High complexity of Glutamine synthetase regulation in *Methanosarcina mazei*: Small protein 26 interacts and enhances glutamine synthetase activity

Miriam Gutt[1] (iD), Britta Jordan[1], Katrin Weidenbach[1], Mirja Gudzuhn[1], Claudia Kiessling[1], Liam Cassidy[2], Andreas Helbig[2], Andreas Tholey[2], Dennis Joshua Pyper[3], Nina Kubatova[3], Harald Schwalbe[3] and Ruth Anne Schmitz[1] (iD)

1 Institute for General Microbiology, Christian-Albrechts-University, Kiel, Germany
2 AG Proteomics & Bioanalytics, Institute for Experimental Medicine, Christian-Albrechts-University, Kiel, Germany
3 Institute of Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance (BMRZ), Johann Wolfgang Goethe University, Frankfurt am Main, Germany

Small ORF (sORF)-encoded small proteins have been overlooked for a long time due to challenges in prediction and distinguishing between coding- and noncoding-predicted sORFs and in their biochemical detection and characterization. We report on the first biochemical and functional characterization of a small protein (sP26) in the archaeal model organism *Methanosarcina mazei*, comprising 23 amino acids. The corresponding encoding leaderless mRNA (spRNA26) is highly conserved on nucleotide level as well as on the coded amino acids within numerous Methanosarcina strains strongly arguing for a cellular function of the small protein. spRNA26 level is significantly enhanced under nitrogen limitation, but also under oxygen and salt stress conditions. Using heterologously expressed and purified sP26 in independent biochemical approaches [pull-down by affinity chromatography followed by MS analysis, reverse pull-down, microscale thermophoresis, size-exclusion chromatography, and nuclear magnetic resonance spectroscopy (NMR) analysis], we observed that sP26 interacts and forms complexes with *M. mazei* glutamine synthetase ($GlnA_1$) with high affinity (app. $K_D = 0.76$ μM $± 0.29$ μM). Moreover, seven amino acids were identified by NMR analysis to directly interact with $GlnA_1$. Upon interaction with sP26, $GlnA_1$ activity is significantly stimulated, independently and in addition to the known activation by the metabolite 2-oxoglutarate (2-OG). Besides, strong interaction of sP26 with the PII-like protein $GlnK_1$ was demonstrated (app. $K_D = 2.9$ μM $± 0.9$ μM). On the basis of these findings, we propose that in addition to 2-OG, sP26 enhances $GlnA_1$ activity under nitrogen limitation most likely by stabilizing the dodecameric structure of $GlnA_1$.

Abbreviations
2-OG, 2-oxoglutarate; ACS, acetyl-CoA-decarbonylase/synthase complex; bp, base pair; GS/GOGAT, glutamine synthetase/glutamate synthase; GS, glutamine synthetase; HEPES, N-(2-hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid); LC-MS/MS, liquid chromatography-mass spectrometry / mass spectrometry; MOPS, 3-morpholino-propanesulfonic acid; MST, microscale thermophoresis; MW, molecular weight; N, nitrogen; NADH, nicotinamide adenine dinucleotide (reduced); NMR, nuclear magnetic resonance spectroscopy; NrpR, nitrogen regulator protein R; nt, nucleotides; RNAseq, RNA sequencing; sORF, small ORF; sP, small protein; spRNA, small protein RNA; sRNA, small RNA.

Introduction

Modern genomic and transcriptomic technologies combined with systematic genome-wide approaches have uncovered an unexpected genome complexity in prokaryotes. Besides genes encoding larger proteins and genes of noncoding RNAs (ncRNAs), global approaches have over the past decade discovered a wealth of hidden small genes containing short open reading frames (sORFs) in many prokaryotic genomes [1–3]. These sORFs often encode proteins smaller than 50 amino acids (aa) in length and have been typically missed in genome annotations by automated gene predictions due to too strict assumptions and traditionally considering only standard proteins in the automated gene annotation tools [4,5]. This led to neglecting the existence of an additional layer of complexity represented by small proteins. Besides, small proteins have been difficult to detect biochemically due to technical limitation. Though, nowadays new technologies are emerging, which enable their global profiling in genome-wide approaches, and systematic approaches for global identification are used, for example, by sophisticated bioinformatic predictions, peptidomics, ribosome profiling, and combinations of those [5–15]. Due to their small size, the small proteins are frequently predicted to modify the activity of larger proteins or complexes via physical interactions or interact with the membranes. However, information regarding specific physiological role(s) of verified small proteins is lacking for the majority of confirmed small proteins. Only a fraction of small proteins experimentally identified in diverse prokaryotes have been functionally characterized. This characterization demonstrated that they can play important roles in different functional scenarios and have a broad range of function from cell division, signal transduction, modeling membrane protein recruitment, or (membrane) protein activity to modulation or being part of a larger mostly membrane integrated protein complex (reviewed by Ref. [2,16,17]). In archaea, only a few small proteins have been reported, the majority of which is regulated in response to specific stress. However, for most of them verified functional analysis is scarce or missing [7,8,18–24].

Methanosarcina mazei strain Gö1 belongs to the methylotrophic methanogens of the order *Methanosarcinales*, which have the most versatile substrate spectrum within the methanogenic archaea and significantly contribute to the production of the greenhouse gas [25]. *M. mazei* is able to fix molecular nitrogen under nitrogen (N) limitation. The regulation of the N metabolism particularly of the nitrogen fixation as well as the glutamine synthetase is well studied on the transcriptional and post-transcriptional levels [26–31]. Overall regulation occurs in response to the N availability, where N limitation is in general perceived internally by sensing the internal 2-oxoglutarate (2-OG) pool, which increases under N limitation due to reduced consumption by the ammonium-dependent glutamate dehydrogenase and increasing glutamine synthetase/glutamate synthase pathway (GS/GOGAT) for glutamate synthesis [32]. We recently showed that post-transcriptional regulation by small RNA_{154}, originally identified in a global RNA sequencing (RNAseq) approach [27], plays a central role in N regulation of several components of the N-cycle including glutamine synthetase [28]. Besides, RNAseq and term-seq approaches as well as unpublished RNAseq data sets identified not only high numbers of small noncoding RNAs but also numerous small mRNAs in *M. mazei* containing putative sORFs [27,33]. Aiming to elucidate whether the predicted sORFs identified are translated *in vivo*, the full cytosolic proteome of *M. mazei* was examined using reversed-phase LC-MS MS approaches (bottom-up and top-down strategies) as described in Ref [7,8,19,23]. In these comprehensive studies, overall 68 small proteins were detected and experimentally validated with high or mid confidence. In this present study, the small protein 26 (sP26), identified with low confidence in our previous study but verified by a recent ribosome-profiling analysis (R. A. Schmitz & M. Gutt, unpublished data), was selected aiming to identify its physiological role, due to the fact that the respective encoding mRNA (spRNA26) is upregulated under N starvation [27]. Moreover, the genomic location of spRNA26 downstream of the operon encoding the acetyl-CoA-decarbonylase/synthase (ACS) complex [27] suggested ACS, which itself is upregulated in response to N starvation on the post-transcriptional level [26], as an attractive potential target for sP26. Very recently, nuclear magnetic resonance spectroscopy (NMR) spectroscopy analysis demonstrated that chemically synthesized sP26 is unstructured, whereas bioinformatic tools predict that unstructured sP26 can potentially fold into a structure upon complex formation with a target [21]. Here, we demonstrate by several independent biochemical approaches that sP26 interacts with glutamine synthetase ($GlnA_1$) and enhances its activity in *M. mazei*. Glutamine synthetase, the central component of the N-cycle, is known to be strictly regulated in response to the cellular nitrogen status by various molecular mechanisms in other Prokaryotes. They range from feedback inhibition, post-translational modifications, and allosteric inhibitors to interactions with other proteins [34,35].

Results

Aiming to gain insight into the physiological role of small protein sP26 in *M. mazei,* we addressed our hypothesis that sP26 is directly or indirectly involved in N regulation. First we examined, whether the archaeal small protein interacts with other *M. mazei* proteins and verified the identified interacting partner glutamine synthetase using various approaches.

spRNA26 and the corresponding small protein are highly conserved in Methanosarcina strains and upregulated under N limitation in *M. mazei*

spRNA26 was identified using a differential RNAseq (dRNA-seq) approach and shown to be expressed under N limitation, whereas under N sufficiency no transcript was detectable [27]. It represents an ~ 70 nucleotide (nt) long leaderless RNA encoding a small protein of 23 amino acids. A tryptic peptide from this protein was identified in the cell extraction of cells grown under N starvation conditions via a 2DLC-MS bottom-up proteomic analysis (annotated spectra of the peptide Fig. S1) [8]. The sp26-derived tryptic peptide was identified with a lower level of stringency than applied in previously reported publications [7,8,19], with a protein-level false discovery rate of 5%, and an MS2 fragment ion tolerance of 0.05 Da, however, can be considered as unambiguously identified. This is further supported by verifying translation in ribosome-profiling approaches (Schmitz and Gutt, unpublished data). spRNA26 is located within the 316-bp intergenic region between *MM2083* encoding orotidine 5′-monophosphate decarboxylase and the operon encoding the acetyl-CoA-decarbonylase/synthase (ACS) complex. Its promoter (TATA box and BRE box) was identified between nt 26 and 40 upstream of the transcriptional start site (+1) based on the distance to the start site and their consensus sequence in *M. mazei* (Fig. 1A). Two further alternative BRE boxes can be predicted (indicated with a dashed line). The repressor-binding motif of the general transcriptional regulator nitrogen regulator protein R (NrpR) [29] was not detected. However, the repetitive elements identified upstream of the BRE box (indicated with gray boxes) might represent a weak binding site for the transcriptional activator nitrogen regulator protein A which has been identified in the upstream promoter region of the *nif* operon [31]. Multiple alignments of the respective spRNA26 and homologs in other Methanosarcina species showed that the sORF experimentally verified in *M. mazei* is highly conserved with regard to amino acid as well as the nucleotide sequence (Fig. 1B,C). Further, high conservation of the promoter and the 5' upstream region was identified among the *M. mazei* strains of which four are missing the BRE box identified for *M. mazei* Gö1 and thus might use one of the alternative BRE boxes (Fig. 1B). All homologs are predicted to start with valine, indicating that a noncanonical start codon is used for translation initiation (GTG). Interestingly, at the third position of the small protein three *Methanosarcina barkeri* strains contain methionine (ATG) in contrast to all other strains encoding valine (GTG) at the third position. This might indicate the presence of an alternative translation start. Unfortunately, the peptide detected by the LC-MS MS analysis starts with the 5th position (lysine) of the predicted sORF (see Fig. 1C, indicated with a box in the sequence logo), thus cannot be used to identify the native start of the small protein. In the *M. thermophila* strains and in *M. falvescens,* apparently a deletion of nt 33 occurred causing the generation of a translational stop codon subsequently resulting in a very short version of sP26 (Fig. 1C). Overall, the high conservation of the amino acid as well as the nucleotide sequence argues for a potential physiological role of sP26.

Northern blot and Illumina RNAseq analysis confirmed the reported transcriptional start site (+1) and upregulation of spRNA26 under N limitation (~ 1.5 to 2.5 fold) as well as under oxygen and salt stress (Fig. 2). In agreement with the RNAseq data, spRNA26 appears to be processed into three shorter fragments (65, 63, and 61 nt) in varying amounts depending on the growth phase and stress conditions (Fig. 2A,C).

To study potential effects of the small protein sP26 *in vivo,* the respective sORF was additionally expressed in *M. mazei* from a plasmid under the control of the constitutive promoter *pmcrB* with the ribosome-binding site of *mcrB* (pRS1242, plasmid map, see Fig. S2d). Growth analysis under N sufficient and N limiting conditions showed no significant phenotype; growth rates comparable to the control containing the empty pWM321 vector were obtained independent of additional sP26 synthesis (μ (+N) = 0.080 h^{-1} vs. 0.086 h^{-1}, μ (−N) = 0.082 h^{-1}). Cell extract preparations and subsequent western blot analysis using antibodies raised against synthetic sP26 demonstrated that chromosomally expressed sP26 is induced under N starvation. The majority is in the soluble fraction independently from the N availability but in a higher oligomeric state or bound to a protein with higher molecular mass (Fig. 3).
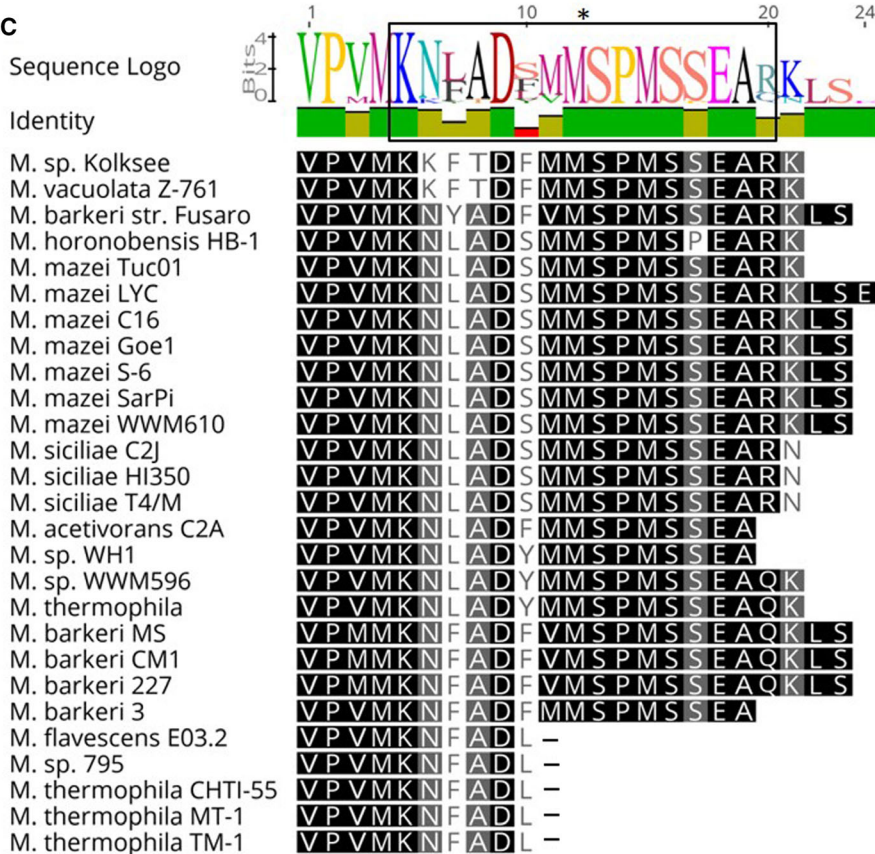
A
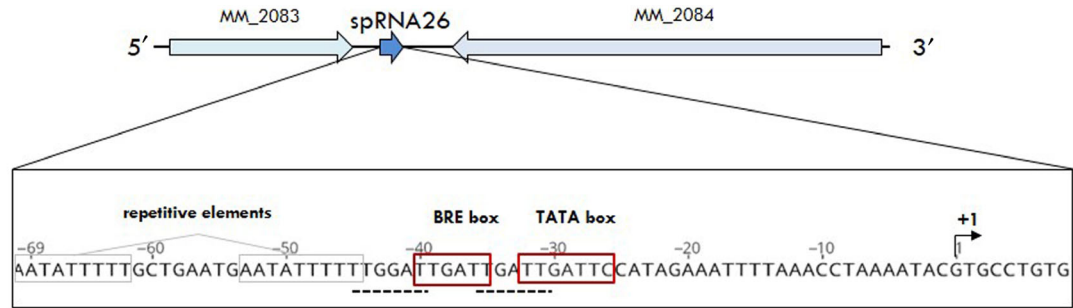
B
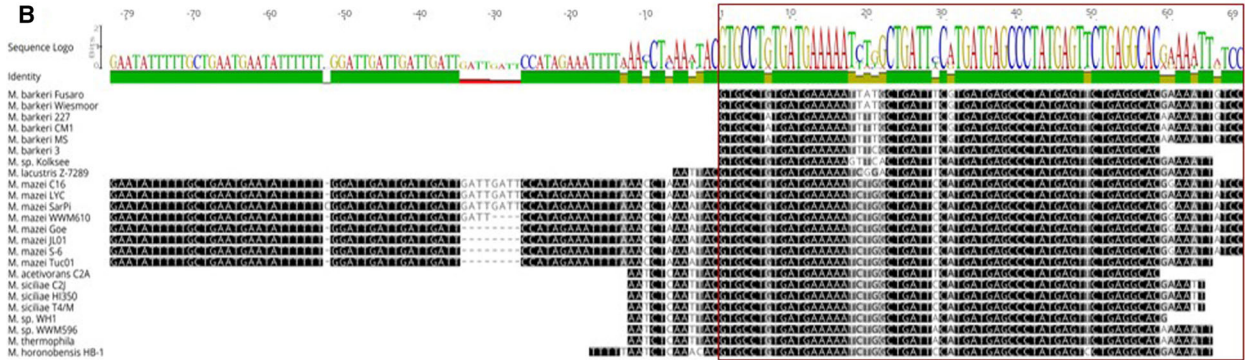
C

