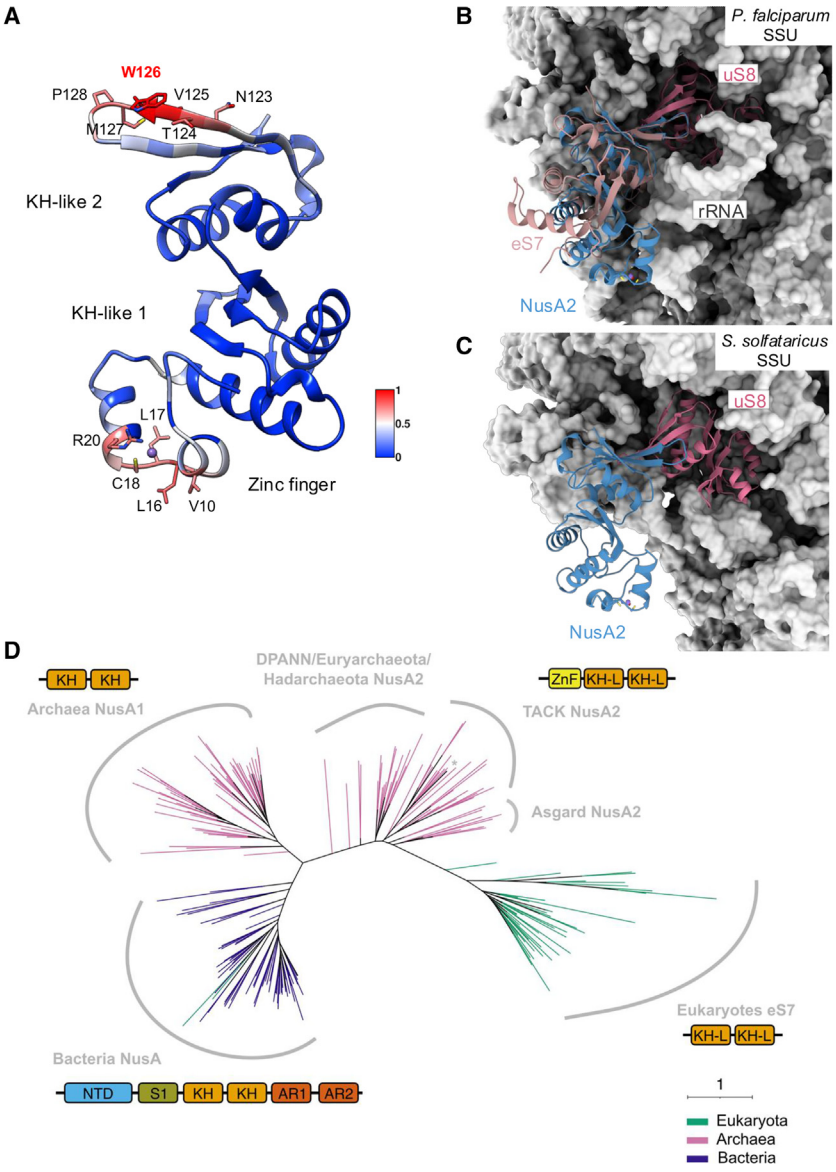


Figure 3. NusA2 evolved into ribosomal protein eS7 in eukaryotes

(A) Protein-protein interface prediction obtained combining seven webservers and normalized between 0, lowest probability, and 1, highest probability of a residue being part of an interface. The prediction output is visualized in a color gradient from blue to red. The residues with a score higher than 0.7 are shown in sticks on the cartoon structure. The per-residue output is reported in Table S1.

(B) Superimposition of Sso NusA2, in blue, with eS7, in pink, in the context of the *Plasmodium falciparum* 40S subunit (PDB code 3j7a[35]) which is shown in surface representation with the exception of uS8 that is displayed as dark pink cartoon and eS7.

(C) Model of the hypothetical Sso NusA2 binding site on the ribosome. The model was obtained by using the superposition shown in panel B and superimposing the Sso 30S subunit (PDB code 6tmf[36]), as gray surface, with uS8 of the *P. falciparum* 40S subunit. For clarity, the *P. falciparum* 40S subunit is not shown. Note that the eukaryote-specific rRNA stem-loop is missing in archaea ribosomes. Subunit and proteins are shown and color coded as in (B). Source data are available at https://doi.org/10.5281/zenodo.13694103.

(D) Unrooted maximum-likelihood phylogenetic tree of NusA, NusA1, NusA2 and eS7 KH/KH-like domains. Terminal branches are colored according to the taxonomic domains (green for eukaryotes, pink for archaea, and purple for bacteria). Representative domain architectures are shown next to the major groups. The asterisk marks the position of *S. solfataricus* NusA2.

first 20 top scoring structures (reported in Table S2) and pruned them to remove (1) outliers with a root-mean-square deviation (RMSD) score higher than 5 Å, (2) orthologs with lower *Z* score, and (3) proteins with a sequence identity lower than 10%. The pruning led us to 6 structures (Marked with asterisks in Table S2), among which only three shared features with both NusA2 KH-like domains, and no hits for the zinc finger, which appears to be a unique structural feature of archaeal NusA2. The three proteins were the *Plasmodium falciparum* ribosomal protein eS7[35] (top-scoring with a *Z* score of 11.2 and 17% sequence identity), the *M. tuberculosis* NusA[38] (*Z* score 9.4) and *A. pernix* NusA1[9] (*Z* score 7.0). The superimposition of NusA2 and eS7 demonstrated an extensive degree of structural conservation between the KH-like 2 of NusA2 and the equivalent domain in eS7 (Figure 3B). The two C-terminal beta strands, predicted as a potential interface (Figure 3A), perfectly overlap with their counterpart in eS7, and form a beta-addition by extending the beta sheet with uS8 via a well conserved patch of hydrophobic residues in the context of eS7 in the 40S ribosomal subunit (Figures S3A and S3B). Residue W126 of NusA2, which had the highest score in the interaction predictor analysis (Figure 3A, highlighted in red), is highly conserved as is the hydrophobic patch in the eS7 interaction partner uS8 (Figure S3B). Moreover, superimposing the generated NusA2/40S model into the Sso 30S ribosomal subunit[36] revealed a perfect fit without any sterical clash (Figure 3C). Finally, in this last model, the NusA2 zinc finger is solvent exposed available to interact with other partners. These findings suggest that NusA2 may be transiently associated with the small ribosomal subunit in archaea. To probe for NusA2 binding partners and the incorporation of NusA2 into larger complexes, we size-fractionated cell lysates on a sucrose gradient and detected NusA2 by western blotting (Figure S3C). NusA2 eluted in the low molecular weight fractions indicating that, at least during exponential growth, NusA2 is predominantly present in its free, unbound form in the cell.

To further explore the evolutionary landscape of archaeal NusA and NusA-like factors in all domains of life, we calculated maximum-likelihood phylogenetic trees based on conserved residues of their KH and KH-like domains. The trees revealed three major clades with distinct taxonomic distributions as follows: (1) bacterial NusA, (2) archaeal NusA1, (3) archaeal NusA2, with eukaryotic eS7 originating from the NusA2 clade (Figure 3D). The

Please cite this article in press as: Phung et al., Archaeal NusA2 is the ancestor of ribosomal protein eS7 in eukaryotes, Structure (2024), https://doi.org/10.1016/j.str.2024.10.019

position of the eS7 within the larger NusA2/eS7 clade, branching off from Asgard NusA2 (90% bootstrap confidence), is well in line with the current understanding of eukaryotic evolution from archaea,[39] and suggests that NusA2 has evolved into a ribosomal protein and stably incorporated in eukaryotic ribosomes.

