

# Biological functions, genetic and biochemical characterization, and NMR structure determination of the small zinc finger protein HVO\_2753 from *Haloferax volcanii*

Sebastian Zahn<sup>1</sup>, Nina Kubatova<sup>2</sup>, Dennis J. Pyper<sup>2</sup>, Liam Cassidy<sup>3</sup>, Krishna Saxena<sup>2</sup>, Andreas Tholey<sup>3</sup>, Harald Schwalbe<sup>2</sup>  and Jörg Soppa<sup>1,4</sup> 

<sup>1</sup> Institute for Molecular Biosciences, Goethe-University, Frankfurt, Germany

<sup>2</sup> Institute for Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance, Goethe-University, Frankfurt/Main, Germany

<sup>3</sup> Systematic Proteome Research & Bioanalytics, Institute for Experimental Medicine, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

<sup>4</sup> Johann Wolfgang Goethe-Universität, Frankfurt am Main, Germany

## Keywords

$\mu$ -protein; *Haloferax volcanii*; NMR solution structure; small protein; zinc finger

## Correspondence

J. Soppa, Institute for Molecular Biosciences, Goethe-University, D-60438 Frankfurt, Germany

Tel: +49-69-798 29564

Email: soppa@bio.uni-frankfurt.de

H. Schwalbe, Institute for Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance, Goethe-University, Frankfurt/Main, Germany

Tel: +49-69-798 29737

Email: schwalbe@nmr.uni-frankfurt.de

Sebastian Zahn, Nina Kubatova and Dennis J. Pyper are joint first authors.

[Correction added on 08 January 2021, after first online publication: Johann Wolfgang Goethe-Universität, Frankfurt am Main was added for Jörg Soppa]

(Received 16 April 2020, revised 26 June 2020, accepted 2 September 2020)

doi:10.1111/febs.15559

The genome of the halophilic archaeon *Haloferax volcanii* encodes more than 40 one-domain zinc finger  $\mu$ -proteins. Only one of these, HVO\_2753, contains four C(P)XCG motifs, suggesting the presence of two zinc binding pockets (ZBPs). Homologs of HVO\_2753 are widespread in many euryarchaeota. An *in frame* deletion mutant of HVO\_2753 grew indistinguishably from the wild-type in several media, but had a severe defect in swarming and in biofilm formation. For further analyses, the protein was produced homologously as well as heterologously in *Escherichia coli*. HVO\_2753 was stable and folded in low salt, in contrast to many other haloarchaeal proteins. Only haloarchaeal HVO\_2753 homologs carry a very hydrophilic N terminus, and NMR analysis showed that this region is very flexible and not part of the core structure. Surprisingly, both NMR analysis and a fluorimetric assay revealed that HVO\_2753 binds only one zinc ion, despite the presence of two ZBPs. Notably, the analysis of cysteine to alanine mutant proteins by NMR as well by *in vivo* complementation revealed that all four C(P)XCG motifs are essential for folding and function. The NMR solution structure of the major conformation of HVO\_2753 was solved. Unexpectedly, it was revealed that ZBP1 was comprised of C(P)XCG motifs 1 and 3, and ZBP2 was comprised of C(P)XCG motifs 2 and 4. There are several indications that ZBP2 is occupied by zinc, in contrast to ZBP1. To our knowledge, this study represents the first in-depth analysis of a zinc finger  $\mu$ -protein in all three domains of life.

## Abbreviations

CSP, chemical shift perturbation; GST, glutathione S-transferase; MAE, mean square error; MS, mass spectrometry; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; RF, open reading frame; RMSE, root mean square error; ZBP, zinc binding pocket.

## Introduction

For a long time, the existence and importance of very small proteins of less than 100 amino acids were severely underestimated. The main reason is that the annotation of protein-coding genes in newly sequenced genomes typically had a lower limit of 100 codons, and, thus, smaller proteins could not be addressed in similarity searches or proteome analyses. In recent years, it has become clear that many very small proteins exist in all three domains of life, archaea, bacteria, and eukaryotes. In addition to the lack of annotation, also experimental work with very small proteins is very challenging and they are often overlooked using approaches that had been optimized for medium-sized and large proteins; for example, they are lost in standard protein isolation procedures or gel and staining systems.

Mainly two recent experimental developments have enabled the systematic identification and analysis of very small proteins: (1) Ribosomal profiling allows quantifying the positions of all ribosomes on all transcripts of a given species [1–3]. When ribosomes are found on small transcripts with very short open reading frames (sORFs), this is experimental proof for the translation of these sORFs into very small proteins. (2) Optimization of several steps of sample preparation, mass spectrometry, and bioinformatics analysis of MS spectra transformed proteomics into the new field of peptidomics [4,5].

Several recent reviews give an overview of the current knowledge about very small proteins in eukaryotes [2,6,7] and in prokaryotes [1,8–10]. Up to now a generally accepted terminology does not exist. 50 amino acids, 70 amino acids, or 100 amino acids are used as the upper limit for very small proteins. Very small proteins are denoted as 'small proteins', ' $\mu$ -proteins', 'peptides', 'micropeptides', and 'sORF-encoded peptides/proteins' (SEPs). Here, we will use the term  $\mu$ -proteins, to distinguish these very small proteins against 'small proteins', which might 'only' be smaller than an average protein of 300 amino acids, and against 'peptides', which might be processed from larger precursors.

The number of studies concentrating on  $\mu$ -proteins is much higher with eukaryotic and bacterial than with archaeal species. However, already more than 10 years ago the 'low molecular weight proteome' of the halophilic archaeon *Halobacterium salinarum* was characterized [11]. Several steps of the workflow were optimized for the application of small proteins. 380 proteins of less than 20 kDa were identified. The majority of 62% had no assigned function. Among these proteins were 20 proteins that had two C(P) XCG motifs (the P is in brackets, because it is present

in only one of the two motifs). It was predicted that the four cysteines complex a zinc ion, and, thus, that these proteins contain zinc fingers [11].

Zinc fingers are very versatile interaction domains, which can mediate interactions to DNA, RNA, proteins including membrane proteins, lipids, and other molecules [12,13]. Zinc finger proteins are involved in many cellular processes, including regulation of transcription, signal transduction, protein degradation, and many others [14]. Therefore, the identification of zinc binding domains does not allow the prediction of putative interaction partners or biological functions of the respective proteins. The zinc ion can be complexed by four cysteines (C4), by three cysteines and one histidine (C3H), or by two cysteines and two histidines (C2H2, the 'classical' zinc finger of eukaryotic transcription factors). The zinc fingers exhibit high structural variability and have been classified into eight structural groups [15].

Zinc finger proteins have first been discovered in eukaryotes and were long thought to be confined to this domain [16]. Later, it was discovered that zinc finger proteins are also present in bacteria and archaea. However, a bioinformatic analysis revealed that zinc finger motifs are only present in 1.5% of bacterial small proteins of less than 100 amino acids, but in 8% of archaeal small proteins [17]. Therefore, the importance of small zinc finger proteins seems to be much higher in archaea than in bacteria.

The genome of the halophilic archaeon *Haloferax volcanii* contains 282 genes for  $\mu$ -proteins of up to 70 aa, 43 of which contain zinc finger motifs [18]. Single *in frame* gene deletion mutants were generated for 16 of these genes. Phenotypic analyses under eight different conditions revealed differences to the wild-type for 12 of the 16 mutants, underscoring the high biological importance for zinc finger  $\mu$ -proteins for haloarchaea. A deletion mutant of a further gene for a  $\mu$ -protein without a zinc finger grew better than the wild-type under iron limitation [19]. The NMR solution structure of the protein was solved, and it turned out that HVO\_2922 is a symmetrical dimer [19]. Solution state nuclear magnetic resonance (NMR) spectroscopy is the method of choice in the structural study of small proteins. It is particularly powerful for the elucidation of the structure and dynamics of proteins in isolated form and within complexes with their biological targets [20,21]. It should be noted that it is not typical that  $\mu$ -proteins are structured. A systematic analysis of 27  $\mu$ -proteins from six archaeal and bacterial species revealed that only 5 of them were fully structured, while seven had partial structure (molten globule), 8 were fully unstructured, and 7 could not be produced

[22]. For the proteins in the molten globule state, it was predicted that they probably need an interaction partner for the formation of a fixed structure. Four of the five structured proteins were from *H. volcanii*, and, thus, it seems that haloarchaeal  $\mu$ -proteins are especially well suited for structural analyses.

Here, we present the in-depth characterization of the  $\mu$ -protein HVO\_2753 from *H. volcanii*. An *in frame* deletion mutant was generated and characterized. The protein was produced homologously in *H. volcanii* and heterologously in *E. coli* for biochemical and structural analyses. Its zinc content was quantified, and single amino acid mutants were generated to evaluate the importance of the C(P)XCG motifs. NMR analyses were used to determine the secondary structure and the temperature-dependent dynamics of the protein. Last, this report presents the 3D high-resolution solution structure, which was elucidated using conventional NOE-based restraints derived from 3D NOESY experiments.