Fig. 1. Genomic localization of spRNA26 and homologs in *Methanosarcina* species. (A) Location and promoter of spRNA26, predicted BRE and TATA boxes are indicated as well as alternative BRE boxes with an interrupted line. Additional repetitive elements are boxed (pot. binding sides of NrprA). MM_2083, encoding OMP decarboxylase and MM_2084, encoding γ subunit of acetyl-CoA decarbonylase/synthase. (B) Nucleotide alignment (ClustalW) of spRNA26 homologs in *Methanosarcina* species with the predicted sORF indicated with a box. The similarity is shown in black-gray-white boxes (black symbols 100% similarity); additionally, the identity is shown in a gray bar and a nucleotide logo above the nucleotide alignment. (C) Amino acid alignment of sP26 homologs in *Methanosarcina* species. The protein sequence was generated by blasting the nucleotide sequence, creating every possible translation frame, and aligning every possible sequence to sp26 using ClustalW and further evaluated by hand. The similarity is shown in black-gray-white boxes (black symbols 100% similarity), and the identity is shown in a gray bar and a nucleotide logo above the alignment; *, indicates the peptide identified by LC-MS/MS; -, indicates a stop. All alignments were generated using Geneious Prime 2020.1.2 (https://www.geneious.com) and sequence data from NCBI with following Accession numbers: *M. acetivorans* C2A: AE010299, *M. barkeri* 227: CP009530, *M. barkeri* 3: CP009517, *M. barkeri* CM1: CP008746, *M. barkeri* MS: CP009528, *M. barkeri* Fusaro: CP000099, *M. barkeri* Wiesmoor: CP009526, *M. flavescens* E03.2: CP032683, *M. horonobensis* HB-1: CP009516, *M. mazei* C16: CP009514, *M. mazei* Goe1: AE008384, *M. mazei* LYC: CP009513, *M. mazei* S-6: CP009512, *M. mazei* SarPi: CP009511, *M. mazei* JL01: CP029709, *M. mazei* Tuc01: CP004144, *M. mazei* WWM610: CP009509, *M. siciliae* C2J: CP009508, *M. siciliae* HI350: CP009507, *M. siciliae* T4/M: CP009506, *M.* sp. 795: CP011449, *M.* sp. Kolksee : CP009524, *M.* sp. WH1: CP009504, *M.* sp. WWM596: CP009503, *M. thermophila*: AF358260, *M. thermophila* CHTI-55: CP009502, *M. thermophila* MT-1: AP017646, *M. thermophila* TM-1: CP009501, *M. vacuolata* Z-761: CP009520.
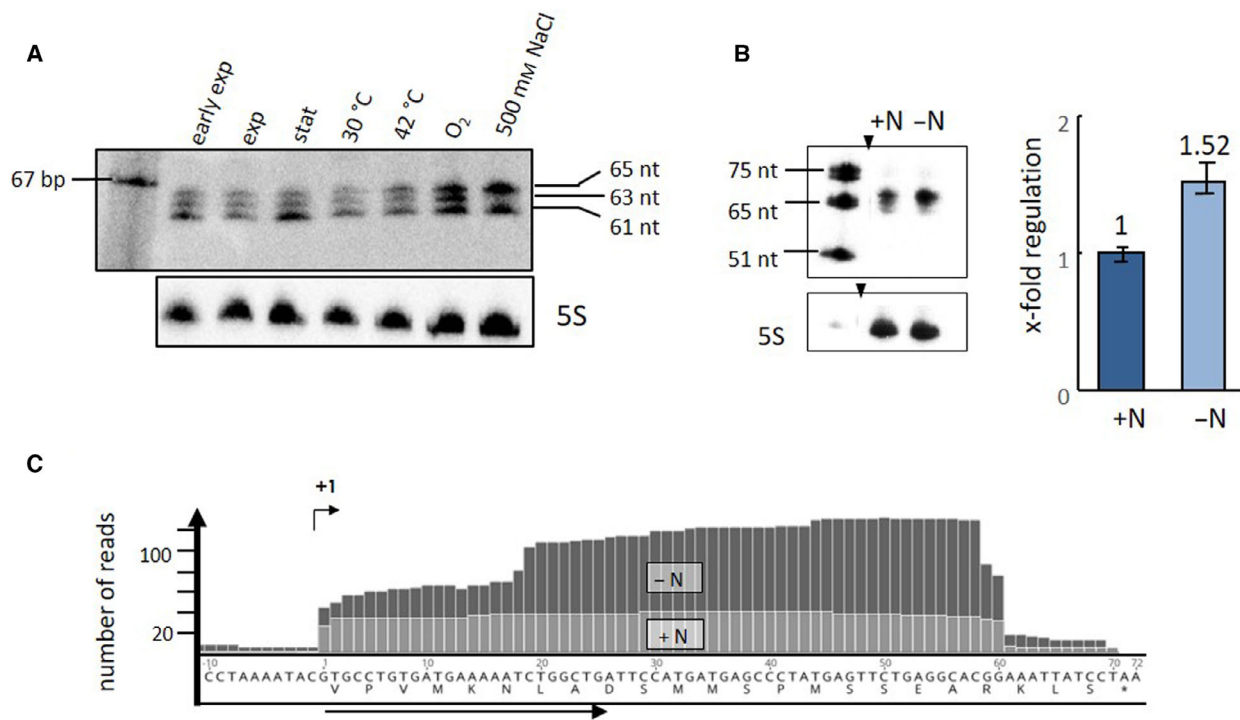

Fig. 2. spRNA26 transcription under different growth conditions. (A, B) Northern blot analysis under various conditions. (A) Northern blot analysis was performed as described in Methods using RNA derived from 50 mL cultures grown under N sufficiency at 37 °C to different growth phases (early exponential, exponential, and stationary) and exponential under different stress conditions as indicated. pGEM™ Conventional DNA Digest Marker PR-G1741 (Promega™, Fitchburg) was used as marker. (B) *M. mazei* grown exponential under N sufficiency (+N) and under N limitation (-N), the regulation was calculated based on eight independent biological replicas (error bars: SD, *n*: 12), one Northern blot is exemplary shown (triangles indicate splicing marks). (C) Differential RNAseq analysis using the Illumina technique was performed with RNA of *M. mazei* grown under N limitation and sufficiency (exponential growth phase) and aligned to the genome: darker plots, under N limitation; lighter plots, under N sufficiency.
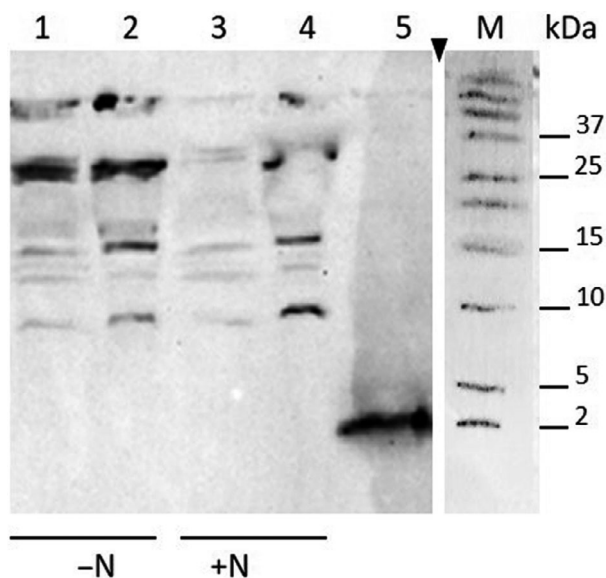
Fig. 3. Localization of sP26 under different nitrogen availabilities. Cells grown under nitrogen limitation (−N) and sufficiency (+N) were harvested during the exponential growth, disrupted, and centrifuged resulting in an insoluble fraction and crude extract. For detection of the localization, a western blot analysis using antibodies against sP26 was performed. Lanes 1 and 3 insoluble fractions; lanes 2 and 4 crude extracts, M Precision Plus Protein™ Dual Xtra Prestained Protein Standards (Bio-Rad, Feldkirchen). Lane 5 synthesized sP26 (control) (triangle indicates splicing mark).

sP26 is interacting with glutamine synthetase

In order to identify cell proteins directly interacting with sP26, we studied potential complex formation between sP26 N-terminally fused to a SUMO-tag (Sumo-His$_6$-sP26) and *M. mazei* cell extract proteins by pull-down experiments using affinity chromatography on Ni-NTA agarose for detecting complexes. His$_6$-SUMO-His$_6$-sP26 was heterologously expressed in *E. coli* (pRS1229, see Fig. S2c) and purified to an apparent homogeneity of 98% by Ni-NTA affinity chromatography as described in Methods. Purified His$_6$-SUMO-His$_6$-sP26 (1 mg) was incubated for 1 h at 4 °C with ~ 27.5 mg cell extract protein of *M. mazei* cells grown under N limitation. Subsequently, the protein mixture was applied to Ni-NTA agarose. After washing the chromatography material to remove all cell extract proteins, unspecifically binding to His$_6$-SUMO-His$_6$-sP26 or the Ni-NTA agarose, His$_6$-SUMO-His$_6$-sP26, and potential specifically interacting proteins were eluted in the presence of 250 mM imidazole. The respective elution fractions were analyzed by denaturing SDS/PAGE and subsequent silver staining. As exemplarily shown in Fig. 4A, only one distinct additional protein band corresponding to

an ~ 42 K_D protein was detected in at least three independent biological replicates. The elution fractions were concentrated and reanalyzed on an SDS/PAGE stained with Coomassie blue (Fig. 4A, panel 2). The additional protein band was analyzed, and the presence of glutamine synthetase (GlnA$_1$) was demonstrated in two independent biological experiments by LC-MS/MS analysis (see Materials and methods). Unspecific binding of *M. mazei* proteins to the SUMO-fusion protein or the affinity chromatography material was excluded by using purified SUMO-fusion protein and *M. mazei* cell extract and loading *M. mazei* cell extract to the affinity chromatography material followed by elution (see Fig. 4A panel 3). To confirm complex formation with GlnA$_1$, a reverse pull-down was performed. Strep-tagged GlnA$_1$ (pRS375) and SUMO-His$_6$-sP26 (pRS1229) were individually heterologously expressed in *E. coli*. Strep-GlnA$_1$ was purified to homogeneity by affinity chromatography using Strep-Tactin® sepharose® as described in Materials and methods. 1 mg purified Strep-GlnA$_1$ was incubated with the respective *E. coli* cell extract with SUMO-His$_6$-sP26 expressed (~ 50 mg) for 30 min at 4 °C followed by affinity chromatography using Strep-Tactin® sepharose®. Strep-GlnA$_1$ eluted from the column mainly in elution fraction 2, subsequent western blot analysis using antibodies directed against the His$_6$-tag confirmed the co-elution of SUMO-His$_6$-sP26 in the Strep-GlnA$_1$ elution fractions (Fig. 4B).

To obtain further experimental evidence for the interaction between sP26 and GlnA$_1$, microscale thermophoresis (MST) analysis was performed and the dissociation constant (K_D) determined. GlnA$_1$ was purified as N-terminal (His)$_6$-fusion protein as described by Ehlers *et al.* (2005) [32]. sP26 was chemically synthesized (ssP26) and nt-647-NHS labeled as described in Methods. Interactions between the labeled ssP26 (20 nM, based on the monomeric MW) and purified His$_6$-GlnA$_1$ in the range of 0.13 nM to 4.15 μM (calculated based on the monomeric MW) were analyzed by MST. Significant binding was observed in two independent biological experiments, each representing completely independent protein expressions and purifications and containing up to three technical replicates. One biological experiment is depicted in Fig. 5A; the K_D of 75 ± 40 nM was calculated based on integrating at least three technical replicates (panel 1), and panel 2 shows the respective calculation with one technical replicate. To verify the interaction, His$_6$-sP26 was expressed in *E. coli* (pRS1245, see Fig. S2e) and purified by fractionated ammonium precipitation followed by affinity chromatography on HisPur™ cobalt superflow agarose, and size exclusion using
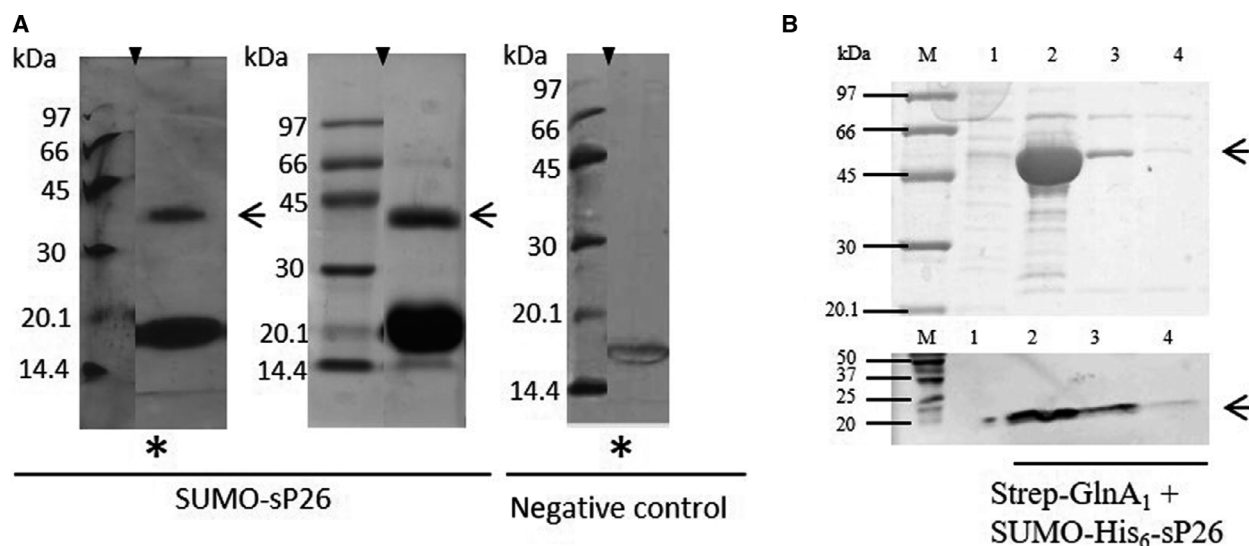
Fig. 4. Target identification of sP26 using cell extracts of *M. mazei* grown under nitrogen limitation. (A) Co-elution analysis of immobilized SUMO-sP26 and *M. mazei* cell extract proteins. One milligram purified SUMO-sP26 fusion protein was incubated with 27.5 mg of *M. mazei* cell extract (derived from 1 L culture grown under N limitation) for 1 h at 4 °C. As control served the SUMO-protein. Proteins were applied to 0.5 mL Ni-NTA agarose. Following the washing steps, SUMO-His$_6$-sP26 fusion protein and potentially interacting cell extract proteins were eluted from the affinity material in the presence of 250 mM imidazole and analyzed by SDS/PAGE of the respective co-elution fractions using silver (*) or Coomassie staining. The prominent band represents the SUMO-fusion protein (~ 20 kDa), and the potential interacting partner with a calculated size of 42 kDa is indicated with an arrow (triangles indicate splicing marks). (B) Reverse cochromatography analysis with purified Strep-GlnA$_1$: 1 mg purified Strep-GlnA$_1$ protein was pre-incubated with 50 mg *E. coli* cell extract containing overexpressed SUMO-His$_6$-sP26 and applied to 1 mL Strep-Tactin® sepharose® matrix (IBA). After 30-min incubation at 4 °C, the column was washed followed by elution of Strep-GlnA$_1$ and potentially interacting proteins in three elution fractions (1–3), which were analyzed by SDS/PAGE (upper panel) and western blot analysis (lower panel) using antibodies against the His-tag (Qiagen). Lane M (upper panel): Low-molecular-weight marker (GE Healthcare, Freiburg, Germany); Lane M (lower panel): Precision Plus Protein™-Dual Xtra Standard (Bio-Rad, Feldkirchen); lane 1, wash fraction; lanes 2-4, elution 1-3.