DISCUSSION

Archaea encode a small protein, NusA2, that is distantly related to the KH domains of the bacterial transcription elongation factor NusA. Despite its evolutionary conservation and abundance in the cell (Figure 1), and the essential nature of *nusA2* for *Sulfolobus* viability, the biological function of NusA2 is unknown. Here, we break open the field with a structural, biochemical and computational analysis of *Sulfolobus* NusA2.

The crystal structure revealed that NusA2 has two highly divergent KH-like domains that, despite sequence related to bacterial and archaeal NusAs, have lost key KH domain signatures (Figure 2A). However, despite the loss of the canonical RNA-binding GXXG motifs, NusA2 retains the ability to bind ssRNA (Figure 2D). The KH domains of bacterial NusA interact with the *nut* site in the rRNA leader regions, which plays a role in transcription antitermination and possibly also in the role of NusA as a chaperone in rRNA processing.[8,18,29] However, the *nut* sites are not conserved in archaea,[40] and antitermination is likely facilitated by idiosyncratic factors and mechanisms. Archaeal NusA factors appear as streamlined versions of their bacterial homologs, and reduced to two KH domains, but NusA2 has made up for this by gaining an additional N-terminal zinc finger domain, indicative of alternative functions. The zinc finger contributes to the high thermostability of NusA2, and its disruption by mutagenesis or metal chelation compromises the structure of the protein (Figures 2G and 2H). Enhanced sampling MD simulations of the zinc-bound NusA2 provided a plausible underlying mechanism, as metal coordination leads to a significant energetic barrier that would prevent the domain from easily reaching misfolded states. In the absence of zinc, the MD simulations indicate that this barrier is absent, the protein unfolding commences from the N-terminus and the folded states are less stable with respect to the unfolded ensemble. Zinc fingers are well characterized interaction domains facilitating the binding of proteins, nucleic acids and other ligands.

Both sequence and structural analyses strongly indicate that NusA2 and eS7 share a common ancestor. Moreover, the eS7 binding site is well conserved in archaea, opening the possibility that NusA2 may conditionally interact with the *Sulfolobus* ribosome in a similar fashion as eS7 in eukaryotes (Figures 3C, S3A, and S3B). The RNA components of archaeal ribosomes (16S, 23S, and 5S) are closer related to bacteria than eukaryotes by lacking several rRNA extensions and the 5.8S rRNA entirely, but the complement of ribosomal proteins is eukaryote-like.[41–43] Surprisingly, several archaeal ribosomal proteins show promiscuous binding patterns on the 30S and 50S ribosome subunits, and some can be lost during fractionation and enrichment.[44,45] eS7 is a regulatory component of the 40S ribosomal subunit in eukaryotes that is present in mono- or poly-ubiquitinylated states in the cell. Ubiquitination of ribosomal proteins plays an important role in mRNA surveillance and ribosome-associated quality control (RQC). However, the ubiquitination of eS7 plays a specific role in the endoplasmic reticulum stress response to protein misfolding, where it has been shown to counteract protein aggregation and increase the translation efficiency,[46–49] challenges that Sulfolobales certainly encounter in their hot environmental niche. Our size fractionation analysis did not document a permanent association of NusA2 with *Sulfolobus* ribosomes, but we cannot rule out that either NusA2 was lost during purification or alternatively that the stable binding of NusA2 is context-dependent. For example, archaeal Mbf1, homolog to the eukaryotic ribosome collision factor EDF1, binds to the ribosome in a transient manner such that no stable interaction with the 30S ribosomal subunit could be detected in fractionated cell lysates, but the majority of Mbf1 was found bound to actively translating ribosomes when chemical crosslinking was applied.[50]

NusA is evolutionary ancient and was present in the last universal common ancestor (LUCA). During early evolution of the archaeal branch of life, the NusA ancestor duplicated (Figure 3D). While NusA1 kept the sequence and structural traits of canonical NusA KH domains, NusA2 speciated. Due to the relaxed selection pressure, NusA2 evolved at accelerated rates that led to the loss of key KH domain signatures, including the GXXG motifs, and gain of a zinc finger domain. Later in eukaryotes, NusA2 likely lost the zinc finger again and became stably incorporated into ribosomes as eS7. This eventually enhanced the co-translational protein folding properties of the ribosome, considering the role of eS7 to counter ER stress. Overall, this event is analogous to a "transcription factor capture" during which a transiently associated regulatory factor becomes stably incorporated into RNAP as it provides a selective advantage for transcription becoming a bona fide RNAP subunit. This happened for TFIIE and TFIIF in eukaryotes, which transiently engage with RNAPII but have become stably integrated into RNAPI (as A49/A34.5, yeast RNAP subunit nomenclature) and III (as C53/C37 and C82/C34).[1,51] Only here, the transcription factor is recruited to the ribosome instead. Future work will focus on probing the association of NusA2 to ribosomes in response to a range of growth and stress conditions and focusing on interactors with the NusA2-specific zinc finger domain.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Finn Werner (f.werner@ucl.ac.uk).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The crystal structures determined in the current study are available in the Protein Data Bank repository, https://www.rcsb.org/, with the following PDB codes: 9F9U for *S. solfataricus* NusA2, and 9FJL for *S. acidocaldarius* NusA2, and are publicly available as of the date of publication. Accession codes are also listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

Research in the RNAP laboratory at UCL in London is funded by a Wellcome Investigator Award in Science "Mechanisms and Regulation of RNAP

transcription" to F.W. (WT 207446/Z/17/Z), and BBSRC grant BB/X017028/1. Research in Francesco Gervasio's team in Geneva is funded by the Swiss National Science Foundation and Bridge, project numbers: 200021_204795, CRSII5_216587, and 40B2−0_203628. F.L.G. and L.H. acknowledge the Swiss National Supercomputing Centre (CSCS) for generous supercomputer time allocations projectID: s1228. The authors would like to thank Diamond Light Source for beamtime (proposal mx17201), and the staff of beamline I03 for assistance with crystal data collection. We do also acknowledge DESY (Hamburg, Germany), a member of the Helmholtz Association HGF, for the provision of experimental facilities. The data collection was carried out at PETRA III and we would like to thank all staff for assistance in using beamline P14.

AUTHOR CONTRIBUTIONS

Conceptualization, F.W.; investigation, D.K.P., S.P., L.H., D.M., F.B., and N.P.; validation, D.K.P., S.P., L.H., D.M., and F.B.; formal analysis, D.K.P., S.P., L.H., D.M., and F.B.; writing – original draft, F.W., D.K.P., S.P., D.M., F.L.G., and L.H.; writing – review & editing, F.W., D.K.P., S.P., and F.B.; visualization, D.K.P., S.P., F.B., L.H., D.M., and N.P.; funding acquisition, F.W. and F.L.G.; resources, F.W. and F.L.G.; supervision, F.W. and F.L.G.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Cloning strategy, expression and purification
  - RNA polymerase Chromatin immunoprecipitation sequencing and RNA-seq data analysis
  - Heat stability test
  - Western blotting
  - Electrophoretic mobility shift assay
  - Protein crystallization, data collection and processing
  - KH sequence- and structure-based searches
  - Protein-protein interface prediction
  - Structure-based analysis
  - Sequence alignments
  - Phylogenetic analysis
  - Genome context analysis
  - Cell lysate sucrose gradient fractionation
  - Molecular dynamics simulations
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.str.2024.10.019.