## Results

### Characteristic features, distribution, and phylogeny of the $\mu$ -protein HVO\_2753

The genome of *H. volcanii* contains 43 genes encoding  $\mu$ -proteins with less than 70 amino acids and at least two C(P)XCG motifs, which can safely be predicted to form zinc fingers. Only one of these proteins, HVO\_2753, contains four C(P)XCG motifs, and, thus, should contain two zinc fingers. This solitary feature was the reason to choose HVO\_2753 for a detailed characterization. The sequence of the 59 amino acid protein is shown in Fig. 1. The four C(P)XCG motifs make up one third of the protein, and they are more or less evenly distributed over the sequence, leaving only short amino acid patches between the motifs and at the N terminus and C terminus. The fraction of charged and hydrophilic amino acids is very high. Table S1 shows the amino acid composition of HVO\_2753 in comparison with that of the *H. volcanii* proteome. Aliphatic amino acids are underrepresented, and the protein does not contain a single tryptophan. At least twofold overrepresented are serine,

methionine, asparagine, and, of course, cysteine. The high serine content prompted us to look for possible posttranslational modifications (see below).

BLAST searches with the HVO\_2753 sequence revealed that the protein is widely distributed in hundreds of euryarchaeota. Distant homologs were found in only eight species of bacteria, which probably have obtained the gene by lateral transfer from archaea. A HVO\_2753 homolog is not present in any eukaryotic species.

Figure S1 shows a multiple sequence alignment of HVO\_2753 and 20 selected species that represent various genera of halophilic and methanogenic archaea. Overall, the sequence is highly conserved. A few positions are conserved only in haloarchaea and differ in methanogenic archaea, for example, the glutamic acid at position 51 and the aspartic acid at position 54. The replacement of the glycine in the third C(P)XCG motif with an arginine is highly conserved in all species. Notable, all haloarchaea contain an N-terminal highly hydrophilic extension of about 10 amino acids, which is absent in methanogenic archaea. It is tempting to speculate that this hydrophilic extension is important for the solubility of the protein in the high-salt cytoplasm of haloarchaea.

Phylogenetic trees with 1000 bootstrap replications were constructed using both the maximum parsimony approach and the maximum-likelihood approach (Fig. S2A + B). Both trees are very similar to one another as well as to 16S rRNA-based species trees. This indicates that the last common ancestor of halophilic and methanogenic archaea already contained a HVO\_2753 homolog, which has been vertically inherited since then.

### Genomic localization of the HVO\_2753 gene and its expression

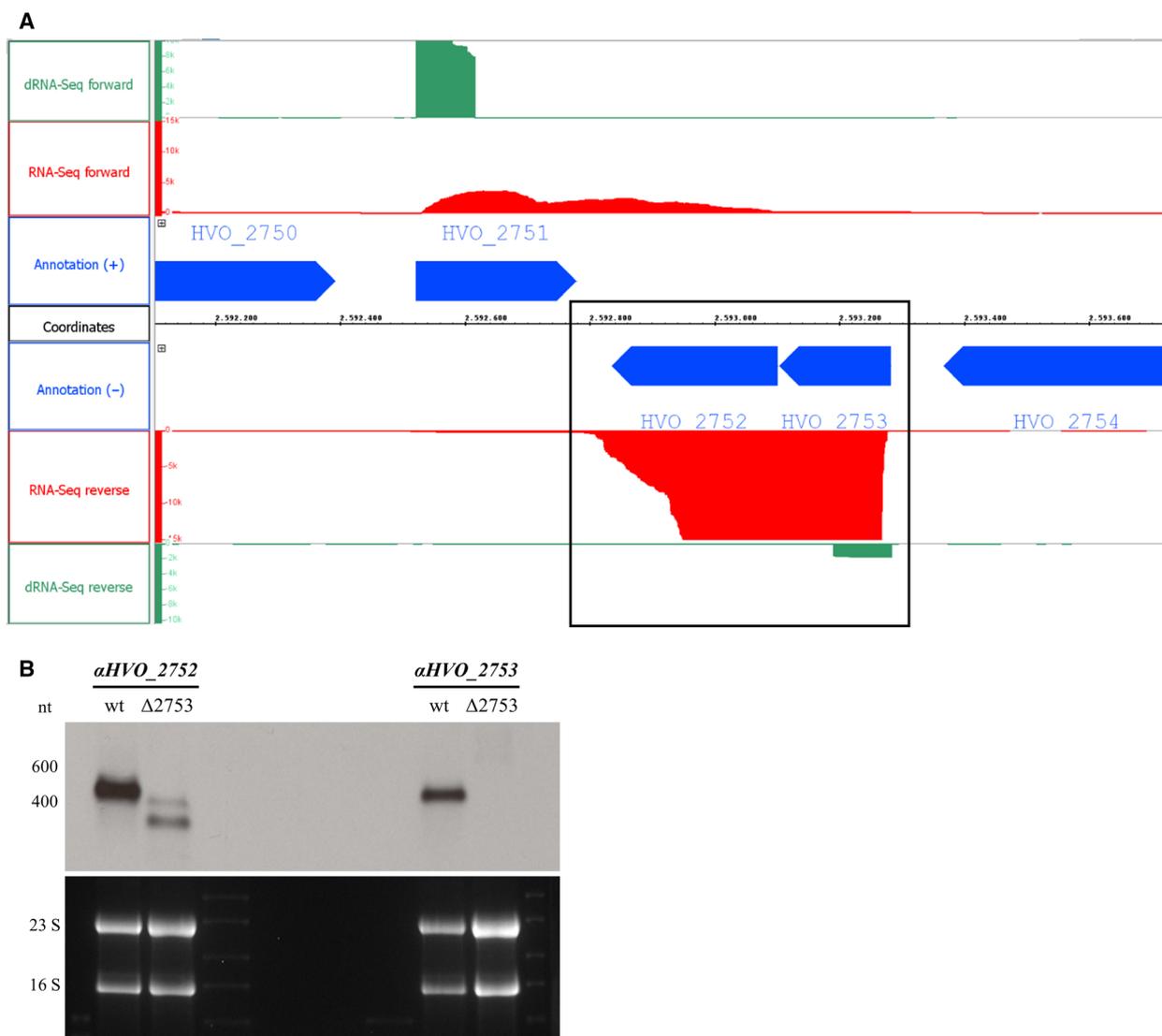
Figure 2A gives an overview of the genomic organization of the HVO\_2753 gene and its neighborhood. It is very close to the gene HVO\_2752 with an intergenic distance of only 2 nt, making co-transcription of the two genes very likely. The results from a recent RNA-Seq [23] study also indicate co-transcription of the two genes (red signal), and a dRNA-Seq study [24] found a

```

1      10      20      30      40      50      59
MSESEQRHAHQCVSCGINIAGMSAATFKCPDCGQEI SRCSKCRKQSNLYECPDCGFMGP
MSESEQRHAHQCVSCGINIAGMSAATFKCPDCGQEI SRCSKCRKQSNLYECPDCGFMGP

```

**Fig. 1.** Sequence of the protein HVO\_2753. Top: The four C(P)XCG motifs are highlighted in red. Bottom: Acidic amino acids are highlighted in red, basic residues in blue, residues with amide side chains in yellow and residues with hydroxy groups in green. The accession number in the UniProtKB database is D4GWB3\_HALVD.

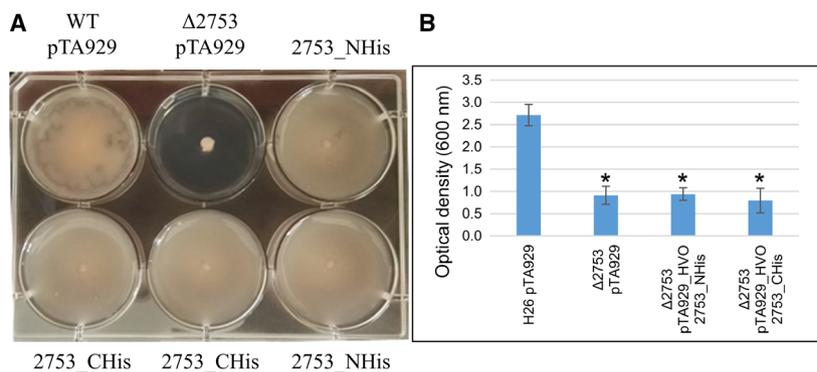


**Fig. 2.** (A) Overview of the genomic neighborhood of the *HVO\_2753* gene. A screenshot from the Integrated Genome Browser is shown, which includes the genomic numbering in the middle, the annotated protein-coding genes in blue, the results from a RNA-Seq study [23] in red, and the results from a dRNA-Seq study [24] in green. (B) Northern blot analysis of the expression of the genes *HVO\_2752* (left) and *HVO\_2753* (right). The top panel shows the hybridization signals, the bottom part a stained RNA gel showing the 16S and 23S rRNA. Both the wild-type and a *HVO\_2753 in frame* deletion mutant were analyzed, as indicated. The northern blot analyses were performed once.

single transcription start signal in front of *HVO\_2753* (green signal). For a further transcript analysis, northern blot analyses were performed. The probes for both genes gave a single signal about 500 nt in size, which fits well to common size of the two genes (448 nt) and a 3'-UTR. Together, the results provide experimental proof that *HVO\_2752* and *HVO\_2753* are transcribed into one bicistronic transcript. *HVO\_2752* encodes the beta subunit of translation elongation factor aEF1, indicating that *HVO\_2753* might also be involved in translation.