10 kDa filters as described in detail in Methods resulting in an ~ 95% purified protein fraction. When using purified and nt-647-NHS-labeled GlnA$_1$ (20 nM, monomeric) and purified His$_6$-sP26 in the range of 4 nM to 140 µM, a K_D of 54 µM ± 18 µM (monomeric) was calculated based on several technical replicates (Fig. 5B, panel 1), which was validated in a second independent biological replicate with three technical replicates. The specific binding with high affinity was further verified using purified untagged sP26 generated from purified SUMO-sP26 by removing the SUMO-fusion protein obtaining an even higher affinity for the tagless sP26 (see Fig. 5C, depicting one of three biological replicates). The fact that the K_D of the untagged sP26 version is two orders of magnitude lower indicates that the His-tag is hampering binding.

We performed NMR analysis of purified ^{15}N-labeled-untagged sP26 in the presence and absence of the target protein GlnA$_1$ and 2-OG (~ 12 mM). As depicted in the 2D ^{1}H ^{15}N HSQC spectrum (Fig. 6) showing signals of NH-sites in sP26 in the absence of the target, 18 out of 20 expected amide signals of sP26 are visible (shown in black); the spectrum is in complete agreement with the respective one of chemically synthesized sP26 [21]. In the presence of GlnA$_1$ (shown in red), seven amino acids of sP26 cannot be detected. They built the interaction surface, and their signals vanish either due to faster relaxation or due to chemical exchange-induced line broadening, further confirming specific binding of sP26 to GlnA$_1$ and allowing to identify the amino acids of sP26 that directly interact with GlnA$_1$ to be K5, S10, M11, M15, S16, S17, and E18. In addition, backbone amide signals of L7, M12, and K21 do not completely vanish but become significantly weaker upon addition of GlnA$_1$.

sP26 effects glutamine synthetase (GlnA$_1$) activity in addition to activation by 2-oxoglutarate

To identify the effects of sP26 on its target GlnA$_1$, we evaluated the glutamine synthetase activity in the absence and presence of the small protein. His$_6$-GlnA$_1$
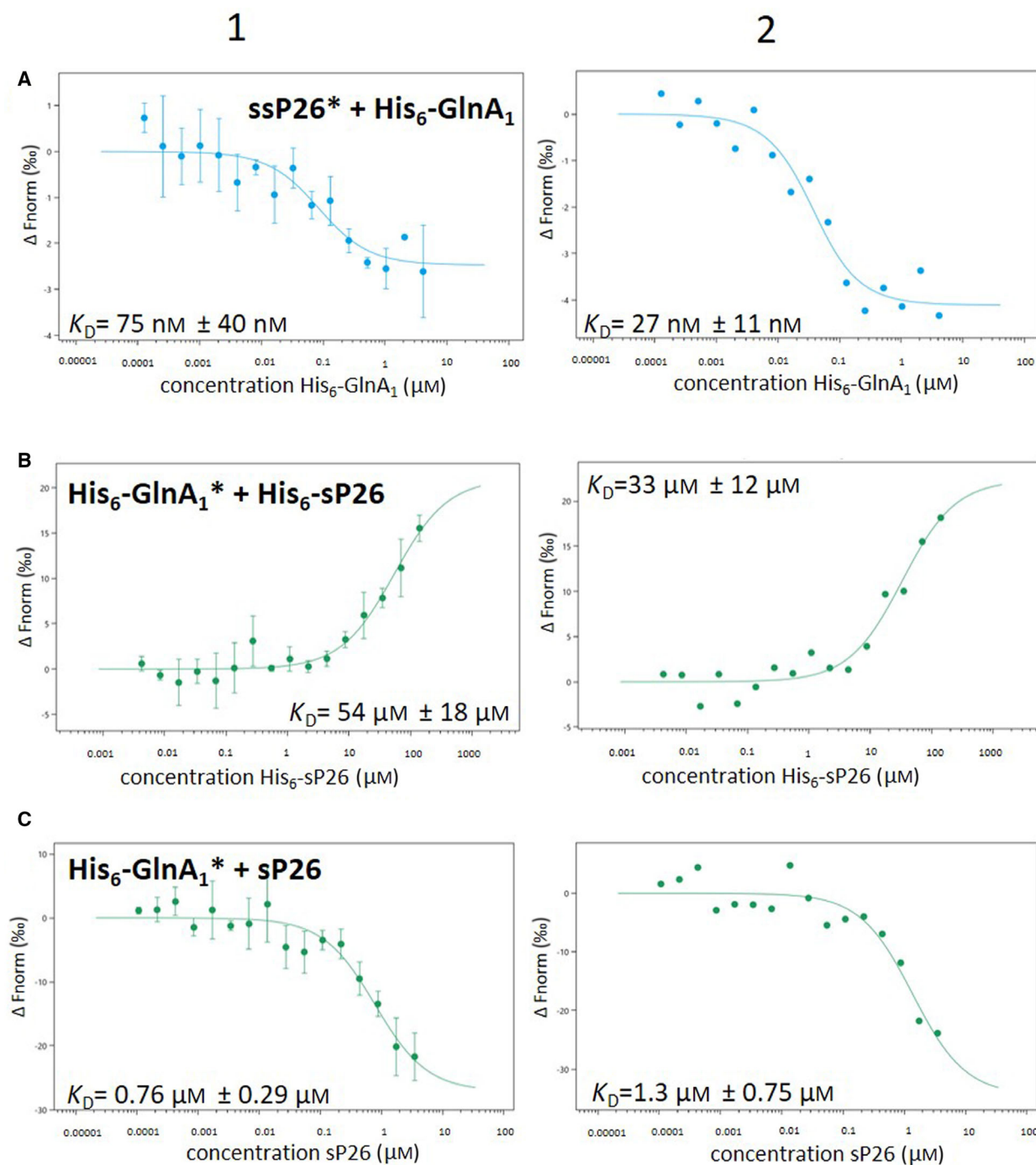
Fig. 5. Interaction studies of sP26 and His_6-$GlnA_1$ by MST analysis. (A) Synthesized sP26 (Davids Biotechnologie) (ssP26, 20 nM) was applied to a dilution series of purified and nonlabeled binding partner His_6-$GlnA_1$ whose concentration was varied between 0.13 nM and 4.15 μM (calculated based on the monomeric MW), resulting in a calculated K_D(His_6-$GlnA_1$): 75 nM ± 40 nM. (B, C) Purified and labeled His_6-$GlnA_1$ was analyzed with different variants of sP26. (B) His_6-sP26 was varied between 4 nM and 140 μM, calculated K_D(His_6-sP26): 54 μM ± 18 μM. (C) Untagged sP26 (derived from SUMO-sP26) was varied between 0.107 nM and 23 μM; calculated K_D(sP26): 0.76 μM ± 0.29 μM. In general (A–C), all measurements were performed as described in Methods, $n = 3$, error bars represent the standard deviation of one exemplarily chosen biological replicate (panel 1), with three technical replicates, one of which is depicted in panel 2.

Fig. 6. NMR analysis of ^{15}N-labeled sP26 in the presence of target protein GlnA$_1$. 2D ^{1}H^{15}N HSQC spectra of sP26 recorded at 298 K on a 600 MHz spectrometer. Spectrum of the pure protein is shown in black. Addition of unlabeled target protein GlnA$_1$ leads to the spectrum shown in red. In total, ten signals disappear or are significantly reduced in intensity. These include amino acids K5, L7, S10, M11, M12, M15, S16, S17, E18, and K21.
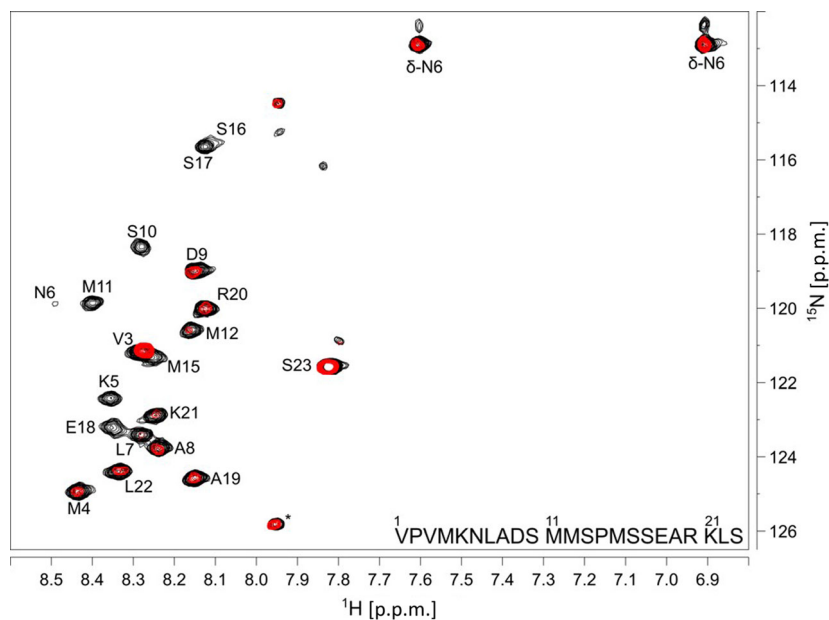
as well as His$_6$-sP26 and untagged sP26 were purified by affinity chromatography after heterologous expression in *E. coli*. GlnA$_1$ activity was determined by using the coupled optical test assay described by Shapiro [36], which couples the consumption of ATP by the conversion of ammonium and glutamate to glutamine catalyzed by glutamine synthetase to the oxidation of nicotinamide adenine dinucleotide (reduced) (NADH) by lactate dehydrogenase. In the course of optimizing the test assay, we observed that the buffer system is crucial and significantly effects the specific activity of glutamine synthetase. Using the originally reported MOPS buffer, which was used in the previous report on GlnA$_1$ activity [32], significantly lower specific activities were observed in comparison with other buffer systems. N-(2-hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid) (HEPES) buffer was proven to show the highest specific activities of GlnA$_1$ in the absence but also in the presence of the metabolite 2-OG, which has been demonstrated to significantly enhance GlnA$_1$ activity [32]. Due to the higher specific activity, we consequently exclusively used the HEPES buffer-based assay in the following experiments and purified GlnA$_1$ in HEPES buffer. In the presence of increasing amounts of purified His$_6$-sP26 (1.8–18 μM) (Fig. 7A), which were pre-incubated with 0.95 μM His$_6$-GlnA$_1$ (calculated based on the monomeric MW) for 5 min at RT in the test assay before starting by supplementing ATP, glutamine synthetase activity was significantly enhanced. The positive effect of His$_6$-sP26 on GlnA$_1$ activity was confirmed by at least five further independent biological replicates using