Received: July 5, 2024
Revised: September 6, 2024
Accepted: October 10, 2024
Published: November 5, 2025

REFERENCES

1. Werner, F., and Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. Nat. Rev. Microbiol. *9*, 85–98. https://doi.org/10.1038/nrmicro2507.
2. Blombach, F., Fouqueau, T., Matelska, D., Smollett, K., and Werner, F. (2021). Promoter-proximal elongation regulates transcription in archaea. Nat. Commun. *12*, 5524. https://doi.org/10.1038/s41467-021-25669-2.
3. Craven, M.G., and Friedman, D.I. (1991). Analysis of the Escherichia coli nusA10(Cs) allele: relating nucleotide changes to phenotypes. J. Bacteriol. *173*, 1485–1491. https://doi.org/10.1128/jb.173.4.1485-1491.1991.
4. Nakamura, Y., Mizusawa, S., Tsugawa, A., and Imai, M. (1986). Conditionally lethal nusAts mutation of Escherichia coli reduces transcription termination but does not affect antitermination of bacteriophage lambda. Mol. Gen. Genet. *204*, 24–28. https://doi.org/10.1007/BF00330182.
5. Greive, S.J., Lins, A.F., and von Hippel, P.H. (2005). Assembly of an RNA-protein complex. Binding of NusB and NusE (S10) proteins to boxA RNA nucleates the formation of the antitermination complex involved in controlling rRNA transcription in Escherichia coli. J. Biol. Chem. *280*, 36397–36408. https://doi.org/10.1074/jbc.M507146200.
6. Wang, C., Molodtsov, V., Firlar, E., Kaelber, J.T., Blaha, G., Su, M., and Ebright, R.H. (2020). Structural basis of transcription-translation coupling. Science *369*, 1359–1365. https://doi.org/10.1126/science.abb5317.
7. Webster, M.W., Takacs, M., Zhu, C., Vidmar, V., Eduljee, A., Abdelkareem, M., and Weixlbaumer, A. (2020). Structural basis of transcription-translation coupling and collision in bacteria. Science *369*, 1355–1359. https://doi.org/10.1126/science.abb5036.
8. Bubunenko, M., Court, D.L., Al Refaii, A., Saxena, S., Korepanov, A., Friedman, D.I., Gottesman, M.E., and Alix, J.H. (2013). Nus transcription elongation factors and RNase III modulate small ribosome subunit biogenesis in Escherichia coli. Mol. Microbiol. *87*, 382–393. https://doi.org/10.1111/mmi.12105.
9. Shibata, R., Bessho, Y., Shinkai, A., Nishimoto, M., Fusatomi, E., Terada, T., Shirouzu, M., and Yokoyama, S. (2007). Crystal structure and RNA-binding analysis of the archaeal transcription factor NusA. Biochem. Biophys. Res. Commun. *355*, 122–128. https://doi.org/10.1016/j.bbrc.2007.01.119.
10. Belogurov, G.A., and Artsimovitch, I. (2015). Regulation of Transcript Elongation. Annu. Rev. Microbiol. *69*, 49–69. https://doi.org/10.1146/annurev-micro-091014-104047.
11. Fouqueau, T., Blombach, F., Cackett, G., Carty, A.E., Matelska, D.M., Ofer, S., Pilotto, S., Phung, D.K., and Werner, F. (2018). The cutting edge of archaeal transcription. Emerg. Top. Life Sci. *2*, 517–533. https://doi.org/10.1042/ETLS20180014.
12. Zhang, C., Phillips, A.P.R., Wipfler, R.L., Olsen, G.J., and Whitaker, R.J. (2018). The essential genome of the crenarchaeal model Sulfolobus islandicus. Nat. Commun. *9*, 4908. https://doi.org/10.1038/s41467-018-07379-4.
13. Misra, S., Miller, G.J., and Hurley, J.H. (2001). Recognizing phosphatidylinositol 3-phosphate. Cell *107*, 559–562. https://doi.org/10.1016/s0092-8674(01)00594-3.
14. Sievers, Q.L., Petzold, G., Bunker, R.D., Renneville, A., Słabicki, M., Liddicoat, B.J., Abdulrahman, W., Mikkelsen, T., Ebert, B.L., and Thomä, N.H. (2018). Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN. Science *362*, eaat0572. https://doi.org/10.1126/science.aat0572.
15. Lu, D., Searles, M.A., and Klug, A. (2003). Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. Nature *426*, 96–100. https://doi.org/10.1038/nature02088.
16. Krois, A.S., Ferreon, J.C., Martinez-Yamout, M.A., Dyson, H.J., and Wright, P.E. (2016). Recognition of the disordered p53 transactivation domain by the transcriptional adapter zinc finger domains of CREB-binding protein. Proc. Natl. Acad. Sci. USA *113*, E1853–E1862. https://doi.org/10.1073/pnas.1602487113.
17. Grishin, N.V. (2001). KH domain: one motif, two folds. Nucleic Acids Res. *29*, 638–643. https://doi.org/10.1093/nar/29.3.638.

18. Beuth, B., Pennell, S., Arnvig, K.B., Martin, S.R., and Taylor, I.A. (2005). Structure of a Mycobacterium tuberculosis NusA-RNA complex. EMBO J. *24*, 3576–3587. https://doi.org/10.1038/sj.emboj.7600829.

19. Siomi, H., Choi, M., Siomi, M.C., Nussbaum, R.L., and Dreyfuss, G. (1994). Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. Cell *77*, 33–39. https://doi.org/10.1016/0092-8674(94)90232-1.

20. Braddock, D.T., Louis, J.M., Baber, J.L., Levens, D., and Clore, G.M. (2002). Structure and dynamics of KH domains from FBP bound to single-stranded DNA. Nature *415*, 1051–1056. https://doi.org/10.1038/4151051a.

21. Guo, X., Myasnikov, A.G., Chen, J., Crucifix, C., Papai, G., Takacs, M., Schultz, P., and Weixlbaumer, A. (2018). Structural Basis for NusA Stabilized Transcriptional Pausing. Mol. Cell *69*, 816–827.e4. https://doi.org/10.1016/j.molcel.2018.02.008.

22. Sheldrick, G.M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr. D Biol. Crystallogr. *66*, 479–485. https://doi.org/10.1107/S0907444909038360.

23. Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. *6*, 458–463. https://doi.org/10.1038/8263.

24. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674. https://doi.org/10.1107/S0021889807021206.

25. Braddock, D.T., Baber, J.L., Levens, D., and Clore, G.M. (2002). Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: solution structure of a complex between hnRNP K KH3 and single-stranded DNA. EMBO J. *21*, 3476–3485. https://doi.org/10.1093/emboj/cdf352.

26. Hardwick, S.W., Gubbey, T., Hug, I., Jenal, U., and Luisi, B.F. (2012). Crystal structure of Caulobacter crescentus polynucleotide phosphorylase reveals a mechanism of RNA substrate channelling and RNA degradosome assembly. Open Biol. *2*, 120028. https://doi.org/10.1098/rsob.120028.

27. Krishna, S.S., Majumdar, I., and Grishin, N.V. (2003). Structural classification of zinc fingers: survey and summary. Nucleic Acids Res. *31*, 532–550. https://doi.org/10.1093/nar/gkg161.

28. Valverde, R., Edwards, L., and Regan, L. (2008). Structure and function of KH domains. FEBS J. *275*, 2712–2726. https://doi.org/10.1111/j.1742-4658.2008.06411.x.

29. Arnvig, K.B., Pennell, S., Gopal, B., and Colston, M.J. (2004). A high-affinity interaction between NusA and the rrn nut site in Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. USA *101*, 8325–8330. https://doi.org/10.1073/pnas.0401287101.

30. Barducci, A., Bussi, G., and Parrinello, M. (2008). Well-tempered metadynamics: a smoothly converging and tunable free-energy method. Phys. Rev. Lett. *100*, 020603. https://doi.org/10.1103/PhysRevLett.100.020603.