### Generation of an *in frame* deletion mutant and its phenotypic characterization

An *in frame* deletion mutant was generated to study the potential biological roles of *HVO\_2753*. *H. volcanii* is highly polyploid; therefore, Southern blot analysis and multiple cycle PCR were used to verify that the deletion mutant is homozygous. In addition, northern blot analysis was used to compare gene expression in the mutant with that of the wild-type. The *HVO\_2753* probe did not detect any transcript in the mutant,



**Fig. 3.** Phenotypic characterization of the wild-type, the HVO\_2753 deletion mutant, and a complementation strain. All strains contained either the expression plasmid or the respective empty vector pTA929. (A) Swarming assay. The swarming assay was performed in a six-well plate. 5  $\mu$ L of each strain was injected at the center of each well, and the plate was incubated for 48 h. The identity of the strains is indicated at the top and the bottom, respectively. One typical plate of three biological replicates is shown. (B) Biofilm assay. The biofilm assay was performed in 96-well plates. Strain designation is shown at the bottom. Average values of three biological and four technical replicates and their standard deviations are shown. The statistical significance was determined using an unpaired two-tailed t-test. \* - *P*-value of less than 0.05.

while the HVO\_2752 probe detected a transcript with the expected size difference to the bicistronic transcript of the wild-type (Fig. 2B).

The phenotypes of the *in frame* deletion mutant and the wild-type were compared under various conditions. Unexpectedly, the mutant did not exhibit a growth defect, neither in complex medium nor in synthetic medium with glucose as sole carbon and energy source (Fig. S3). This result excludes that HVO2753 has a critical role in translation, which would lead to a decrease in growth rate. In stark contrast, the lack of HVO\_2753 led to a complete loss of swarming (Fig. 3A). In addition, the ability to form a biofilm was severely compromised (Fig. 3B). Therefore, it turned out that HVO\_2753 is essential for surface-depended lifestyles of *H. volcanii*, whereas it is dispensable in suspension culture. The open reading frame (ORF) of the HVO\_2753 gene was introduced into an expression vector to complement the phenotypes of the deletion mutant. The swarming defect of the mutant could be fully complemented, both with an N-terminally and with a C-terminally tagged version of the protein (Fig. 3A). Unexpectedly, the biofilm defect could not be complemented with any of the two tagged protein versions (Fig. 3B), indicating that either the required protein concentration or the interaction partners are different for the two lifestyles.

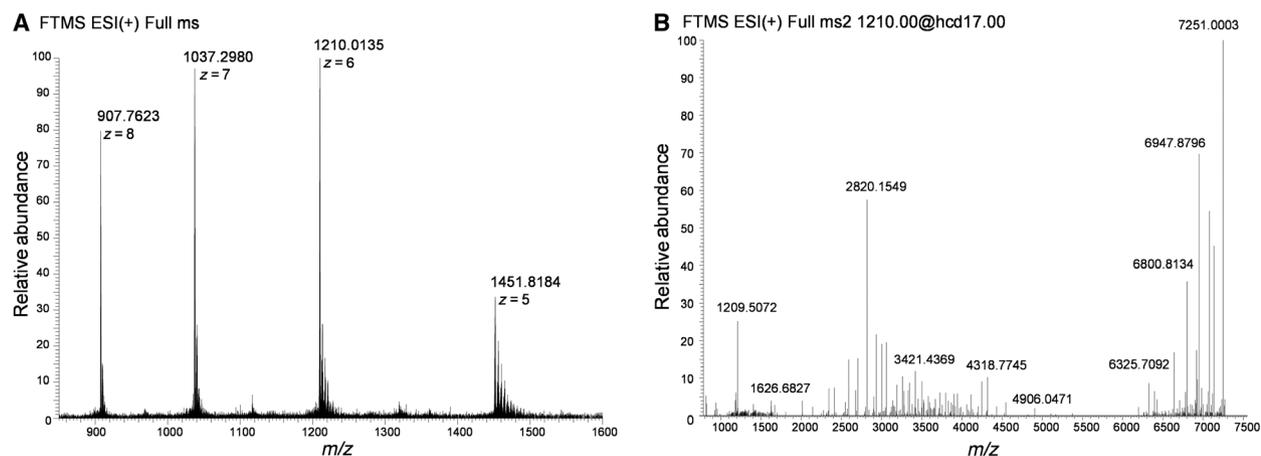
### Homologous production and purification of HVO\_2753

A his<sub>6</sub>-tagged version of HVO\_2753 was introduced into the haloarchaeal expression vector pSD1/R1-6

[25], which enables homologous overproduction of the protein. The protein was isolated under native salt conditions using affinity purification followed by size exclusion chromatography. Figure S4 shows that this two-step procedure resulted in the isolation of a very pure protein of the expected size, which was used for further analyses (see below). The yield was 3.5 mg of protein per liter culture after the two-step isolation procedure. Notably, HVO\_2753 forms a monomer in solution, in contrast to another recently characterized  $\mu$ -protein from *H. volcanii* [19].

### Analysis by mass spectrometry

The homologously produced protein was analyzed by both top-down and bottom-up mass spectrometry (MS/MS). Top-down analysis of the protein determined the monoisotopic mass of the protein to be 7249.9940 Da ( $z = 5$  to  $z = 8$ ) (Fig. 4). This is in excellent agreement with the theoretical monoisotopic mass of 7250.1003 Da ( $\Delta$ mass  $-14.66$  ppm) for HVO\_2753 including the his<sub>6</sub>-tag. In addition, the combined use of both CID and HCD activation types across a range of NCE values resulted in 63% of possible inter-residue cleavages being observed (Fig. S5A). In spite of the high fraction of amino acids that could potentially be modified, the MS analysis revealed that the native protein lacks any posttranslational modifications, at least during exponential growth in complex medium. Further confirmation toward the absence of any non-transient modifications was determined via bottom-up MS analysis. The latter, in concert with the top-down assessment, resulted in 100% sequence coverage of the



**Fig. 4.** Analysis of HVO\_2753 via top-down mass spectrometry. (A) Intact protein measurement via direct infusion onto an LTQ Orbitrap Velos. The mass of the most abundant peak and related charge state of the detected ions ( $z$  values) are indicated for the respective peaks. (B) A representative deconvoluted and deisotoped monoisotopic MSMS spectrum for a 1210.00  $m/z$  6 + precursor ion. The mass values are indicated at the respective peaks.

HVO\_2753\_His<sub>6</sub>-tagged protein and further confirmed the identification of the protein without modifications (Fig. S5B).

### Heterologous production in *E. coli*

For NMR analyses, HVO\_2753 was also produced heterologously in *Escherichia coli*. It was produced as a fusion protein with a GST tag, which enabled affinity purification as a first isolation step. The tag was removed by treatment with TEV protease, and the protein was further purified by size exclusion chromatography (Fig. S6). The two-step procedure resulted in an extremely pure protein, which lacked the N-terminal methionine and instead had four additional amino acids at its N terminus (GAMG), which were numbered from  $-4$  to  $-1$  to ensure an identical numbering of the protein obtained after homologous and heterologous production. The production in *E. coli* offers the advantage of established procedures for labeling the protein with stable isotopes, which is essential for several NMR approaches.

### Structural characterization of the zinc finger protein

The comparison of the HVO\_2753 protein heterologously expressed in *E. coli* and isolated from *Haloferax volcanii* shows consistent chemical shifts for the two differently expressed proteins and those indistinguishable structures (Fig. S7). Further structural investigation via heteronuclear multidimensional NMR experiments of the zinc finger protein was conducted