independent protein purifications. Further, the presence of contaminating adenosintriphosphatase activities in purified His$_6$-sP26 fractions that could interfere with the test assay was excluded. In general, stimulation of glutamine synthetase activity up to ninefold was obtained by His$_6$-sP26 depending on the respective protein purification (its quality condition) and consequential saturation level (see summarizing Table S1). One representative set resulting in 2.1-fold stimulation is exemplarily shown in Fig. 7B. Elucidating the effects of His$_6$-sP26 on glutamine synthetase activity in the presence of the known activator 2-OG (5 mM) clearly demonstrated that the obtained activating effect of sP26 is independent of the activation by 2-OG and stimulates the activity in addition to the activation due to 2-OG (see Fig. 7C, showing one exemplarily chosen biological activity set). Several independent biological replicates (independent protein purifications) confirmed the additional stimulation of GlnA$_1$ activity by His$_6$-sP26 in the presence of 2-OG up to fivefold (Table S1). For one fraction of purified His$_6$-sP26, we additionally tested the positive effects in the presence of 5 mM 2-OG until achieving apparent saturation by His$_6$-sP26, indicating that ~ 10 μM His$_6$-sP26 is saturating when using 0.95 μM GlnA$_1$ in the test assay (calculations based on GlnA$_1$ monomers) (Fig. 7F). Consequently, we propose that only at an excess free concentration of 10 μM His$_6$-sP26, all GlnA$_1$ subunits bound each one sP26 molecule. The stimulatory 2-OG-independent effect on GlnA$_1$ activity was further confirmed when using the purified untagged sP26 generated from SUMO-sP26. The untagged sP26,
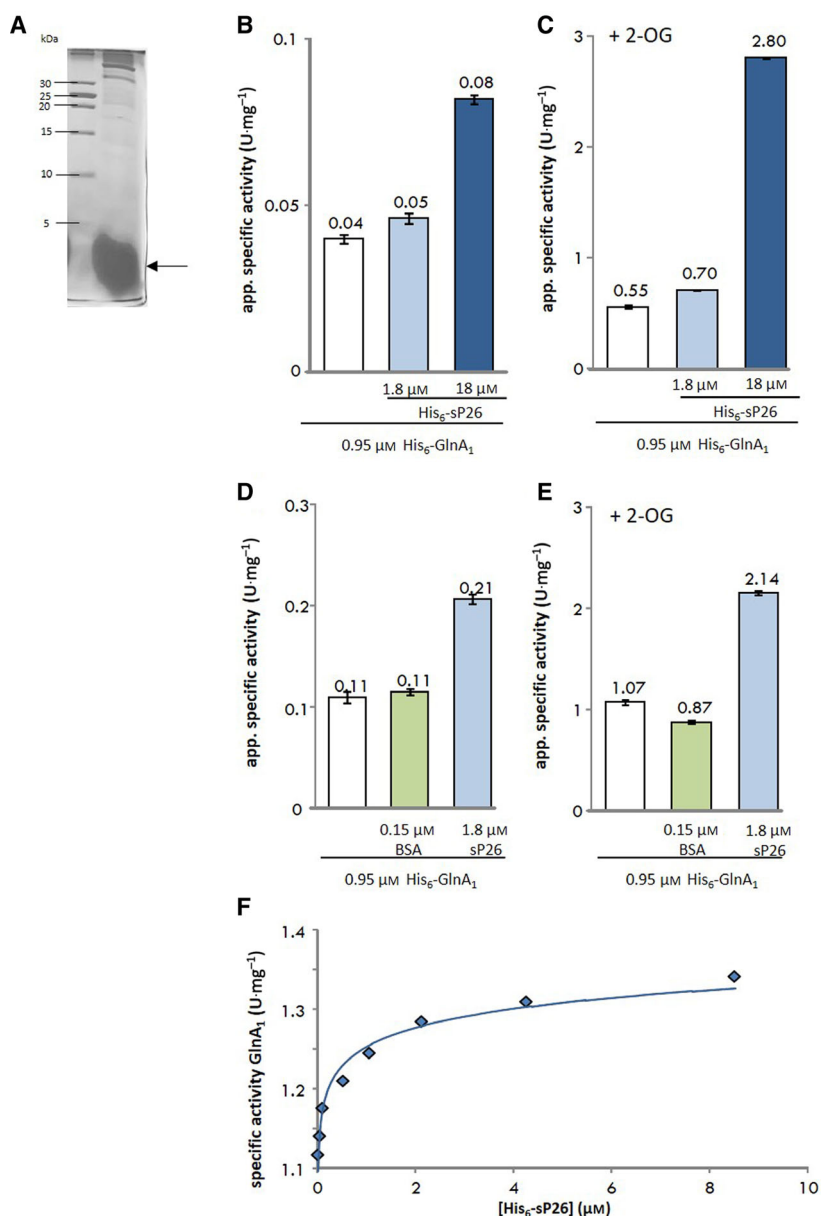
Fig. 7. Modulation of His_6-$GlnA_1$ activity by sP26 and 2-OG. (A) Purified His_6-sP26 was analyzed by SDS/PAGE (18% Tricine gel stained with colloidal Coomassie). (B, C, D, E) Glutamine synthetase activity of purified heterologously expressed His_6-$GlnA_1$ was determined at RT using the coupled optical enzyme assay described by Shapiro and Stadtman [36] but in a 50 mM HEPES system in the presence of varying amounts of His_6-sP26. (B) 0.95 μM purified His_6-$GlnA_1$ and varying amounts of His_6-sP26 (0, 1.8, and 18 μM) were pre-incubated in the absence of 2-OG for 5 min at RT prior to activity analysis (in general concentrations calculated based on the monomeric MW) (Error bars: SD, n:8). (C) The activity was determined under the same conditions as in (b) but in the presence of 5 mM 2-OG (Error bars: SD, n:7). The interaction was confirmed in numerous biological replicates summarized in Table S1. (D) 0.95 μM purified His_6-$GlnA_1$ and 3.98 μM purified untagged sP26 (derived from SUMO-sP26) or 1.5 μM BSA as control were pre-incubated in the absence of 2-OG for 5 min at RT prior to activity analysis (Error bars: SD, n:7). (E) The activity was determined under the same conditions as in (D) but in the presence of 5 mM 2-OG (Error bars: SD, n:14). In general, depicted are exemplarily data sets from one representative purification batch of $GlnA_1$ and His_6-sP26 / untagged sP26 of several independent biological replicates. (F) Saturation of the positive effect of sp26 on $GlnA_1$ activity. The activity of purified 0.95 μM His_6-$GlnA_1$ (calculated based on the monomeric MW) in the presence of 5 mM 2-OG was measured as described in Methods. Various amounts of purified His_6-sP26 were added and pre-incubated for 5 min prior the measurement. All data points were generated using the same His_6-$GlnA_1$ and His_6-sP26 purification batch.

however, showed a tenfold-enhanced activity compared with the His$_6$-tagged version of the protein (Fig. 7D, E), indicating that the tag is hampering the activity or the binding to GlnA$_1$.

To address the predicted higher oligomeric state of GlnA$_1$ in the presence of 2-OG [32], we analyzed the conformation and oligomeric state of purified GlnA$_1$ by structural integrity analysis using the Tycho system (NanoTemper, Munich, Germany) as described in Methods. The analysis showed two different transformation events first the unfolding of the monomer at 63 °C (black line) in the absence of 2-OG and additionally in the presence of the metabolite the dissociation of the oligomer at 48 °C (light gray line) (Fig. 8A). The addition of increasing concentrations of 2-OG resulted in increasing amounts of the oligomeric conformation of GlnA$_1$ and complementary decreasing amounts of the monomeric fraction (Fig. 8A: 5 mM 2-OG light gray line; Fig. 8B: 5, 7, 10, and 15 mM 2-OG from light gray to dark gray). Besides, binding of 2-OG generally resulted in higher stability of both conformations indicated by the significant shift to higher denaturation temperatures (e.g., shifting from 60 to 70 °C for the monomer). Evaluating different independent purifications, this variation in the oligomeric fraction of GlnA$_1$ might explain variability in activity and the range of stimulation observed by sP26 (Table S1). At present however, it is unclear whether the reason for this variability is based on biological or technical variables in the preparation.

sP26 is interacting with GlnK$_1$ and stimulates GlnA$_1$ activity in addition to activation by GlnK$_1$

Previous studies have shown that the PII-like protein GlnK$_1$ is modulating the GlnA$_1$ activity in *M. mazei* and GlnA$_1$ inhibition was predicted due to upshifts in N availability [32]. Consequently, we included GlnK$_1$ in our studies evaluating sP26 effects on GlnA$_1$ activity. Interaction studies between purified proteins by MST analysis clearly showed and verified the predicted specific interaction between GlnA$_1$ and GlnK$_1$. Using His$_6$-GlnA$_1$ and nt-647-NHS-labeled His$_6$-GlnK$_1$, a K_D of 1.5 μM ± 0.7 μM was calculated (always based on the monomeric MW) and verified in additional independent biological experiments including a reverse labeling experiment (nt-647-NHS-labeled His$_6$-GlnA$_1$ and His$_6$-GlnK$_1$, K_D = 17 μM ± 3 μM) (Fig. 9D,E). For all those experiments, the presence of 5 mM 2-OG did not change the binding affinity (K_D values, data not shown). Interactions between the second PII-like protein in *M. mazei* (GlnK$_2$) and GlnA$_1$ could not be detected by MST analysis, strongly arguing for GlnA$_1$-specific interaction with GlnK$_1$.

Elucidating potential interactions between GlnK$_1$ and sP26 showed in two independent biological experiments that nt-647-NHS-labeled His$_6$-GlnK$_1$ besides interacting with GlnA$_1$ similarly interacts with His$_6$-sP26 with high affinities and as well independent of 2-OG (see Fig. 9B, K_D = 2.5 μM ± 1.6 μM). The interaction was further confirmed with untagged sP26
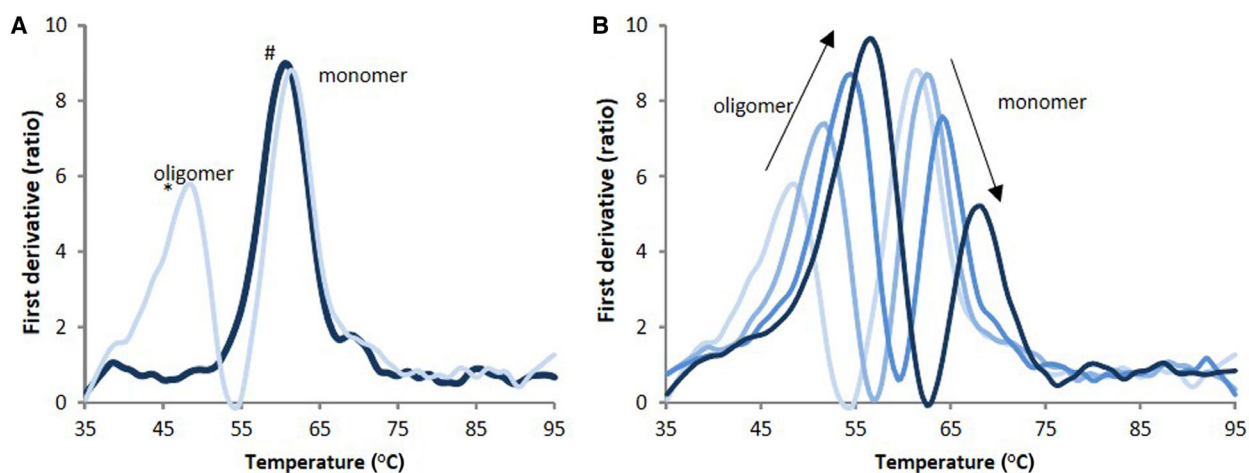
Fig. 8. Structural integrity analysis of purified His$_6$-GlnA$_1$ dependent on 2-OG. (A) 5.7 μM His$_6$-GlnA$_1$ was analyzed in the absence (black line) and presence of 5 mM 2-OG (gray line) using Tycho analysis (NanoTemper). In the presence of 2-OG, two transformation events occurred the dissociation of the oligomeric form (48 °C) and unfolding of the monomer (63 °C). (B) 5.7 μM His$_6$-GlnA$_1$ from the same purification was measured with increasing concentrations of 2-OG (5, 7, 10, and 12.5 mM —from light gray to dark gray line). With increasing concentration of 2-OG, the transformation events of the oligomeric and monomeric forms shift to higher temperatures indicating a more stable state.

Fig. 9. Interaction studies between His_6-$GlnK_1$ and sP26 by MST analysis and resulting modulation of His_6-$GlnA_1$ activity: (A) Labeled synthesized sP26 (20 nM) and purified His_6-$GlnK_1$ varied between 0.13 nM and 4.15 μM (calculated based on the monomeric MW) were measured resulting in an estimated $K_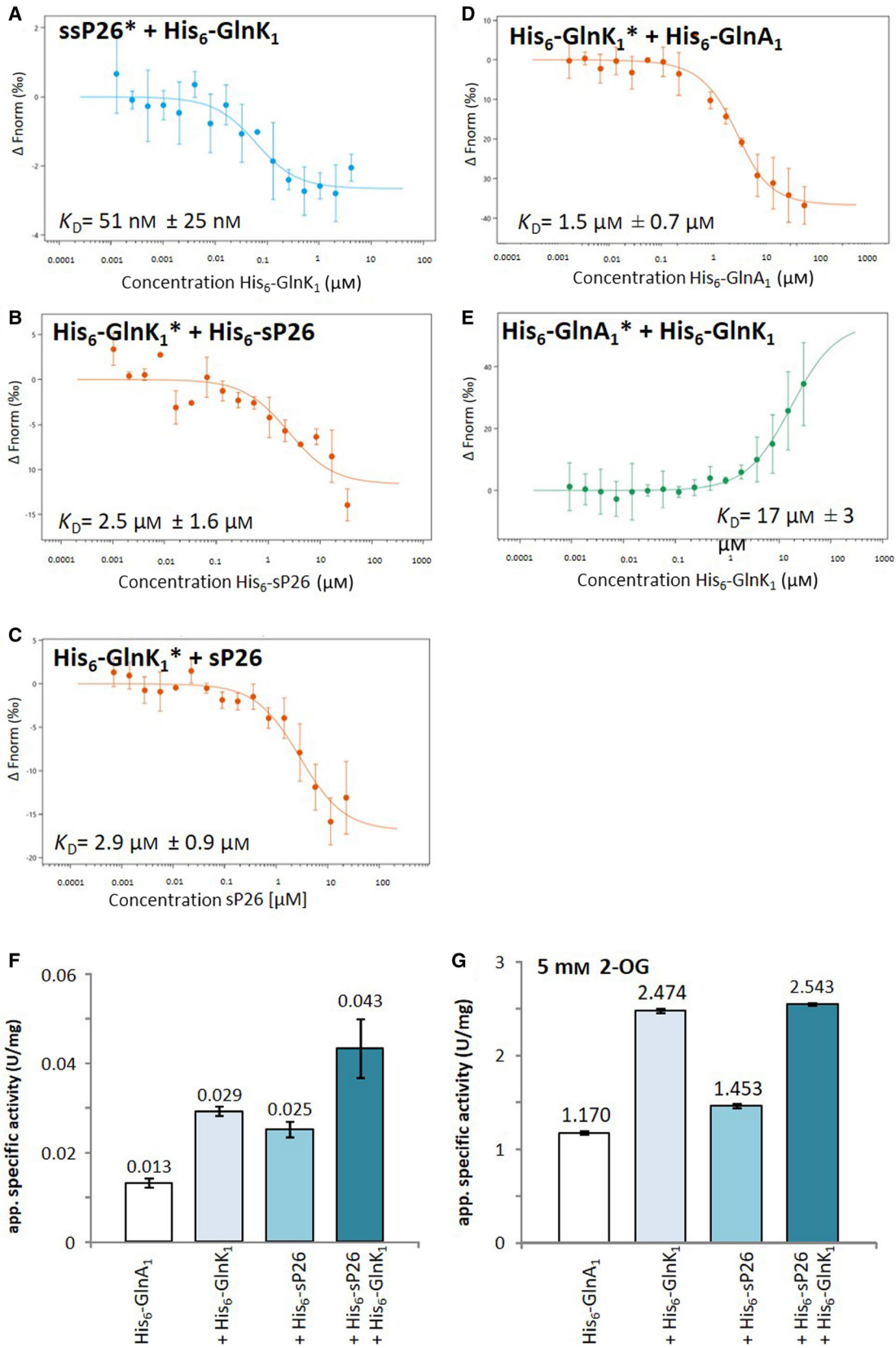D$($His_6$-$GlnK_1$): 51 nM ± 25 nM ($n = 2$). (B) Labeled His_6-$GlnK_1$ (20 nM) and purified His_6-sP26 varied between 4.3 nM and 34 μM; estimated K_D (His_6-sP26): 2.5 μM ± 1.5 μM ($n = 2$). (C) Labeled His_6-$GlnK_1$ (20 nM) and untagged sp26 varied between 0.107 nM and 23 μM; estimated K_D (sP26): 2.9 μM ± 0.9 μM ($n = 3$). (D) Purified and labeled His_6-$GlnK_1$* was applied to purified His_6-$GlnA_1$ whose concentration varied between 1.7 nM and 55.5 μM (calculated based on the monomeric MW). Calculated K_D (His_6-$GlnA_1$): 1.5 μM ± 0.7 μM (with $n = 2$, error bars represent the standard deviation of 1 biological experiment with 2 technical replicates). (E) Purified labeled His_6-$GlnA_1$* (20 nM) was analyzed with purified His_6-$GlnK_1$ with a concentration range from 6 nM to 105 μM. Calculated K_D (His_6-$GlnK_1$): 17 μM ± 3 μM ($n = 3$, error bars represent the standard deviation of 1 biological experiment with three technical replicates). In general (A, B, C, D, E), out of several biological replicates, one respective biological replicate is exemplarily shown in panel 1—based on two or three technical replicates. (F, G) Glutamine synthetase activity of His_6-$GlnA_1$ (0.95 μM) determined as described in Fig. 7 pre-incubated at RT for 5 min, with His_6-sP26 (2.9 μM), His_6-$GlnK_1$ (0.65 μM), and His_6-sP26 (2.9 μM) plus His_6-$GlnK_1$ (0.65 μM) in the absence (F) or presence of 5 mM 2-OG (G) (generally, concentrations calculated based on the monomeric MW). Depicted are exemplarily data from one representative purification batch of His_6-$GlnA_1$, His_6-sP26, and His_6-$GlnK_1$ (standard deviation of three technical replicates). The interaction was confirmed in a second biological replicate (including three technical replicates) (Table S1).