31. Piana, S., Robustelli, P., Tan, D., Chen, S., and Shaw, D.E. (2020). Development of a Force Field for the Simulation of Single-Chain Proteins and Protein-Protein Complexes. J. Chem. Theory Comput. *16*, 2494–2507. https://doi.org/10.1021/acs.jctc.9b00251.

32. Li, P., and Merz, K.M., Jr. (2016). MCPB.py: A Python Based Metal Center Parameter Builder. J. Chem. Inf. Model. *56*, 599–604. https://doi.org/10.1021/acs.jcim.5b00674.

33. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Petersson, G.A., Nakatsuji, H., et al. (2016). Gaussian 16 (Rev. C.01).

34. Maheshwari, S., and Brylinski, M. (2015). Predicting protein interface residues using easily accessible on-line resources. Brief. Bioinform. *16*, 1025–1034. https://doi.org/10.1093/bib/bbv009.

35. Wong, W., Bai, X.C., Brown, A., Fernandez, I.S., Hanssen, E., Condron, M., Tan, Y.H., Baum, J., and Scheres, S.H.W. (2014). Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. Elife *3*, e03080. https://doi.org/10.7554/eLife.03080.

36. Nurenberg-Goloub, E., Kratzat, H., Heinemann, H., Heuer, A., Kotter, P., Berninghausen, O., Becker, T., Tampe, R., and Beckmann, R. (2020). Molecular analysis of the ribosome recycling factor ABCE1 bound to the 30S post-splitting complex. EMBO J. *39*, e103788. https://doi.org/10.15252/embj.2019103788.

37. Holm, L. (2020). Using Dali for Protein Structure Comparison. Methods Mol. Biol. *2112*, 29–42. https://doi.org/10.1007/978-1-0716-0270-6_3.

38. Gopal, B., Haire, L.F., Gamblin, S.J., Dodson, E.J., Lane, A.N., Papavinasasundaram, K.G., Colston, M.J., and Dodson, G. (2001). Crystal structure of the transcription elongation/anti-termination factor NusA from Mycobacterium tuberculosis at 1.7 A resolution. J. Mol. Biol. *314*, 1087–1095. https://doi.org/10.1006/jmbi.2000.5144.

39. Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature *541*, 353–358. https://doi.org/10.1038/nature21031.

40. Arnvig, K.B., Zeng, S., Quan, S., Papageorge, A., Zhang, N., Villapakkam, A.C., and Squires, C.L. (2008). Evolutionary comparison of ribosomal operon antitermination function. J. Bacteriol. *190*, 7251–7257. https://doi.org/10.1128/JB.00760-08.

41. Ferreira-Cerca, S. (2022). The dark side of the ribosome life cycle. RNA Biol. *19*, 1045–1049. https://doi.org/10.1080/15476286.2022.2121421.

42. Greber, B.J., Boehringer, D., Godinic-Mikulcic, V., Crnkovic, A., Ibba, M., Weygand-Durasevic, I., and Ban, N. (2012). Cryo-EM structure of the archaeal 50S ribosomal subunit in complex with initiation factor 6 and implications for ribosome evolution. J. Mol. Biol. *418*, 145–160. https://doi.org/10.1016/j.jmb.2012.01.018.

43. Lecompte, O., Ripp, R., Thierry, J.C., Moras, D., and Poch, O. (2002). Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. Nucleic Acids Res. *30*, 5382–5390. https://doi.org/10.1093/nar/gkf693.

44. Armache, J.P., Anger, A.M., Márquez, V., Franckenberg, S., Fröhlich, T., Villa, E., Berninghausen, O., Thomm, M., Arnold, G.J., Beckmann, R., and Wilson, D.N. (2013). Promiscuous behaviour of archaeal ribosomal proteins: implications for eukaryotic ribosome evolution. Nucleic Acids Res. *41*, 1284–1293. https://doi.org/10.1093/nar/gks1259.

45. Marquez, V., Frohlich, T., Armache, J.P., Sohmen, D., Donhofer, A., Mikolajka, A., Berninghausen, O., Thomm, M., Beckmann, R., Arnold, G.J., and Wilson, D.N. (2011). Proteomic characterization of archaeal ribosomes reveals the presence of novel archaeal-specific ribosomal proteins. J. Mol. Biol. *405*, 1215–1232. https://doi.org/10.1016/j.jmb.2010.11.055.

46. Panasenko, O.O., and Collart, M.A. (2012). Presence of Not5 and ubiquitinated Rps7A in polysome fractions depends upon the Not4 E3 ligase. Mol. Microbiol. *83*, 640–653. https://doi.org/10.1111/j.1365-2958.2011.07957.x.

47. Takehara, Y., Yashiroda, H., Matsuo, Y., Zhao, X., Kamigaki, A., Matsuzaki, T., Kosako, H., Inada, T., and Murata, S. (2021). The ubiquitination-deubiquitination cycle on the ribosomal protein eS7A is crucial for efficient translation. iScience *24*, 102145. https://doi.org/10.1016/j.isci.2021.102145.

48. Matsuki, Y., Matsuo, Y., Nakano, Y., Iwasaki, S., Yoko, H., Udagawa, T., Li, S., Saeki, Y., Yoshihisa, T., Tanaka, K., et al. (2020). Ribosomal protein S7 ubiquitination during ER stress in yeast is associated with selective mRNA translation and stress outcome. Sci. Rep. *10*, 19669. https://doi.org/10.1038/s41598-020-76239-3.

49. Ikeuchi, K., Ivic, N., Buschauer, R., Cheng, J., Fröhlich, T., Matsuo, Y., Berninghausen, O., Inada, T., Becker, T., and Beckmann, R. (2023). Molecular basis for recognition and deubiquitination of 40S ribosomes by Otu2. Nat. Commun. *14*, 2730. https://doi.org/10.1038/s41467-023-38161-w.

50. Blombach, F., Launay, H., Snijders, A.L., Zorraquino, V., Wu, H., de Koning, B., Brouns, S.J., Ettema, T.G., Camilloni, C., Cavalli, A., et al. (2014). Archaeal MBF1 binds to 30S and 70S ribosomes via its

helix-turn-helix domain. Biochem. J. *462*, 373–384. https://doi.org/10.1042/BJ20131474.