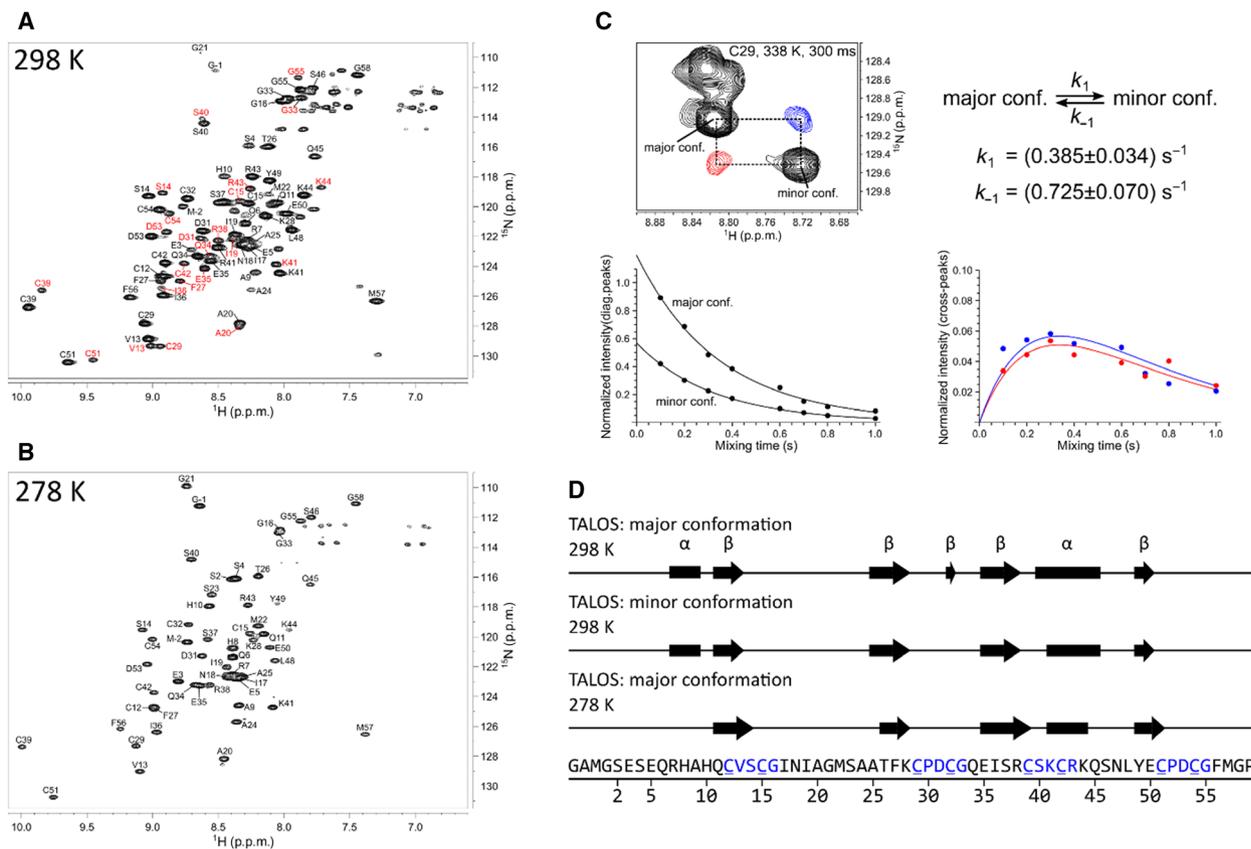
with heterologously expressed protein using <sup>13</sup>C/<sup>15</sup>N isotope labeling schemes.

High-salt conditions are typically required for growth of the halophilic archaeon *H. volcanii*, which is why the influence of 1 M NaCl on the folding motif of HVO\_2753 protein was examined. The chemical shifts of 2D <sup>1</sup>H/<sup>15</sup>N Best-TROSY [26,27] spectra recorded at low (200  $\mu$ M) and high-salt (1 M) concentrations are consistent, indicating no significant effect of NaCl on the protein folding propensities (Fig. S8). Thus, further NMR experiments were conducted at low-salt conditions preferable for NMR spectroscopy.

HVO\_2753 protein adopts a persistent structure rich of  $\beta$ -strand structure elements. At 298 K, 77 amide cross-peaks were observed in 2D <sup>1</sup>H/<sup>15</sup>N BEST-TROSY spectrum, which corresponds to 32% more signals than initially expected (58 expected signals). The additional peaks indicate the presence of a second conformation. The analysis of the signal intensity defines two conformations with a population ratio of approximately 4:1.

### NMR backbone assignment and TALOS prediction

The <sup>1</sup>H, <sup>15</sup>N, and <sup>13</sup>C chemical shift assignment of HVO\_2753 was performed manually using standard double- and triple-resonance NMR experiments at 298 K. The backbone assignment of the major conformation was completed to 97%. 23 residues show a splitting pattern of the signals, indicative for two conformations in slow exchange on the NMR time scale



**Fig. 5.** Temperature dependence of NMR spectra of HVO\_2753 protein. (A, B) 2D  $^1\text{H}^{15}\text{N}$  HSQC spectrum recorded at 800 MHz 298 K (A) and 278 K (B). Assignment of the major conformation is highlighted in black, and the minor conformation is colored in red. The sample contains 25 mM Bis-Tris buffer pH 7, 200 mM NaCl, 5%  $\text{D}_2\text{O}$ , 0.5 mM DSS. C (top)—Expansion of the ZZ-exchange spectra on the example of the residue V13 at mixing times 300 ms at 338 K. (bottom, left) RDecay of the diagonal peaks and (bottom, right) buildup of the exchange cross-peaks between the major and minor conformations. D – TALOS-N secondary structure prediction for major and minor conformations at 298 K and the single major conformation at 278 K. Secondary structure elements are shown as a function of the residue number. Zinc binding motifs are colored in blue.

(Fig. 5A). The population ratio of the coexisting conformers is temperature-dependent (Figs S9 and S10).

In order to analyze the kinetics of the interconversion between the major and the minor conformation, we performed temperature-dependent 2D  $^{15}\text{N}$ -ZZ-exchange NMR experiments [28,29]. Cross-signals correlating the two populations could be observed at 338 K, with a population ratio of two conformations close to 1 : 1 (Fig. 5C). For three residues (C29, C39, and G58), cross-signals between the major and the minor conformation could be resolved (Fig. S11). The exchange rate  $k_{\text{ex}}$ , calculated as the sum of the forward and reverse rate constants, was calculated as the average of the values of the residues C29, C39, and G58 and is around  $1.2 \text{ s}^{-1}$ . With decreasing temperature, the exchange rates increased beyond the detectable range of 0.1 to  $10 \text{ s}^{-1}$  in ZZ-exchange experiment [30].

The presence of the minor conformation hampers the 3D structure calculation of the protein. Multiple signals from the minor conformation overlap with the peaks from the major conformation, and the automated restricted peak picking procedure by SPARKY 3.114 [31] becomes challenging. Even more crucial is the impeded NOESY-based analysis of two populations, where the signal intensity directly relates to the distance restraints. Therefore, we screened for conditions to observe mainly one conformation.

Since HVO\_2753 the major conformation is favored at low temperature (278 K), the chemical shift assignment and eventually the structure calculation were performed at 278 K (Figs 5B, Fig. S9). Chemical shift assignments of the backbone for the well-defined rigid core and the full length were completed to 99% and 97%, respectively, while the overall assignment was achieved to 88% and 86%, respectively.

For structural comparison of different conformations, the backbone resonance assignments of both major and minor population at 298 K and of the single major conformation at 278 K were used to predict the secondary structure elements of the protein with TALOS-N software [32] (Fig. 5D). Besides the N-terminal  $\alpha$ -helical region, which in the major conformation is present at higher temperature and is absent at 278 K, no significant differences in secondary structure elements between the two conformations could be detected.

### Quantification of zinc binding and its influence on protein folding

Typically, zinc ions stabilize the native structure of zinc finger proteins. Since HVO\_2753 protein contains two zinc finger motifs in the sequence and NMR structural analysis showed two temperature-dependent conformations of the protein, the questions how zinc ions are involved in the protein folding pathway and how many zinc ions are incorporated in the binding pockets are particularly important. In order to answer these questions, zinc ions taken up by the protein even without extra addition of zinc to the growth medium were first removed from the protein to allow for titration experiments to determine binding stoichiometry. Structural changes in the protein induced by the zinc ions were monitored by 2D  $^1\text{H}^{15}\text{N}$  HSQC spectra. Complete removal of zinc ions turned out to be nontrivial; it was achieved by incubating the protein with three-fold excess of EDTA and dialysis against EDTA at 338 K for four hours. The high temperature required for EDTA to chelate all the zinc ions indicates higher affinity of the protein for zinc ions compared to EDTA ( $K_D = 10^{-16}$  M) [33]. Eventually, the unfolded metal-free protein was titrated with  $\text{Zn}^{2+}$  ions in 0.25 equivalent steps at 298 K (Fig. 6). Addition of 0.25 equivalents of  $\text{Zn}^{2+}$  ions already induces the refolding of the protein. Remarkably, both conformations refold simultaneously appearing with the same population ratio (4: 1). The native structure of the protein is completely restored by 0.75 equivalents of  $\text{Zn}^{2+}$ . Additional zinc ions do not change the fold or the population ratio of the protein any further. This experiment shows that despite the presence of two zinc finger motifs in the HVO\_2753 sequence, only one zinc ion is bound to the protein.