(Fig. 9C, K_D = 2.9 ± 0.9 μM, representing one out of three biological experiments), as well as by an independent reverse labeling experiment analyzing the interaction between His_6-$GlnK_1$ and nt-647-NHS-labeled-synthesized sP26 (Fig. 9A, K_D = 51 nM ± 25 nM). Interaction between sP26 and $GlnK_2$ was not detected. Overall, these findings strongly argue that $GlnK_1$ modulation has to be taken into further account when analyzing the sP26 effects on $GlnA_1$ activity.

Consequently, we included $GlnK_1$ in evaluating the enhancing effects of sP26 on $GlnA_1$ activity. Aiming to first validate the reported inhibitory effects of $GlnK_1$ on $GlnA_1$ activity in the new buffer system, we reanalyzed the $GlnK_1$ effect in the modified activity assay. Using the originally MOPS-based system, we confirmed the inhibitory effects of $GlnK_1$ on $GlnA_1$ activity; however, changing the buffer system (to NaH_2PO_4 or HEPES) resulted in positive effects of $GlnK_1$ on $GlnA_1$ activity independent of the presence of 2-OG (data not shown). Using the HEPES-based test assay, we obtained strong evidence in two biological independent replicates (each including three technical replicates) that the stimulating effect of sP26 on glutamine synthetase activity is not only independent of 2-OG but also independent and in addition to the stimulating effect of $GlnK_1$ (Fig. 9F, depicting one of two biological replicates). In the presence of 2-OG, the additive effects of $GlnK_1$ and sP26 were not anymore detectable most likely due to limiting concentrations of NADH and ATP in the test assay (saturation of the activity) (Fig. 9G).

In order to evaluate potential complexes formed between the three proteins, purified His_6-$GlnK_1$ (77.7 μg), His_6-$GlnA_1$ (168.5 μg), and His_6-sP26 (24 μg) were pre-incubated at RT for 5 min and analyzed by size-exclusion chromatography (SEC) in the absence and presence of 5 mM 2-OG in the buffer as described in Methods (ENrich 650 column; Bio-Rad, Feldkirchen, Germany). Analysis of the respective SEC fractions was performed via LC-MS/MS as described in Methods. The results summarized in Fig. 10 clearly show that the addition of 2-OG facilitated the identification of both sP26 and $GlnK_1$ in the highest molecular weight (MW) fraction (> 600 kDa) besides $GlnA_1$, providing further evidence toward the close association of these three proteins in the presence of 2-OG.

Discussion

The recently recognized existence of previously overlooked hidden small proteins in bacterial and archaeal genomes became an emerging research field. Here, we report on the physiological role of the first small archaeal protein in *M. mazei*. Concordantly we obtained strong evidence that sP26 interacts specifically and with high affinity with the glutamine synthetase ($GlnA_1$) in *M. mazei* the key enzyme of ammonium assimilation under N limiting growth conditions. Interaction with sP26 directly effects and enhances the activity of $GlnA_1$ independently and in addition to the inducing metabolite 2-OG and the modulating PII-like nitrogen sensory protein $GlnK_1$.

Ammonium assimilation under N starvation via the GS/GOGAT pathway is one of the major intersections in central N metabolism (see Fig. 11, inset). Consequently, in all organisms synthesis and activity of glutamine synthetase (GS) is strictly regulated by N availability [34,37,38]. The GS in prokaryotes and eukaryotes can be classified in three different families, GSI, GS II, and GSIII, all of which are representing large homo-oligomeric complexes with 8, 10, or 12 monomers. The decameric GSII is present in
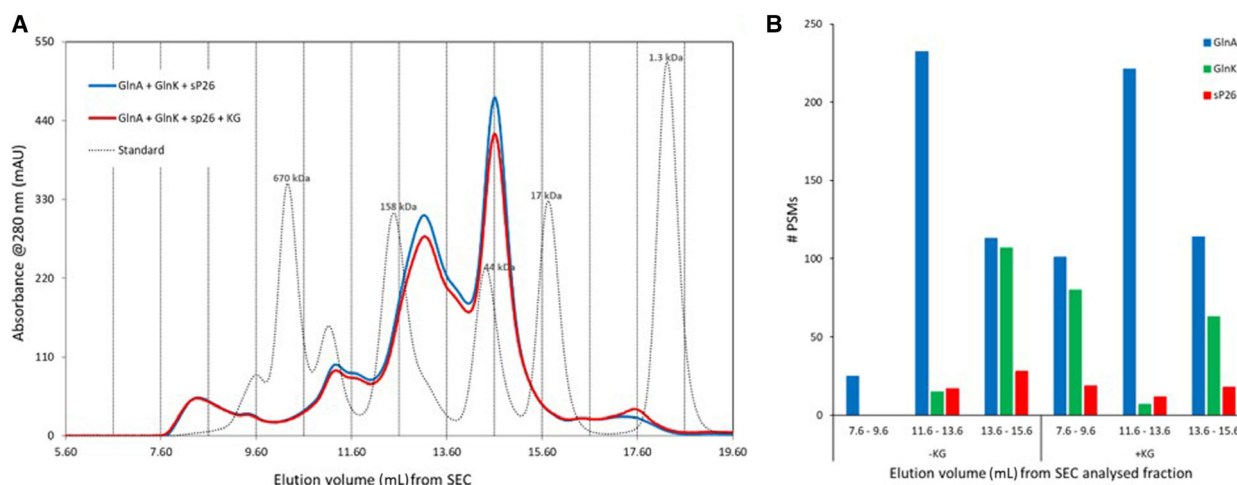
Fig. 10. Gel filtration and LC-MS/MS analysis of complex formation between purified *M. mazei* sp26-His6, His6-GlnA1, and His6-GlnK1 proteins in the absence or presence of 2-OG. (A) Gel filtration analysis was performed on a ENrich 650 column (Bio-Rad) using 50 mM Tris/HCl with 150 mM NaCl and a pH 7.5 as buffer system in the presence and absence of 5 mM 2-OG (KG) and a flow rate of $1.0 \ mL \cdot min^{-1}$. Proteins were detected by monitoring the absorbance at 280 nm. Calibration of the column was performed using the gel filtration mass standard (Bio-Rad Laboratories) containing thyroglobulin (670 kDa), IgG (150 kDa), myoglobulin (44 kDa), ovalbumin (17 kDa), and vitamin B12 (1.35 kDa). Elution volumes of standard proteins are indicated. 3.2 µmol His_6-$GlnA_1$ (52.8 kDa), 5.2 µmol His_6-$GlnK_1$ (14.9 kDa), and 4.4 µmol His_6-sP26 (5.5 kDa) were pre-incubated in a total volume of 250 µL for 5 min at RT before injection. (B) Protein identification for pooled SEC fractions was performed via LC-MS/MS analysis. The peptide spectral match (PSM) counts for the three proteins of interest (His_6-$GlnA_1$, His_6-$GlnK_1$, and His_6-sP26) are displayed indicating in which fractions the proteins were identified and providing a estimation of the potential protein abundance. Co-incubation of the proteins with 2-OG allowed for the identification of both sP26 and $GlnK_1$ in the highest MW fraction (7.6–9.6 mL).

eukaryotes, whereas GSI comprising two subtypes, GSI-α and GSI-β, and GSIII are found in bacteria and archaea mostly forming dodecamers [35]. GSI are typically feedback inhibited by the end products glutamine and AMP. In addition and in contrast to GSI-α, GS of subtype GSI-β are further inhibited in response to N sufficiency by a specific covalent modification of a tyrosine residue near the active side (adenylation by an adenylyltransferase) [35,39,40]. In *M. mazei,* the essential GS, $GlnA_1$, does not exhibit an adenylylation site nor have homologues of adenylyltransferase been identified in the *M. mazei* genome [39,41] excluding post-translational regulation of *M. mazei* GS in response to N by covalent modification [35,39,40]. Accordingly, $GlnA_1$ represents a GS of the GSI-α subdivision. The gene encoding the essential GS in *M. mazei* ($glnA_1$) is under the direct transcriptional control of the global nitrogen repressor NrpR and thus exclusively transcribed under N limitation [27,29]. In addition, the expression of $glnA_1$ is post-transcriptionally regulated by $sRNA_{154}$, which significantly stabilizes $glnA_1$ transcripts under N limitation [28]. In respect to post-translational N-dependent regulation of $GlnA_1$ activity, we have shown in the past that $GlnA_1$ activity is directly stimulated by the cellular metabolite 2-OG [32]. Due to severe reduction in glutamate dehydrogenase activity under N limitation, the cellular concentrations of 2-OG increase drastically; consequently, the cellular 2-OG concentration is considered to reflect the internal signal for N starvation [32]. Perceiving the signal for N starvation as a result of increasing intracellular 2-OG concentrations is also proposed for other autotrophically growing microorganisms as reported, for example, for cyanobacteria [34,42,43], and has been demonstrated for the post-translational modulation of nitrogenase activity by PII-like proteins ($NifJ_1$ and $NifJ_2$) in response to N availability (switch on, switch off mechanism by direct protein interaction) in *Methanococcus maripaludis* [44,45]. As known for other methanoarchaea, the cellular 2-OG concentration in *M. mazei* also mediates the N status to the transcriptional regulatory machinery. Under N limitation and resulting high internal 2-OG concentration, 2-OG binds to the global repressor NrpR, significantly lowering the binding affinity of NrpR to its respective operator. Consequently, NrpR leaves the operator and transcription can be initiated by RNA polymerase [29]. This further emphasizes the central role of 2-OG in sensing and transmitting the internal N status. In addition, a PII-like protein, $GlnK_1$, allows fine-tuning
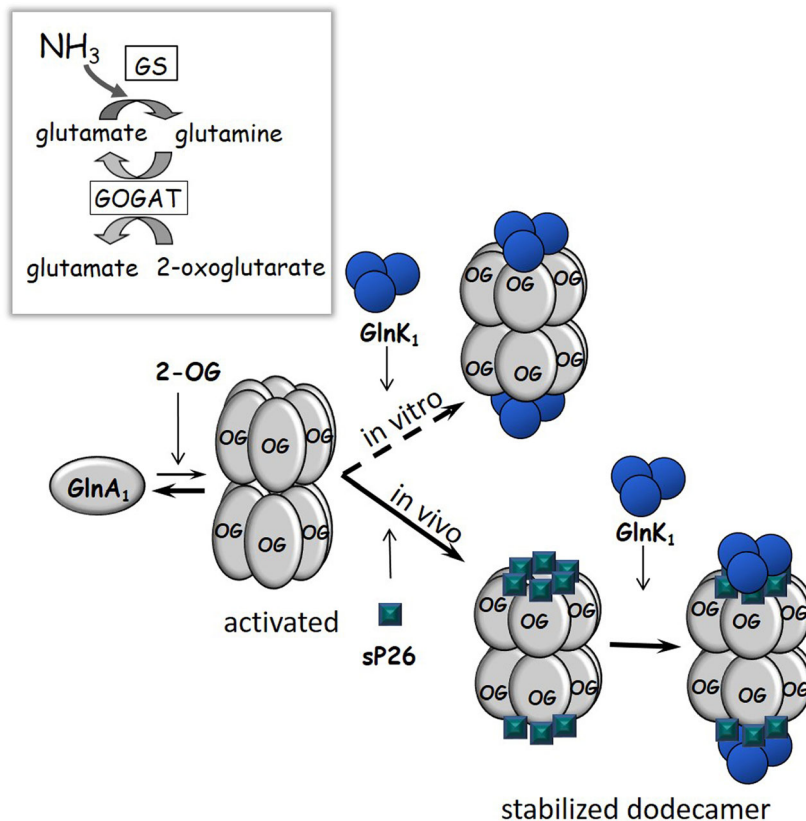
Fig. 11. Hypothetical model of GlnA$_1$ regulation in *M. mazei*. GlnA$_1$ forms an unstable dodecameric oligomer by binding 2-OG. sP26 is proposed to effect GlnA$_1$ oligomeric structure in addition to 2-OG by allowing a higher stability of dodecamers or inducing tighter dodecamers. GlnK$_1$ binds to GlnA$_1$ most likely mediated by interacting with sp26 resulting in further stabilization of the dodecamer. However, at least *in vitro* GlnK$_1$ binds to GlnA$_1$ also in the absence of sP26. sP26 might in addition directly or indirectly effect the GlnA$_1$ coupling with GOGAT, resulting in transferring the amino group of glutamine onto 2-OG, as depicted in the insert.

control of the glutamine synthetase activity under changing N availabilities and was predicted to inhibit GlnA$_1$ activity due to an ammonium upshift after a period of N limitation [32].

Realizing that a wealth of small proteins exists in *M. mazei* identified by systematic global screens for translated small proteins [7,8,19], now unraveled that GlnA$_1$ regulation in *M. mazei* is even more complex and argues for a mechanistically novel post-translational regulation by a small protein. Using different approaches, we obtained conclusive experimental evidence that the small protein sP26 (molecular mass 2.7 kDa), which is upregulated under N limitation, interacts with GlnA$_1$ in *M. mazei* and simultaneously increases its enzyme activity independently and in addition to the stimulating metabolite 2-OG and modulating GlnK$_1$. (a) Pull-down experiments demonstrated that purified *M. mazei* SUMO-His$_6$-sP26 directly interacts with GlnA$_1$ in cell extracts, which was verified by a reverse pull-down, demonstrating coelution of SUMO-His$_6$-sP26 when purifying Strep-GlnA$_1$ (Fig. 4). (b) High binding affinity between sP26 and GlnA$_1$ was observed using MST analysis (including reverse labeling, see Fig. 5). This specific binding was not depending or affected by 2-OG. (c) NMR analysis of sP26 in the presence and absence of GlnA$_1$ identified seven amino acids specifically interacting with GlnA$_1$ (Fig. 6). (d) sP26 significantly enhances glutamine synthetase activity (Fig. 7). The EC50 for sP26 in respect to GlnA$_1$ activity enhancement was calculated to be 0.43 μM, which is in the same range as the K_D determined for the untagged sP26. (e) This activity stimulation was independent and additive to the 2-OG activation and GlnK$_1$ modulation (Fig. 9F). (vi) First SEC analysis provided evidence toward close association of GlnA$_1$, GlnK$_1$, and sP26 in the presence of 2-OG (Fig. 10) indicating a quaternary structure.

Interestingly, in cyanobacteria, where 2-OG is also reflecting the internal N status, small proteins modulating GS activity have been reported several years ago [46,47]. However, these proteins are in the range of 7 and 17 kDa and have been demonstrated to inhibit GS activity by protein–protein interaction due to an ammonium upshift after N starvation [46]. Those so-called inactivating factors (IF7 and IF17) are expressed under N sufficiency combining transcriptional regulation but also post-transcriptional regulation by regulatory RNAs, an antisense RNA (NsiR4) in case of IF7 [48] and a recently discovered glutamine-binding riboswitch in case of IF17 [49].