51. Blombach, F., Salvadori, E., Fouqueau, T., Yan, J., Reimann, J., Sheppard, C., Smollett, K.L., Albers, S.V., Kay, C.W.M., Thalassinos, K., and Werner, F. (2015). Archaeal TFEalpha/beta is a hybrid of TFIIE and the RNA polymerase III subcomplex hRPC62/39. Elife *4*, e08378. https://doi.org/10.7554/eLife.08378.
52. Kabsch, W. (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132. https://doi.org/10.1107/S0907444909047337.
53. Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. *68*, 352–367. https://doi.org/10.1107/S0907444912001308.
54. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., et al. (2018). MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. *27*, 293–315. https://doi.org/10.1002/pro.3330.
55. Agirre, J., Atanasova, M., Bagdonas, H., Ballard, C.B., Baslé, A., Beilsten-Edmands, J., Borges, R.J., Brown, D.G., Burgos-Mármol, J.J., Berrisford, J.M., et al. (2023). The CCP4 suite: integrative software for macromolecular crystallography. Acta Crystallogr. D Struct. Biol. *79*, 449–461. https://doi.org/10.1107/S2059798323003595.
56. Beilsten-Edmands, J., Winter, G., Gildea, R., Parkhurst, J., Waterman, D., and Evans, G. (2020). Scaling diffraction data in the DIALS software package: algorithms and new approaches for multi-crystal scaling. Acta Crystallogr. D Struct. Biol. *76*, 385–399. https://doi.org/10.1107/S2059798320003198.
57. Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr. D Biol. Crystallogr. *62*, 72–82. https://doi.org/10.1107/S0907444905036693.
58. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501. https://doi.org/10.1107/S0907444910007493.
59. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612. https://doi.org/10.1002/jcc.20084.
60. Gouet, P., Courcelle, E., Stuart, D.I., and M√©toz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics *15*, 305–308. https://doi.org/10.1093/bioinformatics/15.4.305.
61. Negi, S.S., Schein, C.H., Oezguen, N., Power, T.D., and Braun, W. (2007). InterProSurf: a web server for predicting interacting sites on protein surfaces. Bioinformatics *23*, 3397–3399. https://doi.org/10.1093/bioinformatics/btm474.
62. Qin, S., and Zhou, H.X. (2007). meta-PPISP: a meta web server for protein-protein interaction site prediction. Bioinformatics *23*, 3386–3387. https://doi.org/10.1093/bioinformatics/btm434.
63. Kufareva, I., Budagyan, L., Raush, E., Totrov, M., and Abagyan, R. (2007). PIER: protein interface recognition for structural proteomics. Proteins *67*, 400–417. https://doi.org/10.1002/prot.21233.
64. Zhang, Q.C., Deng, L., Fisher, M., Guan, J., Honig, B., and Petrey, D. (2011). PredUs: a web server for predicting protein interfaces using structural neighbors. Nucleic Acids Res. *39*, W283–W287. https://doi.org/10.1093/nar/gkr311.
65. Murakami, Y., and Mizuguchi, K. (2010). Applying the Naive Bayes classifier with kernel density estimation to the prediction of protein-protein interaction sites. Bioinformatics *26*, 1841–1848. https://doi.org/10.1093/bioinformatics/btq302.
66. Porollo, A., and Meller, J. (2007). Prediction-based fingerprints of protein-protein interactions. Proteins *66*, 630–645. https://doi.org/10.1002/prot.21248.
67. Segura, J., Jones, P.F., and Fernandez-Fuentes, N. (2012). A holistic in silico approach to predict functional sites in protein structures. Bioinformatics *28*, 1845–1850. https://doi.org/10.1093/bioinformatics/bts269.
68. Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., and Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. *47*, W636–W641. https://doi.org/10.1093/nar/gkz268.
69. Mirdita, M., von den Driesch, L., Galiez, C., Martin, M.J., Söding, J., and Steinegger, M. (2017). Uniclust databases of clustered and deeply annotated protein sequences and alignments. Nucleic Acids Res. *45*, D170–D176. https://doi.org/10.1093/nar/gkw1081.
70. Pei, J., Kim, B.H., and Grishin, N.V. (2008). PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res. *36*, 2295–2300. https://doi.org/10.1093/nar/gkn072.
71. Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. *292*, 195–202. https://doi.org/10.1006/jmbi.1999.3091.
72. Remmert, M., Biegert, A., Hauser, A., and Söding, J. (2011). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat. Methods *9*, 173–175. https://doi.org/10.1038/nmeth.1818.
73. Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
74. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One *5*, e9490. https://doi.org/10.1371/journal.pone.0009490.
75. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.
76. Han, M.V., and Zmasek, C.M. (2009). phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinf. *10*, 356. https://doi.org/10.1186/1471-2105-10-356.
77. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.
78. PLUMED consortium (2019). Promoting transparency and reproducibility in enhanced molecular simulations. Nat. Methods *16*, 670–673. https://doi.org/10.1038/s41592-019-0506-8.
79. Wurtzel, O., Sapra, R., Chen, F., Zhu, Y., Simmons, B.A., and Sorek, R. (2010). A single-base resolution map of an archaeal transcriptome. Genome Res. *20*, 133–141. https://doi.org/10.1101/gr.100396.109.
80. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., et al. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. *51*, D638–D646. https://doi.org/10.1093/nar/gkac1000.

Please cite this article in press as: Phung et al., Archaeal NusA2 is the ancestor of ribosomal protein eS7 in eukaryotes, Structure (2024), https://doi.org/10.1016/j.str.2024.10.019

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-NusA2	This paper	N/A
Goat anti-Rabbit IgG (H+L) Secondary Antibody, DyLight™ 680	Thermo Fisher	Cat#35568 RRID: AB_614946
Bacterial and virus strains		
E. coli DH5α	NEB	Cat#C2987I
E. coli BL21 (DE3) pLysS	TransGen Biotech	Cat#CD701
Chemicals, peptides, and recombinant proteins		
S. solfataricus NusA2 recombinant protein	This paper	N/A
S. acidocaldarius NusA2 recombinant protein	This paper	N/A
S. solfataricus NusA1 recombinant protein	This paper	N/A
Critical commercial assays		
Structure Screen 1 + 2	Molecular Dimensions	Cat#MD1-30
Deposited data		
ChIP-seq and RNA-seq data	Blombach et al.[2]	GEO: GSE141290
Sso NusA2 crystal structure	This paper	PDB: 9F9U
Saci NusA2 crystal structure	This paper	PDB: 9FJL
Alignment and tree files used for phylogenetic analysis	This paper	https://doi.org/10.5281/zenodo.13693274
Experimental models: Organisms/strains		
S. solfataricus P2	DSMZ	RRID:NCBITaxon_273057
S. acidocaldarius DSM639	DSMZ	RRID:NCBITaxon_330779
E. coli DH5α	NEB	Cat#C2987I
E. coli BL21 (DE3) pLysS	TransGen Biotech	Cat#CD701
Oligonucleotides		
Saci_0767-NdeI-Fw GGCGCATATGAAGATACCACTAGATTATAT	This paper	N/A
Saci_0767-XhoI-Rv GGCGCTCGAGTCATTCCACCCTTATTCTAACTT	This paper	N/A
Sso_0172-NdeI-Fw GGCGCATATGATGAAGATGAAGATCCCCCTAGA	This paper	N/A
Sso_0172-XhoI-Rv GGCGCTCGAGTTACTCTACCCTTATTCTAA	This paper	N/A
Sso_0172-11C11A-Fw GGCGCATATGAAGATGAAGATCCCCCT AGATTATGTTGCTGTTAGAAGT	This paper	N/A
Sso_0172-18C18A-Fw GGCGCATATGAAGATGAAGATCCCCCTAGA TTATGTTTGCGTTAGAAGTGGACTTCTCGCT AATAGATGT	This paper	N/A
Sso_0172-21C21A-Fw GGCGCATATGAAGATGAAGATCCCCCTAG ATTATGTTTGCGTTAGAAGTGGACTTCTCT GCAATAGAGCT	This paper	N/A
DNA probe sequence CTCTGTTCGAATTACCAACCTTGCGT	This paper	N/A

(*Continued on next page*)

Please cite this article in press as: Phung et al., Archaeal NusA2 is the ancestor of ribosomal protein eS7 in eukaryotes, Structure (2024), https://doi.org/10.1016/j.str.2024.10.019