The unexpected finding that only one zinc ion is bound despite the two zinc finger motifs was obtained with the protein after heterologous production in *E. coli*. Because it could not be excluded that *E. coli* lacks a potential zinc incorporation chaperone, the

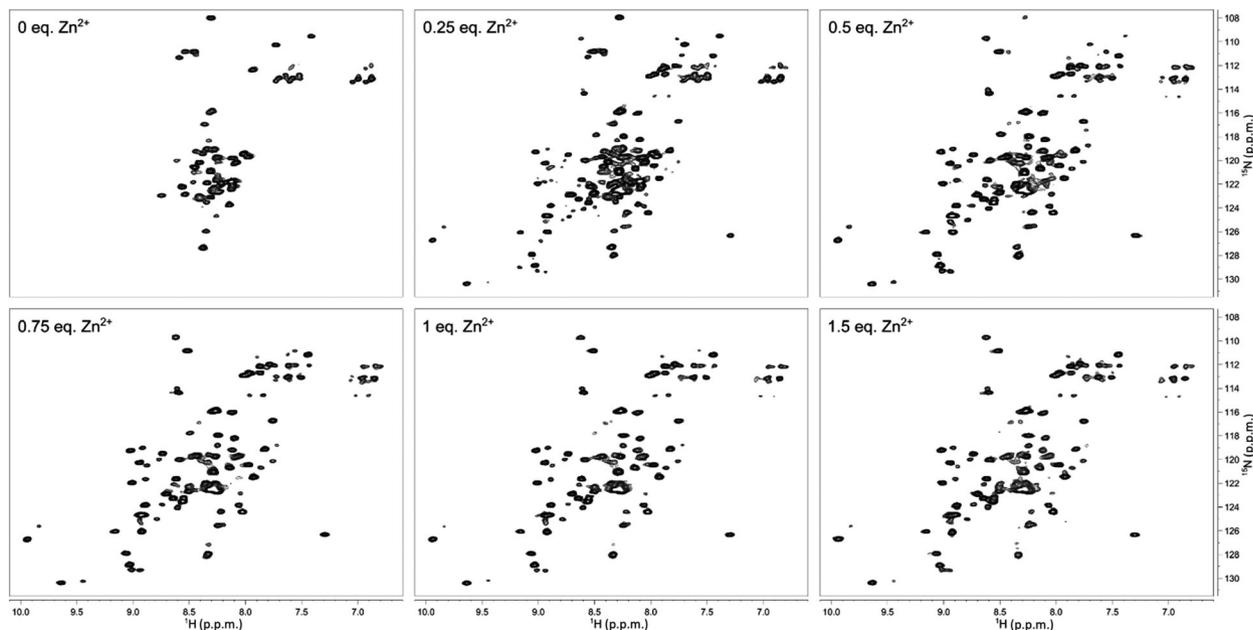
zinc content was also quantified for the protein after homologous production in *H. volcanii*. HVO\_2753 was purified under native conditions (see above) and dialyzed against a low-salt buffer. Zinc was quantified with the fluorophore ZnAF-2F, which is highly specific for zinc and discriminates against all other ions [34]. Only a very small amount of zinc could be detected, underscoring that HVO\_2753 is stable under low-salt conditions, in contrast to many other haloarchaeal proteins. Proteinase K was used to hydrolyze the protein to liberate bound zinc. Quantification with the fluorimetric assay led to the detection of 0.7 equivalents of zinc (Fig. S12), in excellent agreement with the NMR results. Taken together, two very different experimental approaches revealed that HVO\_2753 binds only one zinc ion, in contrast to the expectation based on the presence of four C(P)XCG motifs.

### Point mutations reveal the essentiality of all four C(P)XCG motifs

The binding of only one zinc ion opened the possibility that only two of the four C(P)XCG motifs might be required. To test this possibility, the first cysteine in each of the four motifs was replaced by alanine. The mutant proteins were produced in the HVO\_2753 deletion mutant, and the phenotype of the resulting strains was tested. It turned out that none of the four mutant proteins could complement the deletion mutant, and all four strains were totally devoid of swarming (Fig. S13). It was attempted to purify the four homologically produced mutant proteins and quantify their zinc contents. However, it turned out that the protein levels were very low, indicating that all four mutant proteins were misfolded *in vivo* and rapidly degraded.

Two further amino acids were arbitrarily chosen (Q34 and Y49), and single amino acid mutants were generated. Also, these two mutant proteins could not complement the HVO\_2753 deletion mutant (Fig. S13). This result shows that also residues outside of the C(P)XCG motifs are essential for function, and, in addition, that single amino acid replacements are suitable for a future in-depth scanning mutagenesis approach for HVO\_2753 characterization.

We also induced mutations into the ORF of the HVO\_2753 gene that had been introduced into an expression vector for *E. coli*. In contrast to the homologous production described above, heterologous production and purification turned out to be feasible. However, NMR analysis revealed that all mutations lead to severe loss in structure and stability, and, instead, the mutant proteins adopted molten globule-like conformations (Fig. S14).



**Fig. 6.** Zinc titration series of HVO\_2753 Zn-finger protein. 2D  $^1\text{H}^{15}\text{N}$  HSQC spectrum recorded at 800 MHz, at 298 K. The sample contains 25 mM Bis-Tris buffer pH 7, 200 mM NaCl, 5%  $\text{D}_2\text{O}$ , 0.5 mM DSS. Complete removal of zinc was achieved by incubating the protein with threefold excess of EDTA followed by the dialysis against EDTA at 338 K for 4 h. The unfolded metal-free protein was then titrated with  $\text{Zn}^{2+}$  ions in 0.25 equivalent steps. Both conformations refold simultaneously upon addition of 0.25 equivalents of  $\text{Zn}^{2+}$  ions. 0.75 equivalents of  $\text{Zn}^{2+}$  are enough for the formation of the native structure.

Taken together, characterization of proteins with single amino acid replacements after homologous as well as heterologous production clearly showed that all four C(P)XCG motifs are essential for protein stability, folding, and functionality, despite the fact that only one zinc ion is bound.

### Dynamic studies of HVO\_2753

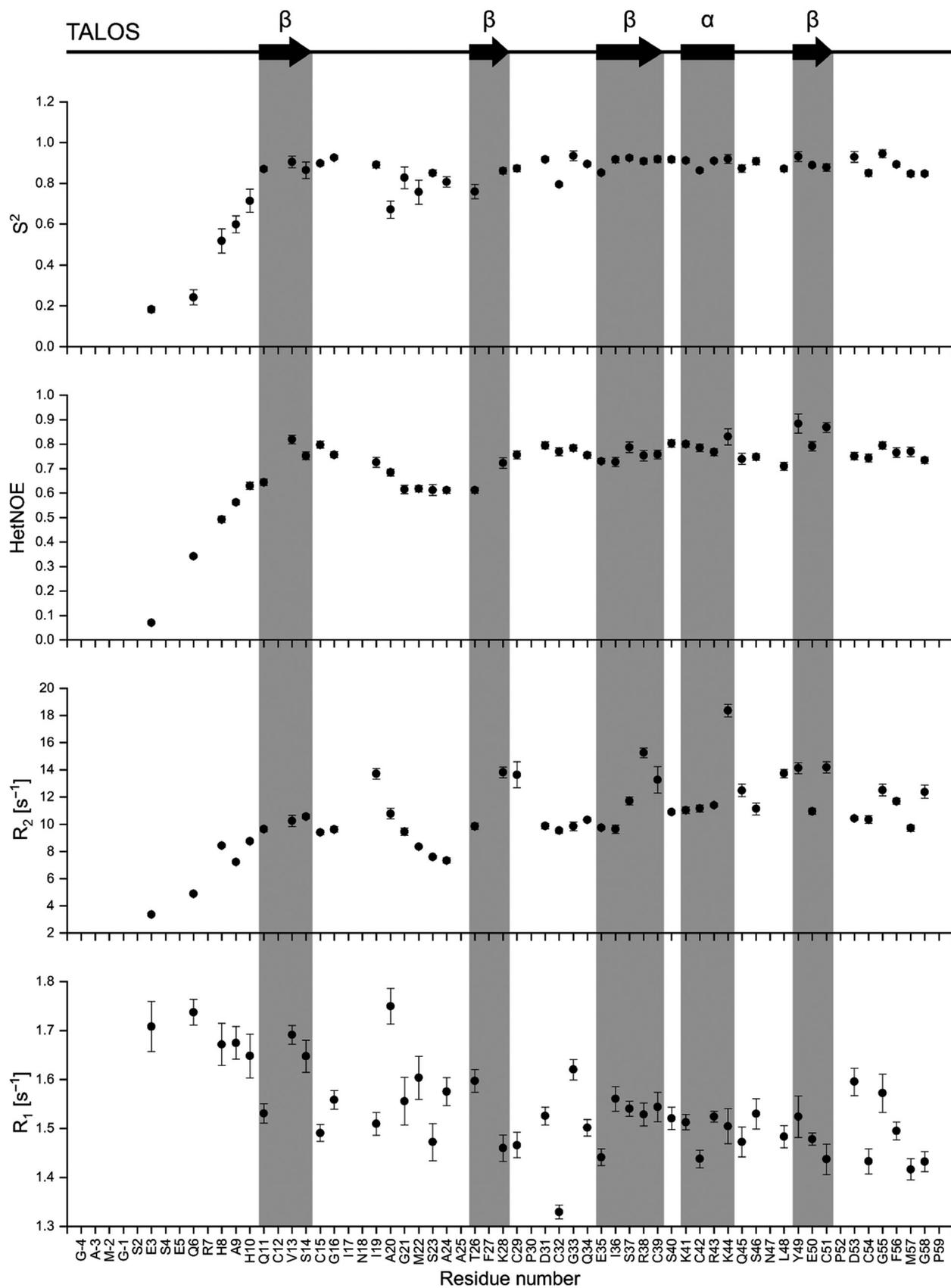
We further characterized the dynamics of HVO\_2753 by conducting relaxation experiments and analysis of intramolecular hydrogen bond formation (Figs 7, S15 and Table S4). Dynamics studied by heteronuclear  $^{15}\text{N}$  relaxation measurements provide values of  $T_1$ ,  $T_2$ , and  $\{^1\text{H}\}$ - $^{15}\text{N}$  het-NOE relaxation parameters, which are then used to determine the Lipari-Szabo order parameter  $S^2$  and the experimental rotational correlation time ( $\tau_c$ ) by the TENSOR2 [35] software.