As already observed by Ehlers *et al.* (2005) [32], the PII protein $GlnK_1$ is able to modify $GlnA_1$ activity. Direct protein interaction has been exclusively shown by pull-down approaches without quantification. Here, we succeeded to demonstrate by MST analysis that $GlnK_1$ binds to $GlnA_1$ with high affinity and further excluded that the interaction is dependent on 2-OG. In contrast however, to the previous report [32] we obtained strong evidence that $GlnK_1$ is activating $GlnA_1$ activity when the enzyme activity is determined in a HEPES buffer-based activity assay (Fig. 9F,G). The observation of contradictory effects of $GlnK_1$ on $GlnA_1$ activity depending on the buffer systems suggests that the buffer system is highly influencing/effecting $GlnA_1$ activity. We can only speculate that the different buffer systems, salt, and/or differences in the pK_S might affect the interaction between $GlnA_1$ and $GlnK_1$, and/or the oligomeric conformation of $GlnA_1$. However, since in all buffers tested except MOPS buffer used in the 2005 [32] $GlnA_1$ activity was enhanced by $GlnK_1$ presence, we conclude that $GlnK_1$ is indeed activating $GlnA_1$ activity. We clearly demonstrated that sP26 also binds to $GlnK_1$ with high affinity and in a 2-OG independent manner (Fig. 9A–C). Thus, in future it will be important to elucidate the structural changes in $GlnA_1$ in the presence of $GlnK_1$ and sP26 (e.g., by NMR analysis), as well as the quaternary complex formation between all three proteins in the presence and absence of 2-OG in more detail and determine the respective ratio as well as the respective oligomeric conformation of $GlnA_1$.

Hypothetical model for post-translational regulation of $GlnA_1$ by sP26

Since strong interactions between sP26 and $GlnA_1$ have been observed, and it is highly unlikely that sP26 displays a catalytic activity for, for example, covalent modification of $GlnA_1$, we propose that the obtained sP26-dependent induction of $GlnA_1$ activity is most likely due to effects on $GlnA_1$ oligomeric structure or stability. Moreover, sP26 might in addition directly or indirectly effect $GlnA_1$ interaction with GOGAT. Potential effects on the $GlnA_1$ /GOGAT interaction by sP26 are particularly attractive, since we have detected peptides derived from both large subunits of GOGAT in several sP26 pull-down experiments in the LC-MS/MS analysis.

Based on our findings, we hypothesize the following model depicted in Fig. 11. In the presence of 2-OG, $GlnA_1$ nearly exclusively expressed under N starvation forms higher oligomeric structures (dodecamers) resulting in $GlnA_1$ activation. However, the dodecamers are not very stable. Only in the additional presence and upon interaction with sP26, which is induced under N starvation, $GlnA_1$ dodecamers change into a more stable (tighter) complex and consequently induce $GlnA_1$ activity in addition to 2-OG. $GlnK_1$ only expressed under N limitation binds to $GlnA_1$ most likely mediated by interacting with sp26 resulting in further stabilization of the dodecamer. *In vivo*, sP26 is always present under N limitation; however, *in vitro* $GlnK_1$ was shown to bind to $GlnA_1$ and stimulate its activity also in the absence of sP26. The first complex analysis by size-exclusion chromatography indicates quaternary complexes of all three proteins are formed only in the presence of 2-OG. Thus, in future studies the quaternary complexes as well as the hypothesized change in the dodecameric structure (most likely two hexameric ring structure face to face) have to be proven, for example, by NMR analysis and in combination with hydrogen-deuterium exchange mass spectrometry (HDX-MS) to identify the interaction sites of all proteins at the protein surfaces. Moreover, since sP26 is highly conserved in the Methanosarcinas, it is attractive to speculate that for other methanoarchaeal glutamine synthetases, modulations by the sP26 homologs might occur.

Materials and methods

Strains and plasmids

All plasmids used have been constructed as follows, confirmed by sequencing and are summarized in Table S2. pRS1242 was generated for overproducing sP26 N-terminally fused to a $(His)_6$-tag (His_6-sP26) in *M. mazei*. A DNA fragment including the respective gene was commercially synthesized (Eurofins Scientific, Nantes, Luxemburg) as follows: The gene was placed under the control of the promoter and the ribosome-binding site of *mcr*B [50], followed by an additional ATG as new translation start and sORF26 with an N-terminal His_6-Tag. Downstream of the sORF26 sequence the methanoarchaeal transcriptional terminator was added. Further, a flanking *Sac*I restriction site (5′) and a *Kpn*I restriction site (3′) were added resulting in a 379-bp fragment. The synthesized DNA fragment was inserted into pEX (Eurofins Scientific), the corresponding plasmid designated pRS1209 (Fig. S2a). Using *Sac*I and *Kpn*I (NEB, Schwalbach, Germany), the respective 373-bp fragment was generated and ligated into the *Sac*I and *Kpn*I-restricted pWM321 [51] resulting in plasmid pRS1242 (Fig. S2d).

For the production of sORF26 fused to a SUMO-tag, pRS1209 as template and the primers sORF26_1 His-for (5′-ATGCATCATCATCATCATCACGTGCCTGTGATG AAAAATCTG-3′) and sORF26_1 rev (5′-AAAAAATT

AAGGATAATTTCCGTGCCTC-3′) were used in a PCR to generate a 100-bp fragment encoding sORF26 with an N-terminal His$_6$-Tag. This fragment was TA-cloned into the pET-SUMO (Invitrogen, Carlsbad, CA, USA) resulting in pRS1229 (Fig. S2c). The respective resulting fusion protein is an N-terminal SUMO-Tag fused to His$_6$-sP26 (16.3 kDa).

pRS1228 was constructed by amplifying sORF26 from pRS1209 using primers sORF26_1 rev (AAAAAATTAAG-GATAATTTCCGTGCCTC) and sORF26_1 for (GTG CCTGTGATGAAAAATCTGGCTG) and TA cloning the respective fragment into pET-SUMO (Invitrogen) vector. In contrast to pRS1229, the respective construct pRS228 provides no additional His$_6$-tag between SUMO and sP26 (Fig. S2d).

pRS1245 for the expression of His$_6$-sORF26 in *E. coli* was generated as follows: sORF26 was PCR amplified using pRS1209 as template and the primers sORF26_1_-for_*Nde*I (5′-CATATGGTGCCTGTGATGAAAAATCTG GCTG-3′) and sORF26_1_rev_*Nde*I (5′-CATATG-GAAAAAATTAGGATAATTTCCGTGCCTC-3′), which add flanking *Nde*I restriction sites (underlined). The resulting 91-bp product was TA-cloned into the vector pCR®II-TOPO® (Life Technologies, Darmstadt, Germany) creating the plasmid pRS1244. After *Nde*I restriction of pRS1244 (NEB), the generated 84-bp fragment was cloned into the *Nde*I-linearized vector pET28a(+) (Merck KGaA, Darmstadt, Germany) resulting in pRS1245 where the N-terminal vector-derived His$_6$-Tag followed by a thrombin site is fused to sORF26 (Fig. S2e). The corresponding small protein has a calculated molecular mass of 5.45 kDa.

pRS375 was constructed by amplifying *glnA$_1$* from pRS196 [32] using a forward primer replacing the N-terminal (His)$_6$-tag by a Streptag (Mm GlnA H/S for 5′-CCATGGGCTGGAGCCACCCGCAGTTCGAAAAAAG CAGCGGCCTGGTGCCGCGC-3′) and adding a *Nco*I restriction site and the reverse primer (Mm GlnA H/S rev 5′-CCGCCGCGCCCGCCGCCCGCCGCAAGCTTGTC-GACGGAGCTCGAATTCGGATCC-3′) adding a *Bam*HI restriction site. The obtained 1.4 kbp PCR product was cloned into pRS196 using *Nco*I and *Bam*HI restriction sites. The resulting plasmid was designated pRS375.

For expression in *E. coli*, all plasmids were transformed into *E. coli* BL21(DE3) (Invitrogen) or *E. coli* BL21 pRIL (Stratagene, La Jolla, Ca, USA) using the method of Inoue *et al.* (1990) [52]. For overproduction in *M. mazei,* the plasmid pRS1242 was transformed into *M. mazei* DSM3647 by liposome-mediated transformation as described by Ehlers *et al.* (2005) [53].

Growth analysis of *M. mazei*

All *M. mazei* strains were grown at 37 °C in 50 mL or 1-L enclosed serum bottles under anaerobic conditions, with the gaseous phase containing N$_2$ and CO$_2$ (80/20) (Air

Liquide, Paris, France). The minimal medium was supplemented with 150 mM methanol and 40 mM acetate as carbon and energy sources, as described in Veit *et al.* (2006) [54]. For growth under N limitation, the N$_2$ from the gas atmosphere served as the sole N source, and for N sufficiency, the medium was supplemented with 10 mM NH$_4$Cl. For different stress conditions, the cultures were grown under various conditions, at 30 °C, at 42 °C, under oxygen stress (20 mL sterile air was added to the gas atmosphere at a turbidity of 0.25), and under salt stress (additional 500 mM NaCl). Growth was generally monitored by determining the turbidity of the cultures at 600 nm.

RNA isolation, northern, and RNAseq analysis

RNA isolation was performed using Rotizol reagent (Carl Roth GmbH, Karlsruhe, Germany) following the manufacturer's protocol. Subsequently, a DNaseI treatment and phenol/chloroform/isoamyl alcohol precipitation were performed as described in Nickel *et al.* (2013) [55]. Northern blot analysis was performed as described by Jager *et al.* (2009) [27]. RNAs were detected using 5′-^{32}P-labeled ssDNA oligo probes against the sORF26 (5′GCCTGTG ATGAAAAATCTGGCTG3′) and against the 5S rRNA (5′CGCACACTTCAGTACAGTAAGGAA3′). The pGEM™ Conventional DNA Digest Marker PR-G1741 (Promega™, Fitchburg, MA, USA) was used as size standard, and the bands were quantified by the AIDA Software (Raytest, Staubenrath, Germany).

The Illumina sequencing data were obtained from *M. mazei* cultures grown under N starvation and N sufficiency (15 mM) which were harvested in the exponential growth phase. RNA isolation was performed as described above, and RNAseq was performed as described by Prasse *et al.* (2017) [28].

Localization of sP26 in *M. mazei* cells

Methanosarcina mazei was grown anaerobically in complex medium under N sufficiency or limitation, as described in Ehlers *et al* 2005 [53]. The optical turbidity was measured at 600 nm with spectrophotometer DU 640 (Beckmann Coulter, Krefeld, Germany). At mid-exponential growth phase, cells were harvested at 2831 *g* and 4 °C for 30 min, resuspended in 50 mM Tris (pH 6.8), and disrupted using a dismembrator (3 min at 1600 r.p.m.). Following centrifugation for 30 min at 15 700 *g* and 4 °C, the supernatant (soluble fraction) and pellet (insoluble fraction) were used for western blot analysis using antibodies directed against the synthetic sP26.

Purification of SUMO-sP26 fusion proteins and sP26, and co-elution analysis

SUMO-His$_6$-sP26 was purified from 1 L culture of *E. coli* BL21 (DE3) carrying the plasmid pRS1229 grown in LB at

37 °C. At a turbidity of 0.6 at 600 nm, the culture was supplemented with 100 μM IPTG and further incubated for 2 h at 37 °C with rigorous shaking. After harvesting (4000 *g* at 4 °C for 30 min) and resuspending in buffer A (50 mM NaH_2PO_4, 300 mM NaCl, pH 8), the cells were disrupted using a French pressure cell two times (at 4.135×10^6 N·m^{-2}) followed by centrifugation at 13 865 *g* at 4 °C for 30 min. The cytosolic supernatant was incubated and stirred with 1 mL Ni-NTA agarose (Qiagen, Hilden, Germany) at 4 °C for 1 h. After pouring the slurry in an empty column, it was washed twice with 8 mL buffer A containing 20 mM imidazole. The SUMO-His$_6$-sP26 protein was eluted with 5x 1 mL buffer A containing 100, 250, and 500 mM imidazole. Elution fractions containing the SUMO-His$_6$-sP26 protein (250 mM imidazole) were combined and dialyzed overnight against buffer A.

SUMO-sP26 was purified from 1 L culture of *E. coli* BL21 (DE3) carrying the plasmid pRS1228 as described above. The SUMO domain from the fusion protein was removed by SUMO protease (Thermo Fisher Scientific™, Waltham, USA, 10 U per 20 μg fusion protein) using buffer B (50 mM HEPES, 300 mM NaCl, pH 7.5) and the remaining tag-less sP26 separated from the SUMO-tag as well as the protease by Co-NTA chromatography (Thermo Fisher Scientific™, Waltham, USA).

For co-elution experiments, 1 mg of purified SUMO-His$_6$-sP26 protein was incubated with 500 μL *M. mazei* crude cell extract grown under N limitation (27.5 mg) and immobilized to 0.5 mL Ni-NTA agarose for 1 h at 4 °C. All nonbinding proteins were washed off with 5 × 1 mL buffer A and the SUMO-His$_6$-sP26 and potentially interacting proteins eluted with 4 × 1 mL buffer A supplemented with 250 mM imidazole. The elution fractions were analyzed with Coomassie- and silver-stained SDS/PAGE and Tris/Tricine/PAGE (see Gel electrophoretic separation). The respective *M. mazei* crude cell extract was generated as follows: 1 L *M. mazei* culture grown exponentially under N limitation was harvested at 4000 *g* and 4 °C for 30 min and the cell pellet resuspended in 500 μL 50 mM Tris/HCl pH 6.8 supplemented with DNaseI (Thermo Fisher Scientific™, Waltham). Cells were twice disrupted by GinoGrinder (SPEX CertiPrep, Metuchen, NJ, USA) on ice at 1300 strokes for 3 min each, followed by centrifugation at 13 865 *g* and 4 °C for 30 min to separate cell debris.

Identification of co-eluting proteins by LC-MSMS analysis

Coomassie blue (R-250)-stained gel bands were excised and destained. The individual bands were reduced with dithiothreitol (10 mM, 56 °C, 1 h) and subsequently alkylated in the presence of iodoacetamide (50 mM, room temperature, 30 min). Enzymatic digestion of was performed overnight at 37 °C by the addition of sequencing grade trypsin (Promega, Madison, WI, USA) (100 ng per sample) in 100 μL of ammonium bicarbonate (ABC) buffer (100 mM pH 7.4). The peptides were extracted via one wash with 60% acetonitrile with 0.1% trifluoroacetic acid and a second wash with 100% acetonitrile. The peptides were then dried via vacuum evaporation prior to LC-MS analysis. For LC-MS analysis, the samples were suspended in UHPLC loading buffer (3% acetonitrile + 0.1% trifluoroacetic acid).