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
RNA probe sequence CUCUGUUCGA AUUACCAACCUUGCGU	This paper	N/A
Recombinant DNA		
pET21a(+)	Novagen	Cat#69740
pET21 NusA1 expression plasmid	This paper	N/A
pET21 NusA2 expression plasmid	This paper	N/A
pET21 NusA2-AAA expression plasmid	This paper	N/A
Software and algorithms		
ChIP-seq and RNA-seq data plotting code	Blombach et al.[2]	https://doi.org/10.5281/zenodo.5346581
XDS	Kabsch[52]	https://xds.mr.mpg.de/html_doc/downloading.html
Phenix v1.20.1 suite (Phaser v2.8.3)	Afonine et al.[53]; McCoy et al.[24]	https://phenix-online.org/download/
MolProbity	Williams et al.[54]	http://molprobity.biochem.duke.edu/
CCP4 suite (Xia2, Aimless, SHELXC/D/E, ARP/wARP v8.0, and Coot v0.9.8.8)	Agirre et al.[55]; Beilstein-Edmands et al.[56]; Evans[57]; Sheldrick et al.[22]; Perrakis et al.[23]; Emsley et al.[58]	https://www.ccp4.ac.uk/download/#os=macos
UCSF Chimera	Pettersen et al.[59]	https://www.cgl.ucsf.edu/chimera/download.html
ESPript 3.0	Gouet et al.[60]	https://espript.ibcp.fr/ESPript/ESPript/
InterProSurf	Negi et al.[61]	https://curie.utmb.edu/prosurf.html
meta-PPISP	Qin et al.[62]	https://pipe.rcc.fsu.edu/meta-ppisp.html
PIER	Kufareva et al.[63]	No longer available
PredUs2.0	Zhang et al.[64]	https://honiglab.c2b2.columbia.edu/hfpd/html/predus.html
PSIVER	Murakami et al.[65]	https://mybiosoftware.com/psiver-prediction-protein-protein-interaction-sites-protein-sequences.html
SPPIDER	Porollo et al.[66]	https://sppider.cchmc.org/
VORFFIP	Segura et al.[67]	http://www.bioinsilico.org/cgi-bin/VORFFIP/htmlVORFFI/home
Normalization equation	This paper	N/A
DALI	Holm et al.[37]	http://ekhidna2.biocenter.helsinki.fi/dali/
Clustal Omega	Madeira et al.[68]	https://www.ebi.ac.uk/jdispatcher/msa/clustalo
MMseqs2	Mirdita et al.[69]	https://github.com/soedinglab/MMseqs2
PROMALS3D	Pei et al.[70]	http://prodata.swmed.edu/promals3d/promals3d.php
PSIPRED	Jones et al.[71]	http://bioinf.cs.ucl.ac.uk/psipred
HHblits	Remmert et al.[72]	https://toolkit.tuebingen.mpg.de/tools/hhblits
trimAl	Capella-Gutierrez et al.[73]	https://vicfero.github.io/trimal/
FastTree 2.1	Price et al.[74]	http://www.microbesonline.org/fasttree/
IQ-TREE 2	Minh et al.[75]	http://www.iqtree.org/
Archeopteryx	Han et al.[76]	https://www.phylosoft.org/archaeopteryx/
Parallel tempering-Metadynamics	Barducci et al.[30]	N/A
Gromacs package v2019.6 (DES-Amber forcefield v1.0)	Abraham et al.[77]; Piana et al.[31]	https://manual.gromacs.org/2019/download.html
PLUMED package v2.7.0	Plumed consortium[78]	https://www.plumed.org/
MCPB.py	Li et al.[32]	https://ambermd.org/AmberTools.php
Gaussian	Frisch et al.[33]	https://gaussian.com/gaussian16/
ChIP-seq and RNA-seq data analysis	Blombach et al.[2]	https://doi.org/10.5281/zenodo.5346581

Please cite this article in press as: Phung et al., Archaeal NusA2 is the ancestor of ribosomal protein eS7 in eukaryotes, Structure (2024), https://doi.org/10.1016/j.str.2024.10.019

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Sulfolobus acidocaldarius DSM639.
Saccharolobus solfataricus P2.
E. coli DH5α was used as host for plasmid production.
E. coli BL21 (DE3) pLysS was used as host for heterologous protein expression.

METHOD DETAILS

Cloning strategy, expression and purification

The genes encoding archaeal Saccharolobus solfataricus NusA1 and NusA2 (respectively Sso_0220 and Sso_0172) were cloned into pET-21a and over-expressed in Escherichia coli BL21 (DE3) pLysS by IPTG induction. The E. coli cell pellets were resuspended in N200 buffer (25 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, 0.1 mM $ZnSO_4$, 10% glycerol) supplemented with 1 tablet of protease inhibitor (Roche), 1mg/mL of RNase A (NEB) and DNase I (Sigma). The cell lysate was obtained by sonication and was heated 10 min at 65°C and 15 min at 75°C with clarification step at 20,000 g between each heating steps. The obtained supernatant was then filtered through 0.2 μm filter prior loaded on HiTrap Heparin column (Cytiva) and Superose 12 size exclusion column (Cytiva) in 25 mM Tris-HCl pH 8.0, 200 mM NaCl, 5 μM $ZnSO_4$ buffer. Finally, the recombinants proteins were concentrated to 5 mg/ml. NusA2-AAA mutant was obtained by amplification of the interested gene with corresponding oligos reported in the key resources table and cloned into pET21a. Expression and purification of NusA2-AAA was proceeded as for the wild type.

RNA polymerase Chromatin immunoprecipitation sequencing and RNA-seq data analysis

Input normalised Chromatin immunopreicipitation sequencing data and RNA-seq occupancy was obtained from NCBI GEO superseries GSE141290.[2] The transcription start site coordinates were obtained from.[79] Data were plotted as described previously[2] with the code for data plotting available at https://doi.org/10.5281/zenodo.5346581.

Heat stability test

25μL of Sso NusA1 (1mg/mL), Sso NusA2 (5mg/mL) and NusA2-AAA (1mg/mL) alone or supplemented with 100 mM EDTA were incubated at 80°C for 15 and 30 min. A control sample for each sample was prepared and kept at room temperature. Then, all samples were centrifugated for 10 min at 20,000 g. The pellets and the supernatants were separated and both resuspended in 25 μL of SDS-PAGE loading buffer to be resolved on a SDS-PAGE.

Western blotting

S. solfataricus cells pellets were obtained at OD_{600} of 0.25, 0.3 and 1. Whole-cell extracts were obtained through sonication, 10 min, in pulse mode (10 s on, 20 s off) with 50% amplitude, followed by centrifugation at 21,000 g for 30 min. On SDS-PAGE, 10 mg of cell extract were loaded alongside with 124, 62, 31.5, and 15.52 ng of recombinant NusA2 protein. Proteins were transferred to a nitrocellulose membrane by a semi dry transblot. The membranes were saturated with PBST buffer (1X PBS, 0.05% Tween-20) containing 5% non-fat dry milk for 1 hour at room temperature. Following saturation, the membranes were washed three times with PBST buffer. The membranes were then incubated with in-house polyclonal rabbit antibodies against NusA2, diluted 1:1,000 in PBST buffer containing 5% non-fat dry milk, for 1 hour. After the primary antibody incubation, the membranes underwent a second washing step, followed by incubation with goat anti-rabbit secondary antibodies conjugated to Dyelight 680 (Invitrogen), diluted 1:10,000 in PBST buffer containing 5% non-fat dry milk, for 1 hour. Following a final series of washes, the fluorescent signal was recorded on a GE FLA 9000 Typhoon imaging system.

Electrophoretic mobility shift assay

EMSA was performed preparing 20 μl reactions with protein concentration of 0.1, 0.25, 0.5, 0.75, 1, 2.5, and 5 μM in buffer 200 mM NaCl, 20 mM MOPS, pH 6.5, 10 mM $MgCl_2$, 0.1 mg/ml BSA and 10% glycerol in presence of 5 nM of radiolabelled probes. Samples were incubated for 10 min at 65°C and 15 μl were loaded onto a native 7% Native Tris-Glycine gel (1X TGOE, 7% Acry/Bis-acry, 3% glycerol, 0.5 mM DTT) and run at 150V for 30 min in 1X TBE (100 mM Tris/100 mM boric acid/5 mM EDTA, pH 7.0). The gels were dried, then exposed to a phosphor screen for 1h and scanned on a GE FLA 9000 Typhoon. Probe sequences are listed in the key resources table.