The N terminus, which includes the first 13 residues, shows low het-NOEs and order parameters, indicating

high flexibility compared to the more rigid core of the protein (Fig. 7). The experimental rotational correlation time ( $\tau_c$ ) of the rigid core is 5.1 ns, which is characteristic of a protein of this size.

Intramolecular hydrogen bond formation was monitored by temperature-dependent amide proton chemical shift perturbations. A series of 2D  $^1\text{H}^{15}\text{N}$  BEST-TROSY spectra was recorded for the temperature range from 278 to 348 K in 10 K increments. The temperature coefficients of the amide protons were calculated from a linear fit of the chemical shift perturbation as a function of the temperature [36]. Temperature coefficients with values more negative than  $-4.5$  ppb/K are characteristic for the rapidly exchangeable and therefore not hydrogen-bonded amide protons, while protons with a temperature coefficient higher than  $-4.5$  ppb/K are involved in hydrogen bonding [36]. The temperature coefficients of the N terminus of the HVO\_2753 protein are consistent with the relaxation data, showing a flexible unstructured part of the protein (Fig. S15).

**Fig. 7.** Backbone dynamic studies of HVO\_2753 protein. On the top, a schematic representation of the NMR chemical shift-based TALOS-N secondary structure prediction is shown. The analysis of the order parameter ( $S^2$ ) was performed using TENSOR2 [35] based on the below presented experimentally measured relaxation parameters:  $\{^1\text{H}\}$ - $^{15}\text{N}$  heteronuclear NOE,  $R_1$ , and  $R_2$ . Relaxation experiments were performed at 278 K, 600 MHz.



### 3D structure determination

The structure calculation of HVO\_2753 protein was performed by fully automated NOESY cross-peak assignment using CYANA [37–39]. NOE cross-peaks were obtained by restricted peak picking with Sparky 3.114 [31]. 32 pairs of  $\varphi/\psi$  backbone dihedral angles obtained from TALOS-N software, 42  $^3J(\text{H}^N, \text{H}^\alpha)$ -coupling constants, and 8 hydrogen bond constraints were used for structure calculation. The 20 NMR structures with the lowest energy represent the 3D NMR solution structure of HVO\_2753 (Fig. 8). The structural statistics are shown in Table 1. The atomic coordinates have been deposited in the protein data bank (PDB ID: 6YDH).

HVO\_2753 consists of a well-folded rigid core and unstructured flexible N terminus. The unstructured part comprises the first 13 amino acids. The core domain shows two twisted antiparallel  $\beta$ -sheets and one helix. The first  $\beta$ -sheet is composed of two  $\beta$ -strands including residues Q34 to R38 and A25 to C29, while the second one includes amino acids Y49 to C51 and F56 to G58. The  $\alpha$ -helix is formed by the amino acids C42 to K44 and prolonged at its ends by two residues forming  $3_{10}$ -helix.

According to the calculated structure, two zinc binding pockets (ZBP) are possible. The first one (ZBP1) involves binding motifs 1 (CVSCG) and 3 (CSKCR), while the second binding pocket (ZBP2) consists of

binding motifs 2 (CPDCG) and 4 (CPDCG). Regarding cis/trans-peptide bond conformation of the proline residues using the experimental chemical shift assignment and the prediction tool Promega [40,41], prolines P30 and P52 in ZBP2 adopt a trans-conformation, while the C-terminal P59 adopts a cis-conformation (Table S3).

In order to identify the occupied binding pocket, CYANA structure calculation was performed separately with manually defined ZBP1 and ZBP2, which were further compared with a structure without a strictly defined binding pocket (Fig. S16). Each binding pocket was defined by setting the additional six upper distance restraints between four coordinating sulfur atoms from the cysteine residues to 4 Å, thus forming a symmetrical tetrahedral ZBP. No significant differences in the RMSD value could be detected between the structures without restricted ZBP and with defined ZBP2 (about 0.25), while the structure with defined ZBP1 shows a larger RMSD value (0.33). Moreover, the structure calculation of the last one was accompanied by Ramachandran plot outliers in all 20 structures. The affected residue is C15, which is directly involved in the coordination of the zinc ion in this structure. We take this as evidence that ZBP2 carries the zinc cation but not ZBP1.

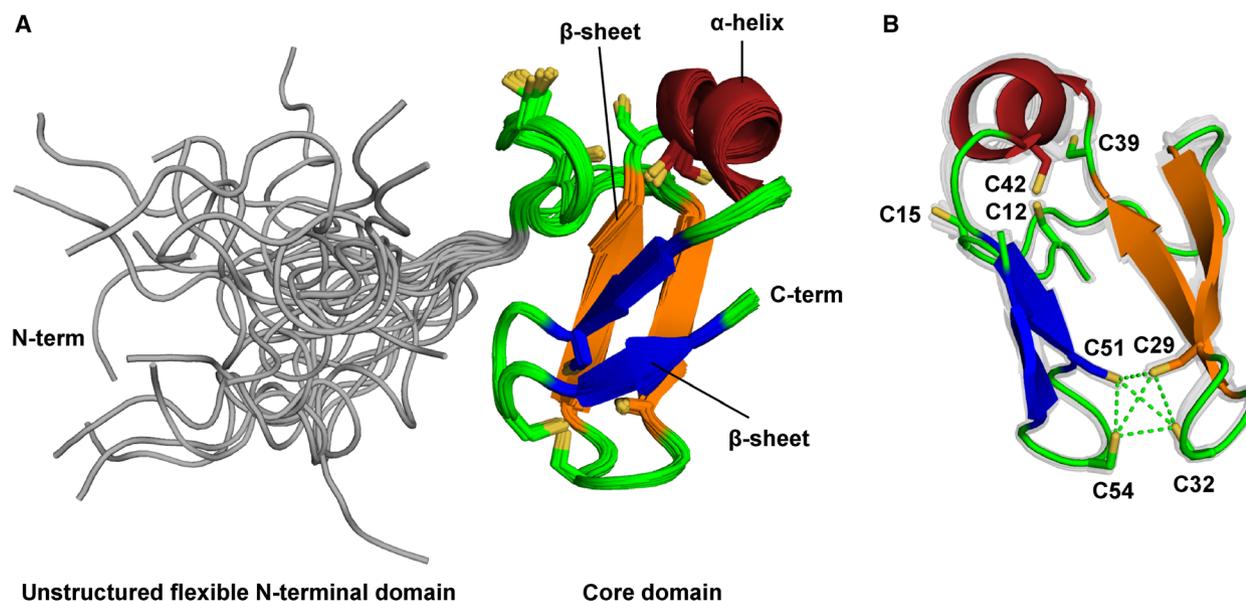
Further indication that ZBP2 contains the metal ion provides  $^{13}\text{C}$  NMR chemical shift analysis of cysteines.  $^{13}\text{C}$   $\text{C}_\alpha$  and  $\text{C}_\beta$  NMR chemical shifts are sensitive to the redox state of the cysteines and can, therefore, be used to predict the reduced and oxidized form of cysteines. Furthermore, chemical shift analysis allows distinguishing between reduced zinc coordinating and noncoordinating cysteines [42]. According to the chemical shift assignment of the cysteines (Fig. S17 and Table S2), all cysteine residues except for C42 are in the zinc-ligated region. Cysteine 42 shows characteristic downfield and upfield shifts of the  $\text{C}_\alpha$  and  $\text{C}_\beta$  atoms, respectively, indicating the reduced metal nonbound form and, therefore, a free binding pocket.