Chromatographic separation was performed on a Dionex U3000 UHPLC system (Thermo Fisher Scientific, Darmstadt, Germany) equipped with an Acclaim PepMap 100 column (3 μm particle size, 75 μm × 150 mm) and μ-precolumn (300 μm × 5 mm) coupled online to LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). The eluents used were as follows: eluent A: 0.05% formic acid (FA) and eluent B: 80% ACN + 0.05% FA. The separation was performed over a programmed 90-min run. Initial chromatographic conditions were 5% B for 5 min followed by an increase to 10% B over 1 min, subsequently a linear gradient from 10% to 40% B over 54 min, a 5-min increase to 95% B, and 10 min at 95% B. Following this, an inter-run equilibration of the column was achieved by 15 min at 5% B. A constant flow rate of 300 nL·min^{-1} was used, and 8 μL of sample was injected per run. Data acquisition on the LTQ Orbitrap Velos mass spectrometer utilized CID activation (NCE 35). A full scan MS acquisition was performed (resolution 60 000) scan range 300–1500 *m/z*, maximum IT 100 ms. Subsequent data-dependent MS/MS (resolution 7500, minimum intensity 500), of the top 10 most intense ions, single charged, and undetermined charged state ions were excluded, and dynamic exclusion was enabled (90-s duration, repeat count 2 in 30 s); internal lockmass was enabled on 445.12003 *m/z*.

MS data files were searched against a FASTA database containing the full *M. mazei* proteome (accessed from Uni-Prot 2016.03.16) plus predicted sORF encoded proteins [27,33], and the cRAP list of commonly occurring laboratory contaminants (version 1.0, 2012.01.01). The searches were performed using the Proteome Discoverer software package (version 1.4.0.288) using the SequestHT search algorithm. A tryptic search was performed (precursor tolerance 10 p.p.m., fragment tolerance 0.04 Da, missed cleavages 2); Variable protein modification: oxidation of methionine and fixed modification: cysteine carbamidomethylation. Strict parsimony criteria were applied with a target FDR of 0.01 (1%) applied at peptide level. In addition, proteins required two high-confident peptides to be considered as identified.

Reverse pull-down analysis of immobilized Strep-GlnA$_1$ and SUMO-His$_6$-sP26

E. coli BL21 pRIL containing plasmid pRS375 was grown at 37 °C to a turbidity of 0.6 at 600 nm when the expression of Strep-GlnA$_1$ was induced by adding 200 μg·L^{-1} tetracycline. After 2-h incubation at 37 °C, cells were

harvested, resuspended in 6 mL W-buffer (100 mM Tris, pH 8.0; 150 mM NaCl; 1 mM EDTA), and disrupted using a French pressure cell at 4135 x 10^6 N·m^{-2}, followed by centrifugation at 20 000 g for 20 min. The cell-free crude extract was incubated with 1 mL W-buffer equilibrated Strep-Tactin® sepharose® matrix (IBA, Göttingen, Germany) incubated for 30 min at 4 °C under slightly swivel. The column was washed with 10 mL W-buffer. Strep-GlnA$_1$ was eluted in the presence of 2.5 mM D-Desthiobiotin (IBA) followed by buffer exchange using an Amicon Centrifugal Filter (30 KDa) (Merck) according to the manufacturer's instructions. To prepare cell-free crude extract containing SUMO-His$_6$-sP26, *E. coli* BL21 pRIL / pRS1229 was grown in 1 L LB at 37 °C to a turbidity at 600 nm of 0.6 and the expression of SUMO-His$_6$-sP26 was induced by adding 100 μM IPTG, followed by further incubation for 2 h. Cells were harvested, and cell-free extract was prepared as described above. Approximately 50 mg of crude extract containing SUMO-His$_6$-sP26 was mixed with 1 mg purified Strep-GlnA$_1$ and incubated with 1 mL Strep-Tactin® sepharose® matrix (IBA) for 30 min at 4 °C under slightly swivel. Nonbinding proteins were subsequently washed from the column with 10 mL W-buffer. Strep-GlnA$_1$ and potential interacting proteins were eluted in the presence of 2.5 mM D-Desthiobiotin (IBA). Aliquots of wash and elution fractions of the reverse cochromatography were separated by SDS/PAGE according to Laemmli (1970), and the elution fractions were further analyzed by western blot analysis as described in Weidenbach *et al.* [30] using commercial antibodies directed against the His-tag following the instructions of the manufacture (Qiagen).

Purification of His$_6$-GlnK$_1$, His$_6$-GlnA$_1$, and His$_6$-sP26

One litre of the respective *E. coli* BL21 / pRIL strains carrying pRS203 (His$_6$-GlnK$_1$), pRS196 (His$_6$-GlnK$_1$), or pRS1245 (His$_6$-sP26) was grown in LB medium at 37 °C under rigorous shaking to a turbidity of 0.7 at 600 nm. The cultures were cooled down to 18 °C followed by supplementing with 10 μM IPTG and further incubation at 18 °C for ~ 18 h. The cells were harvested at 4000 g and 4 °C for 20 min and resuspended in 20 mL buffer B (50 mM HEPES, 300 mM NaCl, pH 7.5). Cytosolic extracts were generated as described above and the His$_6$-tagged proteins purified by metal affinity chromatography and gravity flow using 0.5 mL HisPur™ cobalt superflow agarose (Thermo Fisher Scientific™, Waltham) according to manufacturer's protocol, eluting proteins at 50 mM imidazole. The imidazole of the elution fraction was removed until the concentration of imidazole was below 1 mM using Amicon® centrifugal filters (0.5 mL size, GlnA$_1$: 30 kDa and GlnK$_1$: 10 kDa) (Merck KGaA). His$_6$-sP26 was purified as described above with the following additional steps: The generated cytosolic extract was fractionated by $(NH_4)_2SO_4$

precipitation prior the metal affinity chromatography. His$_6$-sP26 precipitated at ~ 40% to 45% ammonium sulfate. The cytosolic extract was supplemented with 30% $(NH_4)_2SO_4$ (v/v) and incubated under stirring for 1 h at 8 °C. After centrifugation at 13.865 x g and 4 °C for 45 min, the respective supernatant was supplemented with additional 15% ammonium sulfate and processed as described. The generated second pellet (precipitated His$_6$-sP26) was resuspended in 20 mL Buffer B and dialyzed overnight against buffer B. The protein was purified using HisPur™ cobalt superflow agarose (0.5 mL) (Thermo Fisher Scientific™, Waltham). Finally, the His$_6$-sP26 protein was separated from contaminating larger proteins using size-exclusion 10 kDa Amicon® centrifugal filters (0.5 mL size) and the buffer was exchanged to buffer B without imidazole using 3 kDa Amicon® centrifugal filters (0.5 mL size) (both from Merck KGaA).

Gel electrophoretic separation and determination of protein concentrations

The purifications of His$_6$-GlnA$_1$ and His$_6$-GlnK$_1$ were analyzed using standard 12% SDS/PAGE [56], whereas His$_6$-sP26 was analyzed using 18% Tricine/SDS/PAGE [57]. For interaction analyses, the gradient 4–12% Bis–Tris NuPAGE (Thermo Fisher Scientific™, Waltham) according to manufacturer's protocol was used.

Due to the lack of aromatic amino acids within the small protein His$_6$-sP26, the Pierce™ BCA protein assay kit (Thermo Fisher Scientific™, Waltham) was used to determine the concentration of all proteins and was performed according to manufacturer's protocol. Absorbance was measured at 562 nm by the spectrophotometer Ultrospec 2100 UV/VIS (GE Healthcare™, Chicago, IL, USA). The concentration was calculated by using a BSA standard (50–2000 μg·mL^{-1}).

Interaction and affinity analysis by microscale thermophoresis (MST) using Monolith NT.115

Fifty microgram synthesized sP26 (Davids Biotechnologie, Regensburg, Germany) was fluorescently labeled (depicted as *) with the Monolith NT Protein Labeling Kit RED (NanoTemper) according to manufacturer's protocol. Twenty nanomolar labeled sP26 in MST buffer (50 mM Tris/HCl pH 7.4, 150 mM NaCl, 10 mM MgCl$_2$, 0.05% Tween-20) was applied to a dilution series of purified His$_6$-GlnA$_1$ ranging from nM to μM concentrations and measured using standard treated capillaries (NanoTemper), 100% Excitation Power, medium MST-Power, and MST-on of 5 s.

Purified His$_6$-GlnA$_1$ or His$_6$-GlnK$_1$ was fluorescently labeled (depicted as *) using the Monolith Protein Labeling Kit RED-NHS 2nd Generation (NanoTemper) with buffer B (50 mM HEPES, 300 mM NaCl, pH 7.5). A degree of

labeling of 0.95 or 0.87 was achieved, respectively, and for analysis, 20 nM labeled protein, supplemented with 0.05% Tween 20 and 1 mg·mL^{-1} BSA, was applied to nM to μM concentrations of nonlabeled putative interaction partners. The measurements were performed using standard treated capillaries, an excitation power of 20%, high MST-power, and MST-on of 15 s.

In general, the proteins were incubated at RT for 5 min prior to loading the capillaries. K_D values were calculated using the Nanotemper tool MO. Affinity Analysis. Data of three independent pipetted measurements were analyzed (MO. Affinity Analysis software version 2.3, NanoTemper Technology).

Determination of glutamine synthetase activity

Glutamine synthetase (GS) activity was determined by using the coupled optical test assay described by Shapiro [36], which couples the consumption of ATP by the conversion of ammonium and glutamate to glutamine catalyzed by glutamine synthetase to the oxidation of NADH by lactate dehydrogenase. The test assay was optimized by testing several buffer systems at pH 7 instead of the reported MOPS buffer (TES, $HPO_4^{2-}/H_2PO_4^-$, imidazole, Tris/HCl, and HEPES). The optimized assay buffer used contained 90 mM KCl, 50 mM NH₄Cl, 30 mM glutamate (pH 7), 50 mM MgCl₂, and 43 mM HEPES (pH 7) (final concentrations). The test assay was performed as follows: 0.95 μM purified His₆-GlnA₁ was pre-incubated for 5 min at RT in 0.4 mL freshly prepared concentrated assay buffer, supplemented with 20 μL 14 mM NADH (freshly prepared) (Carl Roth®), 10 μL 100 mM PEP (Carl Roth®), 10 μL of the enzymatic mixture (lactic dehydrogenase, 900–1400 U, and pyruvate kinase, 600–1000 U; Sigma-Aldrich, St. Louis, MO, USA) and varying amounts of purified His₆-sP26 or/and His₆-GlnK₁ to a final volume of 0.95 mL. Once the absorbance at 340 nm was constant, the reaction was started by adding 50 μL 72 mM ATP (Roche, Basel, Switzerland). The reaction was continuously monitored over a time course of 300 s by a spectrophotometer (Jasco V-550, Pfungstadt, Germany). In case the assays were performed in the presence of 2-OG, it was supplemented 5 min before the start. The slope was calculated for the linear degression of the absorbance at 340 nm, and the Δc was calculated using the Beer-Lambert law.

Conformation analysis using Tycho NT.6

His₆-GlnA₁ was analyzed using Tycho NT.6 (NanoTemper) by applying a standard capillary (10 μL) with 1.5 mg·mL^{-1} enzyme in buffer B. Thermal unfolding profiles were recorded within a temperature gradient between 35 and 95 °C. In case the assays were performed with the addition

of 2-OG (2.5, 5, 7.5, 10 and 12.5 mM), it was supplemented, respectively, 5 min before start.

Complex analysis using size-exclusion chromatography

All proteins were heterologous expressed in *E. coli* and purified as described above: 168.75 μg His₆-GlnA₁, 77.5 μg His₆-GlnK₁, and 24 μg His₆-sP26 in different combinations (His₆-GlnA₁; His₆-GlnA₁ + His₆-sP26; His₆-GlnA₁ + His₆-GlnK₁; His₆-GlnA₁ + His₆-GlnK₁ His₆-sP26) were incubated together at RT for 5 min in the presence and absence of 5 mM 2-OG prior to loading on the analytical ENrich 650 column (Bio-Rad Laboratories, Inc., Hercules, USA). The column was equilibrated with 50 mM Tris/HCl and 150 mM NaCl and supplemented with 5 mM 2-OG when the complexes were pre-incubated in the presence of 5 mM 2-OG and the applied protein isocratic eluted with the respective buffer with a flow rate of 1 mL·min^{-1}. One millilitre fractions were collected and concentrated by TCA precipitation (15% v/v). The pellets were resuspended in 50 mM ammonium acetate for LC-MS/MS analysis.

Proteomic analysis of SEC fractions

SEC fractions were precipitated before being suspended in 100 mM ammonium acetate. Aliquots (10 μL) from each fraction were made up to a volume of 60 μL with 100 mM ABC buffer (pH 7.4). The proteins were reduced with dithiothreitol (10 mM, 56 °C, 1 h) and subsequently alkylated in the presence of chloroacetamide (50 mM, room temperature, 30 min). Enzymatic digestion was performed overnight at 37 °C by the addition of sequencing grade trypsin (Promega) (100 ng per sample) in 100 μL of ABC buffer. The samples were acidified via the addition of 10 μL of 10% TFA to stop the digestion before being dried down under vacuum (Concentrator plus; Eppendorf, Hamburg, Germany). The dried peptides were stored at −20 °C prior to LC-MS analysis. On the day of LC-MS analysis, the samples were suspended in 15 μL of UHPLC loading buffer (3% acetonitrile + 0.1% trifluoroacetic acid).

Chromatographic separation was performed on a Dionex U3000 UHPLC system (Thermo Fisher Scientific, Darmstadt) equipped with an Acclaim PepMap 100 column (2 μm particle size, 75 μm × 500 mm) and μ-precolumn (300 μm × 5 mm) coupled online to Fusion Lumos mass spectrometer (Thermo Fisher scientific). The eluents used were as follows: eluent A: 0.05% formic acid (FA) and eluent B: 80% ACN + 0.05% FA. The separation was performed over a programmed 60-min run. Initial chromatographic conditions were 4% B for 3 min followed by linear gradients from 4% to 50% B over 30 min, a 1-min increase to 90% B, and 10 min at 90% B. Following this, an inter-run equilibration of the column was achieved

by 15 min at 4% B. A constant flow rate of 300 nL·min⁻¹ was used, and 1 μL of sample was injected per run. Data acquisition on the Fusion Lumos mass spectrometer utilized HCD activation (NCE 30). A full scan MS acquisition was performed (resolution 120 000) scan range 300–1500 m/z, AGC target 4e5, maximum IT 50 ms. In a subsequent data-dependent MS/MS (resolution 30 000, minimum intensity 5e4, maximum IT 100 ms) of the most intense ions for the following 3 s; the MIPS mode was activated for peptides, single charged and > 7+ charged peptides were excluded and dynamic exclusion was activated (60-s duration). The internal barrier mass was activated at 445.12003 m/z.

MS data files were searched against a combined set of fasta databases containing the full *M. mazei* proteome plus all predicted sORF encoded proteins (PTK_92x_Mmazei_Full_Plus_SEP_190425.fasta), and the His-tag protein sequences provided for sp26 and GlnA₁ (up26HisTagged + GlnAHis.fasta). In addition, protein sequences for the expression system, *E. coli* D3, and the cRAP list of common laboratory contaminants were included for all searches (PTK_proteome_Ecoli.fasta and cRAP47.fasta). The searches were performed using the Proteome Discoverer software package (version 2.2.0.388) using the SequestHT search algorithm. A semi-tryptic search was performed (Precursor tolerance 10 p.p.m., fragment tolerance 0.02 Da, missed cleavages 2) with variable modifications; methionine oxidation and protein N-terminal acetylation, fixed modification; cysteine carbamidomethylation.

Strict parsimony criteria have been applied with a target FDR of 0.01 (1%) applied at the PSM, peptide, and protein level. In addition, proteins required two high-confident peptides to be considered identified, or in the case of sP26, single peptides were manually validated to assure quality of the identification.