Protein crystallization, data collection and processing

Sso NusA2. The sample was concentrated up to 5 mg/ml in buffer 25 mM Tris-HCl pH 8.0, 200 mM NaCl, 50 μM ZnSO4. Crystals were grown in a precipitant solution consisting of 0.1 M Bicine pH 9.0, 10% w/v PEG 20,000, 2% 1,4-dioxane (Structure Screen 1 + 2, Molecular Dimension) in sitting drop, vapour diffusion at 20°C on a 96-well plate prepared with Mosquito LCP (SPT Labtech). A single-wavelength anomalous diffraction (SAD) dataset on the zinc ion was collected at wavelength of 1.2826 Å using an EIGER2 CdTe 16M detector at the P14 beamline of PETRA III synchrotron (DESY, Hamburg, Germany). Data were processed using the automated pipeline available at PETRA III. Scaling and merging were carried out in XDS.[52] Structure determination and phasing were performed in SHELXC/D/E,[22] followed by ARP/wARP v8.0 model building.[23] Then, the model obtained from the automated processing was

Please cite this article in press as: Phung et al., Archaeal NusA2 is the ancestor of ribosomal protein eS7 in eukaryotes, Structure (2024), https://doi.org/10.1016/j.str.2024.10.019

initially refined manually in Coot v0.9.8.8[58] followed by refinement in Phenix v1.20.1.[53] Structure validation was carried out using MolProbity webserver http://molprobity.biochem.duke.edu/.[54]

Saci NusA2. Crystals were obtained in 0.1 M Hepes pH 7.0, 0.7 - 0.9 M NaH2PO4/KH2PO4 in a 24-well plate, vapour diffusion, hanging drop, 20°C. The dataset was collected at Diamond Light Source synchrotron (Didcot, UK), beamline I03 at wavelength of 0.9762 Å using a Dectris Pilatus 6M detector. Data were processed using the ccp4 software suite v8.0.012.[55] Scaling was carried out in Xia2[56] and sub-sequentially data were merged in Aimless.[57] Structure determination and phasing were obtained by molecular replacement in Phaser v2.8.3[24] using the Sso NusA2 structure as initial model. The model was finally refined in Coot v0.9.8.8 and Phenix v1.20.1. Structure validation was carried out using MolProbity webserver.

KH sequence- and structure-based searches

KH domain-containing proteins were searched in the UniProt https://www.uniprot.org/, Protein Data Bank https://www.rcsb.org/, Pfam http://pfam.xfam.org/, and SCOP https://scop.berkeley.edu/ databases using KH domain as query. The output was manually filtered to remove redundancy and the pdb files were then manually inspected in Chimera v1.13[59] to confirm the presence of the KH signature and check for conformational changes due to the different techniques used (crystallography, NMR, or cryo-EM) or induced by ligand binding. Established that none of the proteins analysed change conformation, one representative structure for each protein was selected with its own ligand when available. The search and pruning led to a final set of 33 structures from 24 different proteins among which 26 KH domains type I, 4 of them bound to DNA and 22 RNA-binding proteins, and 7 KH domains type II. Then, all additional domains not belonging to the KH domain were deleted and we performed a superimposition using the Chimera's MatchMaker tool with default parameters. The superimposition obtained was used to carry out structure-based alignments. The alignments were generated using a residue-residue distance cutoff of 10 Å and they were split in three sets: DNA- and RNA-binding type I KH domains and KH type II domains, respectively. All figures were prepared with ESPript 3.0[60] using the %MultAlign coloring scheme with a global score of 0.5 as parameters. Below, the list of proteins used for the alignments in alphabetic order with their corresponding pdb codes and chain identifiers in brackets. DNA-binding KH type I: Homo sapiens Far Upstream Element-binding protein 1 (FUBP1 or FBP1 or FUSE binding protein 1) (1j4w-A), H. sapiens Heterogeneous nuclear ribonucleoprotein K (hnRNP K) (1j5k-A), H. sapiens Poly(rC)-binding protein 1 (PCBP1) (1ztg-A). RNA-binding KH type I: Saccharomyces cerevisiae Branchpoint-bridging protein (BBP) (4wal-A), Mus musculus Filia protein (3v69-A), H. sapiens Fragile X mental retardation protein 1 (FMR1) (2qnd-A), Caenorhabditis elegans Female germline-specific tumor suppressor gld-1 (4jvy-A), Gallus gallus Insulin-like growth factor 2 mRNA-binding protein 1 (IMP1 or IGF2BP1) (2n8m-A and 2n8l-A), H. sapiens KH domain-containing, RNA-binding, signal transduction-associated protein 3 (KHDRBS3) (5emo-A), Chaetomium thermophilum KRR1 (5oql-g), H. sapiens KH-type splicing regulatory protein (KSRP or KHSRP) (4b8t-A, and 2opu-A), H. sapiens RNA-binding E3 ubiquitin-protein ligase MEX3C (www-A and 5wwx-A), H. sapiens Neuro-oncological ventral antigen 1 (NOVA-1) (2anr-A), H. sapiens PNO1 (6g5i-x), Caulobacter vibrioides Polyribonucleotide nuclotidyltransferase (PNPase) (4am3-A), H. sapiens Protein quaking (QKI) (4jvh-A), H. sapiens Splicing factor 1 (SF1-Bo isoform) (1k1g-A), M. musculus Tudor and KH domain containing protein (TDRKH) (1we8-A), H. sapiens Vigilin (or HDL binding protein) (2cte-A, and 1vih-A). KH type II: Pyrococcus horikoshii Cleavage and polyadenylation specificity factor subunit (CPSF or putative uncharacterized protein PH1404) (3af5-A), Aquifex aeolicus GTPase Era (3r9w-A), Escherichia coli Nuclear substance A (NusA) (6gov-A), Aeropyrum pernix NusA1 (hypothetical transcription termination factor NusA) (2cxc-A) and S. cerevisiae 40S ribosomal protein S3 (4u4r-S3).

Protein-protein interface prediction

All protein interface prediction programs were run in webserver mode with default parameters using the structure of Sso NusA2 as input with the only exception of PSIVER which requires the protein sequence. The programs' list is the following: InterProSurf,[61] meta-PPISP,[62] PIER,[63] PredUs2,[64] PSIVER,[65] SPPIDER,[66] and VORFFIP.[67] The outputs of meta-PPISP, PIER, PSIVER, SPPIDER and VORFFIP were normalized according to Equation 1 applying a threshold of the 30% prior the normalization to exclude all negative results and buried residues (Table S1). InterProSurf does not provide any score but only a list of 10 residues clusters in order of probability. This order was used to manually assign values to the listed residues from 0.1 to 1. All outputs were then merged with the PredUs2 output, which provides only positive hits with normalized scores by default, by calculating the mean of all values for each residue and normalized according to Equation 1 (last column of Table S1 and Figure 3A).

$$N_x = \frac{N_i - N_{min}}{N_{max} - N_{min}} \quad \text{(Equation 1)}$$

Structure-based analysis

A structure-based search in DALI server[37] was performed using the Sso NusA2 structure as query. From the output, we selected only the first top 20 entries and we discarded all entries having a rmsd score higher then 5 Å, sequence identity lower then 10%, and orthologs to ensure a unique set of protein. The six structures obtained from the pruning were retrieved from the Protein Data Bank and analysed using the structural equivalences information provided on the DALI output, which were reported on Table S2.