In order to validate the structure and in particular the binding pockets, CYANA structure calculation was performed with excluded dihedral angles restraints, derived from the TALOS-N [32] predictions for both cases, with defined ZBP1 and ZBP2. From the resulted bundle of the 20 best structures, the backbone chemical shifts were predicted by SHIFTX2 [43] and further compared with the experimental set of resonances. The analysis was performed by calculation of the root mean square error (RMSE) and the mean square error (MSE) for both structures. RMSE and MSE of the structures with fixed ZBP1 and ZBP2 were

**Table 1.** Structural statistics for the ensemble of 20 best NMR structures of full-length HVO\_2753

<i>Experimental restraints</i>	
NOE distances	1016
short range ( $ i-j  \leq 1$ )	682
medium range ( $1 <  i-j  < 5$ )	74
long range ( $ i-j  \geq 5$ )	260
Hydrogen bonds	8
Backbone torsion angles ( $\varphi, \psi$ )	64
Diffusion anisotropy ( $T_1/T_2$ )	45
Scalar couplings ( $^3J(\text{H}^N, \text{H}^\alpha)$ )	42
Upper length of S-S bond in zinc binding pocket (Å)	4
Number of S-S restraints in zinc binding pocket	6
<i>Ramachandran score (%)</i> <sup>a</sup>	
most favored	73.9
additionally allowed	26.0
generously allowed	0.0
disallowed	0.1
<i>Average rmsd to mean (Å)</i> <sup>b</sup>	
backbone	0.27 ± 0.07
all atom	0.73 ± 0.07

<sup>a</sup>For the ensemble of 20 structures of the full length; <sup>b</sup>For the ensemble of 20 structures of the rigid core



**Fig. 8.** Nuclear magnetic resonance solution structure of HVO\_2753 protein. (A) Ribbon representation of the ensemble of the best 20 structures as a full length. (B) Ribbon representation of the lowest energy structure of the rigid core of the protein rotated by 90°. Flexible N terminus is highlighted in gray,  $\beta$ -sheets are colored in blue and orange,  $\alpha$ -helix is shown in red, sulfur atoms from corresponding cysteine residues are represented as yellow sticks and manually defined binding pocket is shown as green dashed lines. The figure was generated by PyMOL.

calculated for the rigid core and the binding pockets region separately. The rigid core excludes the N-terminal flexible 13 amino acids, while the binding pocket region involves only 20 residues from four binding motifs. The RMSE values of the backbone chemical shifts for the structure with manually defined ZBP1 do not show significant differences to the structure with the fixed ZBP2 (Fig. S18). However, the MSE value of the binding pocket region between fixed ZBP1 and ZBP2 shows overall higher errors for the structure with fixed ZBP1 (Fig. S19). For instance, MSE of the fixed ZBP1 is 22% higher for the CB resonances and by a factor of two higher for the HA chemical shifts. Taken together, the results of the MSE structure validation indicate that ZBP2 is preferred and, therefore, most likely occupied by the zinc ion.

### Analysis of the HVO\_2753 minor conformation

In order to investigate the nature of the minor conformation, we performed CSP (chemical shift perturbation) analysis, which was further displayed by mapping the affected residues on the calculated solution structure of the rigid core domain (Fig. S20). All affected CSPs are located predominantly in the binding motifs. However, according to the TALOS prediction, no significant differences in the secondary structure elements between major and minor

conformations could be detected (Fig. 5D). This indicates that the minor population is characterized by the slight movement of the binding motifs, without affecting the overall structure.

Furthermore, detailed analysis of the relaxation data revealed that several residues exhibit faster relaxation rate  $R_2$  compared to the average value of the rigid core of the protein ( $9.5 \pm 0.2 \text{ s}^{-1}$ ). This contribution to the transverse relaxation indicates the presence of chemical exchange. Almost all amino acids with  $R_2$  rates higher than  $13 \text{ s}^{-1}$  were observed to adopt the second minor conformation at higher temperature. The highest  $R_2$  values show residues K44 and R38 with  $18.4 \pm 0.4 \text{ s}^{-1}$  and  $15.3 \pm 0.3 \text{ s}^{-1}$ , respectively (Fig. S21). Both are involved in the formation of the minor conformation. Interestingly, among those residues are two cysteines, C29 and C51, which build the zinc binding pocket 2 (ZBP2) and are the first amino acids in the CPDCG zinc binding motifs. These residues were found to be essential for the protein stability and function.

### Analysis of the HVO\_2753 binding pockets

The analysis of the intramolecular interactions of the cysteines revealed that each residue is involved in the formation of several long-distance NOEs. The recording of a 3D  $^1\text{H}^1\text{H}^{15}\text{N}$  NOESY-HSQC experiment at a

950 MHz spectrometer with the mixing time of 120 ms allowed us to observe cross-peaks corresponding to a distances of up to 5.7 Å (Figs S22 and S23). Remarkably, C12 and C39 residues, which despite an unoccupied ZBP1 were found to be essential for the protein folding, show two intramolecular distances each. These interactions are important for the tertiary structure and are directly involved in structure formation of the protein.

## Discussion

In this contribution, we describe an in-depth analysis of the  $\mu$ -protein HVO\_2753 from *H. volcanii*. Until now, only very few archaeal  $\mu$ -proteins have been characterized. A very early analysis of the low molecular weight proteome led to the identification of 370 small proteins of less than 20 kDa [11]. However, only two of these proteins were further characterized. A 60 aa  $\mu$ -protein with an unusual C3H zinc finger was found to be important for the expression of the bacterioopsin (bop) gene, and, accordingly, it was named Brz (bacteriorhodopsin-regulating zinc finger protein) [17]. A 55 aa  $\mu$ -protein was named 'bacteriorhodopsin-regulating basic protein' (Brb), because it was found that Brb cooperates with Brz and another protein, Bat, in the transcriptional regulation of the bacterioopsin gene [44].

Optimization of mass spectrometric techniques was used to systematically enhance the identification of  $\mu$ -proteins of *Methanosarcina mazei* Gö1 together with the analysis of the standard proteome [45]. Two different approaches led, in summary, to the simultaneous identification of 28  $\mu$ -proteins and nearly 1900 standard proteins. In another study, LC-MSMS was used to verify the presence of  $\mu$ -proteins in *M. mazei* Gö1 [46]. The concentrations of two  $\mu$ -proteins, Sp36 and Sp41, were found to be elevated during nitrogen starvation, indicating that these  $\mu$ -proteins are involved in (the regulation of nitrogen) metabolism.

The annotation of the microproteome of *H. volcanii*, the species used in this study, is more comprehensive than that of most other archaeal species, because the results obtained with *H. salinarum* discussed above could be transferred for all  $\mu$ -proteins present in both species. The annotation contains 282 proteins with a length of up to 70 aa and 575 proteins with a length of up to 100 aa. Two recent studies characterized several of these  $\mu$ -proteins: 1) 16 single gene deletion mutants of zinc finger proteins were generated and their phenotypes were characterized [18]. For 12 of the mutants, differences to the wild-type were observed. 2) The NMR structure of one  $\mu$ -protein was solved, which was found to be a dimer in solution [19].

It should be noted that the characterization of HVO\_2753, which is presented here, combines genetic, biochemical, and structural approaches and is more comprehensive than studies mentioned above. The protein was structured and stable in low salt, in contrast to typical haloarchaeal proteins, which require high-salt concentrations for stability [47]. However, low-salt stability has been described for several other proteins from *H. volcanii*, for example, the  $\mu$ -protein HVO\_2922 [19] and tRNase Z [48], which, surprisingly, was found to be only active at low salt.

The HVO\_2753 deletion mutant had no growth defect in complex or synthetic media, indicating that HVO\_2753 is not important for central processes like transcription or translation, in contrast to expectations based on its co-transcription with the translation elongation factor HVO\_2752. Interestingly, a quantitative proteomic study has revealed that the differential protein level patterns of the two proteins under different conditions are not the same, but very dissimilar [49]. The HVO\_2753 level is severely downregulated under several different stress conditions, while, in contrast, the HVO\_2752 level is upregulated at low and high temperatures. This indicates differential regulation of translation or/and protein stability of HVO\_2753 and HVO\_2752, which are encoded in one operon.

The HVO\_2753 deletion mutant revealed that the protein is essential for swarming and biofilm formation. The swarming phenotype could be complemented with the wild-type protein, but with none of four C-A mutants, which covered all four C(P)XCG motifs. In addition, protein levels of the four mutant proteins were very low, indicating that they are misfolded and unstable *in vivo*. In excellent agreement with these results, NMR analyses revealed that all four C-A mutant proteins are not structured, but were in the molten globule state.

The essentiality of all four C(P)XCG motifs was expected as long as it was thought that HVO\_2753 binds to zinc ions and has two zinc fingers. However, very unexpectedly, using two very different methods it was revealed that it binds only one zinc ion, in spite of the four motifs. Both methods found about 0.75 equivalents of zinc, which is a somewhat lower than one. The difference is most probably due to a slight overestimation of the protein concentration. The protein does not contain any tryptophan, making a spectroscopic protein quantification impossible due to the low extinction coefficient. Therefore, standard protein quantification assays were applied, which are a bit less accurate. In any case, it can be safely excluded that HVO\_2753 binds two zinc ions.

The NMR solution structure of HVO\_2753 was solved and revealed several interesting features: 1) The

N-terminal 13 is are totally mobile and unstructured. This is in agreement with the fact that they are found only in halophilic, but not in other archaea. This N-terminal extra part is very hydrophilic, and it is probable that it enhances the solubility in the high-salt cytoplasm of *H. volcanii*. Another protein, ferredoxin form *H. salinarum*, also contains an hydrophilic extra domain compared with nonhalophilic homologs [50]. For Fdx, it could be experimentally proven that this extra domain is essential for the halophilic character of the protein, including stability and formation of the Fe-S cluster. 2) The four C(P)XCG motifs form two zinc binding pockets (ZBPs), and at first sight, the 3D localization of the eight cysteines did not explain why only one is able to bind zinc. 3) The two ZBPs are not formed by C(P)XCG motifs 1 + 2 and 3 + 4, as expected, but by clusters 1 + 3 and 2 + 4.