Protein production for NMR analyses

Plasmid pRS1228 carrying the SUMO-tagged sP26 sequence was transformed into *E. coli* BL21 cells and heterologously expressed. The necessary ¹⁵N-labeling scheme for NMR spectroscopy was performed by using M9 minimal medium containing ¹⁵N-enriched NH₄Cl (Cambridge Isotope Laboratories, Cambridge, MA, USA). The transformed cells were incubated at 37 °C until a turbidity at 600 nm of 0.6 was reached and protein expression was subsequently induced by addition of 1 mM IPTG. The cell cultures were further incubated at 18 °C overnight. The cells were harvested by centrifugation (5000 r.p.m., 15 min, 4 °C) and resuspended in purification buffer (50 mM HEPES pH 7, 300 mM NaCl) supplemented with one EDTA-free protease inhibitor tablet (cOmplete™, Roche, Germany).

Protein purification was accomplished by tandem Co-NTA chromatography. Cleavage of the SUMO-tag was performed with SUMO protease at 20 °C overnight. Purity of the protein was confirmed by SDS/PAGE and mass spectrometry (MALDI).

NMR spectroscopic analysis

NMR samples were prepared in NMR buffer containing 50 mM HEPES pH 7, 300 mM NaCl, 1 mM DSS, and 10% D₂O. Spectra were recorded on a 600 MHz Bruker spectrometer at 298 K. Referencing was done by setting the DSS signal to 0.00 p.p.m.; referencing of the indirect dimension was calculated according to the ¹H referencing [58]. Spectra were processed and analyzed using TopSpin 3.5pl7 (Bruker Biospin, Ettlingen, Germany).

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

Author contributions

RAS conceptualized and designed the study; RAS, AT, HS, LC, and KW contributed to methodology; MG, BJ, KW, MG, CK, LC, AH, NK, and DP investigated the study; RAS, KW, HS, and MG visualized the study; RAS and KW wrote—original draft preparation; RAS supervised the study; RAS, HS, and AT involved in funding acquisition.

All authors have read and agreed to the published version of the manuscript.

Peer Review

The peer review history for this article is available at https://publons.com/publon/10.1111/febs.15799.

References

1 Hobbs EC, Yin X, Paul BJ, Astarita JL & Storz G (2012) Conserved small protein associates with the multidrug efflux pump AcrB and differentially affects

antibiotic resistance. *Proc Natl Acad Sci USA* **109**, 16696–16701.

2 Orr MW, Mao Y, Storz G & Qian SB (2020) Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Res* **48**, 1029–1042.

3 Ramamurthi KS & Storz G (2014) The small protein floodgates are opening; now the functional analysis begins. *BMC Biol* **12**, 96.

4 Basrai MA, Hieter P & Boeke JD (1997) Small open reading frames: beautiful needles in the haystack. *Genome Res* **7**, 768–771.

5 Miravet-Verde S, Ferrar T, Espadas-Garcia G, Mazzolini R, Gharrab A, Sabido E, Serrano L & Lluch-Senar M (2019) Unraveling the hidden universe of small proteins in bacterial genomes. *Mol Syst Biol* **15**, e8290.

6 Cardon T, Herve F, Delcourt V, Roucou X, Salzet M, Franck J & Fournier I (2020) Optimized sample preparation workflow for improved identification of ghost proteins. *Anal Chem* **92**, 1122–1129.

7 Cassidy L, Kaulich PT & Tholey A (2019) Depletion of high-molecular-mass proteins for the identification of small proteins and short open reading frame encoded peptides in cellular proteomes. *J Proteome Res* **18**, 1725–1734.

8 Cassidy L, Prasse D, Linke D, Schmitz RA & Tholey A (2016) Combination of bottom-up 2D-LC-MS and semi-top-down GelFree-LC-MS enhances coverage of proteome and low molecular weight short open reading frame encoded peptides of the Archaeon *Methanosarcina mazei*. *J Proteome Res.* **15**, 3773–3783.

9 Chu Q, Ma J & Saghatelian A (2015) Identification and characterization of sORF-encoded polypeptides. *Crit Rev Biochem Mol Biol* **50**, 134–141.

10 Ingolia NT (2014) Ribosome profiling: new views of translation, from single codons to genome scale. *Nat Rev Genet* **15**, 205–213.

11 Lago M, Monteil V, Douche T, Guglielmini J, Criscuolo A, Maufrais C, Matondo M & Norel F (2017) Proteome remodelling by the stress sigma factor RpoS/sigma(S) in Salmonella: identification of small proteins and evidence for post-transcriptional regulation. *Sci Rep* **7**, 2127.

12 Ma J, Ward CC, Jungreis I, Slavoff SA, Schwaid AG, Neveu J, Budnik BA, Kellis M & Saghatelian A (2014) Discovery of human sORF-encoded polypeptides (SEPs) in cell lines and tissue. *J Proteome Res* **13**, 1757–1765.

13 Meydan S, Marks J, Klepacki D, Sharma V, Baranov PV, Firth AE, Margus T, Kefi A, Vazquez-Laslop N & Mankin AS (2019) Retapamulin-assisted ribosome profiling reveals the alternative bacterial proteome. *Mol Cell* **74**, 481–493.e6

14 Saghatelian A & Couso JP (2015) Discovery and characterization of smORF-encoded bioactive polypeptides. *Nat Chem Biol* **11**, 909–916.

15 Weaver J, Mohammad F, Buskirk AR & Storz G (2019) Identifying small proteins by ribosome profiling with stalled initiation complexes. *MBio* **10**, e02819-18.

16 Duval M & Cossart P (2017) Small bacterial and phagic proteins: an updated view on a rapidly moving field. *Curr Opin Microbiol* **39**, 81–88.

17 Storz G, Wolf YI & Ramamurthi KS (2014) Small proteins can no longer be ignored. *Annu Rev Biochem* **83**, 753–777.

18 Jevtic Z, Stoll B, Pfeiffer F, Sharma K, Urlaub H, Marchfelder A & Lenz C (2019) The Response of *Haloferax volcanii* to salt and temperature stress: a proteome study by label-free mass spectrometry. *Proteomics* **19**, e1800491.

19 Kaulich PT, Cassidy L, Weidenbach K, Schmitz RA & Tholey A (2020) Complementarity of different SDS-PAGE gel staining methods for the identification of short open reading frame-encoded peptides. *Proteomics* **20**, 2000084.

20 Kubatova N, Jonker HRA, Saxena K, Richter C, Vogel V, Schreiber S, Marchfelder A & Schwalbe H (2020) Solution structure and dynamics of the small protein HVO_2922 from *Haloferax volcanii*. *ChemBioChem* **21**, 149–156.

21 Kubatova N, Pyper DJ, Jonker HRA, Saxena K, Remmel L, Richter C, Brantl S, Evguenieva-Hackenberg E, Hess WR, Klug G *et al.* (2019) Rapid biophysical characterization and nmr spectroscopy structural analysis of small proteins from bacteria and archaea. *ChemBioChem* **21**, 1178–1187.

22 Nagel C, Machulla A, Zahn S & Soppa J (2019) Several one-domain zinc finger micro-proteins of *Haloferax Volcanii* are important for stress adaptation, biofilm formation, and swarming. *Genes* **10**, 361.

23 Prasse D, Thomsen J, De Santis R, Muntel J, Becher D & Schmitz RA (2015) First description of small proteins encoded by spRNAs in *Methanosarcina mazei* strain Go1. *Biochimie* **117**, 138–148.

24 Zahn S, Kubatova N, Pyper DJ, Cassidy L, Saxena K, Tholey A, Schwalbe H & Soppa J (2020) Biological functions, genetic and biochemical characterization, and NMR structure determination of the small zinc finger protein HVO_2753 from *Haloferax volcanii*. *FEBS J.* https://doi.org/10.1111/febs.15559

25 Ferry JG (1999) Enzymology of one-carbon metabolism in methanogenic pathways. *FEMS Microbiol Rev* **23**, 13–38.

26 Buddeweg A, Sharma K, Urlaub H & Schmitz RA (2018) sRNA41 affects ribosome binding sites within polycistronic mRNAs in *Methanosarcina mazei* Go1. *Mol Microbiol* **107**, 595–609.

27 Jager D, Sharma CM, Thomsen J, Ehlers C, Vogel J & Schmitz RA (2009) Deep sequencing analysis of the *Methanosarcina mazei* Go1 transcriptome in response to nitrogen availability. *Proc Natl Acad Sci USA* **106**, 21878–21882.

28 Prasse D, Forstner KU, Jager D, Backofen R & Schmitz RA (2017) sRNA154 a newly identified regulator of nitrogen fixation in *Methanosarcina mazei* strain Go1. *RNA Biol* **14**, 1544–1558.

29 Weidenbach K, Ehlers C, Kock J, Ehrenreich A & Schmitz RA (2008) Insights into the NrpR regulon in *Methanosarcina mazei* Go1. *Arch Microbiol* **190**, 319–332.

30 Weidenbach K, Ehlers C, Kock J & Schmitz RA (2010) NrpRII mediates contacts between NrpRI and general transcription factors in the archaeon *Methanosarcina mazei* Go1. *FEBS J* **277**, 4398–4411.

31 Weidenbach K, Ehlers C & Schmitz RA (2014) The transcriptional activator NrpA is crucial for inducing nitrogen fixation in *Methanosarcina mazei* Go1 under nitrogen-limited conditions. *FEBS J* **281**, 3507–3522.

32 Ehlers C, Weidenbach K, Veit K, Forchhammer K & Schmitz RA (2005) Unique mechanistic features of post-translational regulation of glutamine synthetase activity in *Methanosarcina mazei* strain Go1 in response to nitrogen availability. *Mol Microbiol* **55**, 1841–1854.

33 Dar D, Prasse D, Schmitz RA & Sorek R (2016) Widespread formation of alternative 3' UTR isoforms via transcription termination in archaea. *Nat Microbiol* **1**, 16143.

34 Leigh JA & Dodsworth JA (2007) Nitrogen regulation in bacteria and archaea. *Annu Rev Microbiol* **61**, 349–377.

35 Brown JR, Masuchi Y, Robb FT & Doolittle WF (1994) Evolutionary relationships of bacterial and archaeal glutamine synthetase genes. *J Mol Evol* **38**, 566–576.

36 Shapiro BM & Stadtman ER (1970) [130] Glutamine synthetase (*Escherichia coli*). In Methods in Enzymology, pp. 910–922. New York, NY, Academic Press.

37 Reitzer L (2003) Nitrogen assimilation and global regulation in *Escherichia coli*. *Annu Rev Microbiol* **57**, 155–176.

38 Reitzer L & Schneider BL (2001) Metabolic context and possible physiological themes of sigma(54)-dependent genes in *Escherichia coli*. *Microbiol Mol Biol Rev* **65**, 422–444, table of contents.

39 Cohen-Kupiec R, Marx CJ & Leigh JA (1999) Function and regulation of glnA in the methanogenic archaeon *Methanococcus maripaludis*. *J Bacteriol* **181**, 256–261.

40 Smith DR, Doucette-Stamm LA, Deloughery C, Lee H, Dubois J, Aldredge T, Bashirzadeh R, Blakely D, Cook R, Gilbert K *et al.* (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J Bacteriol* **179**, 7135–7155.

41 Deppenmeier U, Johann A, Hartsch T, Merkl R, Schmitz RA, Martinez-Arias R, Henne A, Wiezer A, Baumer S, Jacobi C *et al.* (2002) The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. *J Mol Microbiol Biotechnol* **4**, 453–461.

42 Forchhammer K (2004) Global carbon/nitrogen control by PII signal transduction in cyanobacteria: from signals to targets. *FEMS Microbiol Rev* **28**, 319–333.

43 Irmler A, Sanner S, Dierks H & Forchhammer K (1997) Dephosphorylation of the phosphoprotein P(II) in Synechococcus PCC 7942: identification of an ATP and 2-oxoglutarate-regulated phosphatase activity. *Mol Microbiol* **26**, 81–90.

44 Dodsworth JA & Leigh JA (2007) NifI inhibits nitrogenase by competing with Fe protein for binding to the MoFe protein. *Biochem Biophys Res Commun* **364**, 378–382.

45 Dodsworth JA & Leigh JA (2006) Regulation of nitrogenase by 2-oxoglutarate-reversible, direct binding of a PII-like nitrogen sensor protein to dinitrogenase. *Proc Natl Acad Sci USA* **103**, 9779–9784.

46 Garcia-Dominguez M, Reyes JC & Florencio FJ (1999) Glutamine synthetase inactivation by protein-protein interaction. *Proc Natl Acad Sci USA* **96**, 7161–7166.

47 Garcia-Dominguez M, Reyes JC & Florencio FJ (2000) NtcA represses transcription of gifA and gifB, genes that encode inhibitors of glutamine synthetase type I from Synechocystis sp. PCC 6803. *Mol Microbiol* **35**, 1192–1201.

48 Klahn S, Schaal C, Georg J, Baumgartner D, Knippen G, Hagemann M, Muro-Pastor AM & Hess WR (2015) The sRNA NsiR4 is involved in nitrogen assimilation control in cyanobacteria by targeting glutamine synthetase inactivating factor IF7. *Proc Natl Acad Sci USA* **112**, E6243–E6252.

49 Klahn S, Bolay P, Wright PR, Atilho RM, Brewer KI, Hagemann M, Breaker RR & Hess WR (2018) A glutamine riboswitch is a key element for the regulation of glutamine synthetase in cyanobacteria. *Nucleic Acids Res* **46**, 10082–10094.

50 Gernhardt P, Possot O, Foglino M, Sibold L & Klein A (1990) Construction of an integration vector for use in the archaebacterium *Methanococcus voltae* and expression of a eubacterial resistance gene. *Mol Gen Genet* **221**, 273–279.

51 Metcalf WW, Zhang JK, Apolinario E, Sowers KR & Wolfe RS (1997) A genetic system for Archaea of the genus Methanosarcina: liposome-mediated transformation and construction of shuttle vectors. *Proc Natl Acad Sci USA* **94**, 2626–2631.

52 Inoue H, Nojima H & Okayama H (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23–28.

53 Ehlers C, Weidenbach K, Veit K, Deppenmeier U, Metcalf WW & Schmitz RA (2005) Development of genetic methods and construction of a chromosomal glnK1 mutant in *Methanosarcina mazei* strain Go1. *Mol Genet Genomics* **273**, 290–298.

54 Veit K, Ehlers C, Ehrenreich A, Salmon K, Hovey R, Gunsalus RP, Deppenmeier U & Schmitz RA (2006) Global transcriptional analysis of *Methanosarcina mazei* strain Go1 under different nitrogen availabilities. *Mol Genet Genomics* **276**, 41–55.

55 Nickel L, Weidenbach K, Jager D, Backofen R, Lange SJ, Heidrich N & Schmitz RA (2013) Two CRISPR-Cas systems in *Methanosarcina mazei* strain Go1 display common processing features despite belonging to different types I and III. *RNA Biol* **10**, 779–791.

56 Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

57 Schagger H (2006) Tricine-SDS-PAGE. *Nat Protoc* **1**, 16–22.

58 Wishart DS (2011) Interpreting protein chemical shift data. *Prog Nucl Magn Reson Spectrosc* **58**, 62–87.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig, S1. Annotated spectra for the proteotypic peptide.

Fig. S2. Schematic plasmid maps of expression vectors.

Table S1. $GlnA_1$ activity determined as described in Methods using purified proteins of independent purifications.

Table S2. Strains and plasmids used.