Sequence alignments

All sequence-based alignments were carried out in Clustal Omega and rendered using Espript 3.0.[68] The uniprot identifiers for the different protein sequences shown in Figures 2C, S3A, and S3B are as follows: bacterial NusA: *Escherichia coli* str. K12 P0AFF6

and *Mycobacterium tuberculosis* str. H37Rv P9WIV3; archaeal NusA1: *Aeropyrum pernix* Q9YAU4, *Methanocaldococcus jannaschii* Q58447, *Saccharolobus solfataricus* Q980R4, *Sulfolobus acidocaldarius* P11523; NusA2: *Lokiarchaeum* str. GC14_75 A0A0F8Y5E8, *Pyrococcus furiosus* Q8U1J3, *Saccharolobus solfataricus* Q980V4, *Sulfolobus acidocaldarius* Q4JAN9, Ca. *Nitrosotenuis cloacae* A0A3G1B2K6; eS7: *Homo sapiens* P62081, *Saccharomyces cerevisiae* eS7a *P26786, Arabidopsis thaliana* eS7-1 Q9C514, *Plasmodium falciparum* 3D7 Q8IET7;uS8: *Aeropyrum pernix* Q9YF89.2, *Arabidopsis thaliana* P42798.2, *Archaeoglobus fulgidus* O28369.2, *Escherichia coli* B1X6F8.1, *Haloarcula marismortui* P12742.3, *Homo sapiens* P62244.2, *Methanocaldococcus jannaschii* P54041.1, *Mycobacterium tuberculosis* Q9YF89.2, *Plasmodium falciparum* O77395, *Pyrococcus furiosus* Q8U014.1, *Saccharolobus solfataricus* Q9UX92, *Saccharomyces cerevisiae* P0C0W1.2, *Thermotoga maritima* Q9ZAE5.2, *Thermus thermophilus* P0DOY9.1.

Phylogenetic analysis

Sequences of NusA and eS7 homologs were collected with PSI-BLAST (five iterations, 1e-5 E-value profile inclusion threshold) for against the Uniclust90 database (i.e., UniProt clustered at 90% sequence identity with MMseqs2,[69] using five selected representative sequences of tandem KH domains as queries (yeast and microsporidia eS7, E. coli NusA, A. pernix NusA1 and S. solfataricus NusA2). Sequence hits longer than 70 aa and with E-value lower than 0.001 to at least one query were collected and extended to match the entire length of tandem KH domains. The 9985 sequences were clustered at 40% identity with CD-HIT and resulting 281 representatives were further clustered in Cytoscape, using pairwise BLAST E-values lower than 0.001 as edge weights. False positives were removed and remaining representatives (NusA_Uniclust40) were used to retrieve all sequences (NusA_Uniclust90) belonging to the four major clusters (NusA, NusA1, NusA2, and eS7).

The NusA_Uniclust40 sequences were aligned with PROMALS3D[70] guided by seven representative crystal structures (PDB IDs 1cxc, 4mtn, 4u4r, 5lm9, 5xyi, 6rm3 and S. solfataricus NusA2). Secondary structures were predicted with PSIPRED[71] using HHblits-generated MSAs.[72] The resulting multiple sequence alignment (MSA) was manually adjusted based on the predicted secondary structures. For a phylogenetic tree reconstruction, all columns in the MSA with at least 10% gaps were removed with trimAl[73] ("-gt 0.9"). The phylogenetic tree was built with FastTree 2.1[74] (using JTT+CAT model and Shimodaira-Hasegawa test for local support values) and IQ-TREE 2[75] (VT+F+G4 model and both UFBoot and SH-aLRT with 1000 replicates for estimating support values), and visualized with Archeopteryx.[76] The aligment and tree files are available at Zenodo: https://doi.org/10.5281/zenodo.13693274.

Full sequence alignments were obtained by adding respective NusA_Uniclust90 sequences to the seed MSAs with MAFFT ("mafft –add –keeplength").

Genome context analysis

The genome context analysis for archaeal *nusA2* was carried out manually by checking the neighbouring genes of *nusA2* and their respective Pfam database annotations. The analysis was initiated by a search for *aspS* (the neighbouring gene in the *S. solfataricus* P2 genome) in the STRING database[80] that indicated the conserved neighbourhood of *aspS* and a gene coding for a NusA-like protein in Archaea that we verified to be *nusA2*. NCBI genome assemblies and locus tags for the *nusA2* homologues are listed in Table S3.

Cell lysate sucrose gradient fractionation

Sulfolobus solfataricus (Sso) cells were cultivated at 76°C until reaching an optical density (OD_{600}) of 0.5. Subsequently, the cell cultures were chilled to 4°C and harvested by centrifugation at 6000 g for 20 minutes at 4°C. The resulting cell pellets underwent three washes with 1X phosphate-buffered saline (PBS) and were then freeze-stored at −80°C. Upon resuspension, 1 L cell pellets were treated with 6 mL of Lysis buffer (20 mM HEPES pH 7.4, 10 mM Mg acetate, 40 mM NH_4Cl, 1 mM DTT), supplemented with a cocktail of EDTA-free protease inhibitors (cOmpleteTM, Roche). Whole-cell extracts were obtained through sonication, 10 min, in pulse mode (10 s on, 20 s off), and 50% amplitude, followed by centrifugation at 21,000 g for 30 min. 2 mg of total protein from whole-cell extracts was layered onto a linear 10–30% sucrose gradient prepared in lysis buffer in a 14 x 89 mm polypropylene tube (Beckman Coulter). The gradient was then centrifuged for 4 hours at 34,000 rpm at 4°C using the Beckman-Avanti XPN-80 SW41 rotor. Monitoring of the 30S and 50S ribosomal subunit positions was achieved by measuring absorbance at 254 nm with the ISCO UA-6 gradient fraction collector. Each gradient run resulted in 22 fractions of 500 μl each. Subsequently, 20 μL from each fraction was resolved onto SDS-PAGE. Proteins were transferred to a nitrocellulose membrane for immunoblotting analysis with polyclonal antibodies against Sso NusA2.

Molecular dynamics simulations

The molecular dynamics (MD) simulations were performed using the 2019.6 version of the Gromacs package,[77] patched with the 2.7.0 version of the PLUMED package.[78]

The protein structure was based on the crystal structure. For the simulations without the Zn atom, it was simply removed. For the simulations with the Zn atom, it was parametrised using the MCPB.py software[32] and Gaussian.[33] The rest of the protein and the added water molecules and ions (0.15 M NaCl, charge neutralised) were parametrised using the DES-Amber forcefield with 1.0 scaling.[31]

All the simulations used the following equilibration procedure: 1 ns of NVT equilibration with restraints, 10 ns of NPT equilibration with Berendsen thermostat and restraints, and finally 10 ns of NPT equilibration with Parinello-Rahman thermostat without restraints.

Please cite this article in press as: Phung et al., Archaeal NusA2 is the ancestor of ribosomal protein eS7 in eukaryotes, Structure (2024), https://doi.org/10.1016/j.str.2024.10.019

In the well-tempered metadynamics simulations,[30] the deposited hills had a pre-scaling height of 1 kJ and were deposited every 2 ps (1000 simulation steps). The widths of the Gaussians were 0.25 and 0.001 for the contact map distance from the native structure of the zinc finger and K1 domain and the Met1 and Glu45 distance, respectively.

The simulations were coupled with parallel tempering, meaning that there were 20 replicas spanning the range from 80 to 120°C, with exchanges between neighbouring replicas attempted every 2 ps (1000 simulations steps). The total production length of every simulation was 1 μs.

QUANTIFICATION AND STATISTICAL ANALYSIS

X-ray crystallography data collection and refinement statistics are summarised in Table 1.