The presence of two ZBPs with differential zinc binding capacity is highly unusual; in fact, to our knowledge, HVO\_2753 represents the first such case. For the following reasons, we propose that ZBP2 is occupied by a zinc ion, while ZBP1 is empty:

- 1 In proteins with two C(P)XCG motifs, it is very highly conserved that one of the motifs contains a proline residue at the second position, and that both motifs contain a glycine residue at the fifth position. However, ZBP1 is formed by the motifs CVSCG and CSKCR, none of which includes a proline. In addition, motif 3 does not have the highly conserved glycine at the fifth position, but an arginine.
- 2 When additional distance constraints were included in the NMR solution structure to simulate zinc ion binding, for ZBP1 one of the cysteines (C15) became a Ramachandran plot outlier, while this was not the case for ZBP2. Furthermore, implementation of the additional restrains in order to model the occupied ZBP1 to the structure calculation resulted in larger RMSD value compared with an occupied ZBP2.
- 3 The chemical shift analysis showed that C42, which is part of ZBP1, is in the reduced form, indicating that ZBP1 is not occupied by zinc.
- 4 In ZBP2, the zinc ion binding site is formed by two hairpins between antiparallel beta-strands (Fig. 8). This zinc binding pocket is similar to the zinc ribbon folding group, which has been described earlier for zinc finger motifs in large proteins [15].

While the structures of various isolated zinc finger motifs from larger proteins have been reported, which represent very small fractions of the complete proteins, the structures of only very few complete small zinc binding proteins have been solved. Here, we present

the structure of an especially interesting and to date unique protein, which contains 2 ZBPs, only one of which does bind a zinc ion. However, the geometry of the other ZBP, lacking a zinc ion, is also essential for the folding, stability, and functionality of the protein, because all single C-A mutations were detrimental and the cysteines are involved in long-range interaction networks. Future work will unravel interaction partners and the molecular mechanism of HVO\_2753, and it will also include further analyses of the influence of single amino acids on the formation of a folded and functional structure.

## Materials and methods

### Databases and bioinformatic analyses

Homologous sequences to HVO\_2753 were identified in the NCBI nonredundant protein sequence database using BlastP. The top 500 hits were retrieved, and the phylogenetic distribution was analyzed.

Multiple sequence alignments of selected sequences were generated with MEGA-X [51] using the algorithm MUSCLE. The alignments were visualized using Jalview, which enables coloring according to different criteria.

Phylogenetic trees were generated with MEGA-X [51] using, respectively, the maximum parsimony and the maximum-likelihood approach. In both cases, 1000 bootstrap repetitions were calculated, and the fractions of cases (%) each node was retrieved were shown in the resulting consensus tree.

The Integrated Genome Browser [52] was used to visualize the manually curated genome annotation [53] as well as the results of a RNA-Seq study [23] and a dRNA-Seq study [24]. Gene sequences from the *H. volcanii* genome were retrieved from the database HaloLex [54]. The amino acid composition of the *H. volcanii* proteome was provided by Friedhelm Pfeiffer (MPI for Biochemistry, Martinsried, Germany).

### Strains, media, and culture conditions

The *H. volcanii* strain H26 [55] was used as a wild-type in this study. Generation of deletion and complementation strains is described below. The strains were grown in complex medium and in synthetic media with different carbon and energy sources as described [56,57].

The *E. coli* strain XL1-Blue MRF' (Agilent Technologies, Waldbronn, Germany) was used for cloning, and it was grown in standard media [58].

### Northern blot analysis

Northern blot analyses were essentially performed as described [59]. In short, *H. volcanii* cultures were grown in

complex medium to mid-exponential growth phase. Cells were harvested, and total RNA was isolated. 15  $\mu$ g of total RNA from each sample was separated on a denaturing formaldehyde agarose gel. The RNA was transferred by capillary blotting onto a nylon membrane, and fixed by UV-cross-linking. Digoxigenin-labeled single-stranded DNA probes were generated by PCR using DIG-dUTP and a dNTP mix with reduced a dTTP concentration. The primers for probe generation are listed in Table S5. Hybridization was performed overnight at 50 °C. The membrane was washed twice with 2  $\times$  SSC/0.1 % SDS and twice with 1 $\times$  SSC/0.1 % SDS. The probes were detected using an anti-DIG antibody coupled to alkaline phosphatase and the chemiluminescence substrate CDP star according to the instructions of the manufacturer (Roche, Mannheim, Germany). The signals were visualized with X-ray films (GE Healthcare), and the sizes were analyzed with the marker 'RiboRuler Low Range RNA Ladder' (Thermo Fisher Scientific).

### Generation of an *in frame* deletion mutant and its complementation

The *in frame* deletion mutant was generated by the so called Pop-In-Pop-Out method as described previously [55,60]. It should be noted that *in frame* deletions are important for genes that are situated in an operon, because polar effects on the expression of other can be avoided. The primers for the generation of the mutant are listed in Table S5. *H. volcanii* is highly polyploid with about 30 copies of the major chromosome [61], and thus, it is possible that not all copies of the wild-type chromosome are replaced by the deletion copies and heterozygous clones arise. Therefore, the homozygosity of the deletion mutant was verified both by Southern blot analysis and by multicycle PCR with isolated DNA as template.

For the complementation of the mutant, the HVO\_2753 gene was cloned into the expression vector pTA929 [62]. The vector contains a tryptophan-inducible promoter and enables modulation of the expression level. The primers that were used to amplify the gene with genomic DNA as template are listed in Table S5. The sequence of the resulting plasmid was verified by sequencing, and it was introduced into the deletion mutant by PEG-mediated transformation. As a negative control, the deletion mutant was also transformed with the empty vector.

### Analyses of growth, swarming, and biofilm formation

Growth analyses were performed in microtiter plates as described previously [57]. For each condition, 150  $\mu$ L medium was inoculated to an OD<sub>600</sub> of 0.05 from a preculture that had been grown under the respective condition. The cultures were grown on a Heidolph Titramax 1000 rotary shaker (Heidolph, Schwalbach, Germany) with 1100 rpm at

42° C. The OD<sub>600</sub> was determined using the microtiter plate photometer Spectramax 340 (Molecular Devices, Ismaning, Germany). Three biological replicates were performed, and average values and standard deviations were calculated.

The swarming assay was performed as described previously [18]. 5 mL of the respective medium with 0.3% (w/v) agar was filled into each well of six-well plates (Sarstedt, Nümbrecht, Germany) one day prior to start of the assay. The cells from precultures in the respective medium were harvested by centrifugation and resuspended in basal salts (medium without a carbon source) to an OD<sub>600</sub> of 20. 2  $\mu$ L of the cell suspension was injected deeply into the swarm agar in the center of the wells, because *H. volcanii* swarms only at low oxygen concentrations. The plates were incubated at 42°C in a Styrofoam box together with a glass of water to inhibit drying. Three biological replicates were performed, and averages and standard deviations were calculated.

Biofilm formation was quantified in 96-well plates. 150  $\mu$ L of cell suspensions was used for each well, and the plates were incubated at 42 °C without shaking for 24 h and 48 h. The principle of the method is to remove unattached cells, fix and stain cells in the biofilm, wash the biofilm, remove the bound stain, and quantify it spectroscopically. A detailed description has recently been published [18].

### Homologous production and native purification of HVO\_2753

For homologous production of the protein, the HVO\_2753 gene was cloned into the expression vector pSD1/R1-6 [25], which had been modified by the addition of a *NdeI* site. The vector contains a very strong promoter, which leads to a 30-fold higher trimethoprim resistance than a median promoter when it controls the expression of the *dhfr* (dihydrofolate reductase) gene [25]. The primers used to amplify the gene with genomic DNA as template are listed in Table S5. The primers added a *NdeI* site, the sequence for a N-terminal hexahistidine tag, and a *KpnI* site. The resulting plasmid was verified by sequencing and introduced in the  $\Delta$ HVO\_2753 deletion mutant by transformation. A preculture of the  $\Delta$ HVO\_2753 pSD1\_R1/6\_HVO\_2753\_NHis production strain was grown overnight in 30 ml complex media containing 0.4  $\mu$ g·mL<sup>-1</sup> novobiocin. 1.5 ml of the preculture at mid-exponential growth phase was used to inoculate 500 ml of complex medium. The production culture was grown overnight, and the cells were harvested by centrifugation (6500 g, 30 min, 4°C). The pellet was washed with 20 ml of basal salts, and it was suspended in 5 ml binding buffer (2.1 M NaCl, 20 mM HEPES pH 7.5, 20 mM imidazole, 1 mM PMSF). The cells were lysed by sonication on ice. The lysate was clarified by centrifugation at 16000  $\times$  g for 60 min at 4 °C. The supernatant was loaded on a Chelating Sepharose® Fast Flow (GE Healthcare) gravity column, which had

been charged with  $\text{Ni}^{2+}$  and equilibrated in binding buffer. The resin was washed with 20 ml of binding buffer to remove nearly all unspecifically bound proteins. Specifically binding proteins were eluted with 4 ml binding buffer containing 300 mM imidazole.

The subsequent size exclusion chromatography was performed using a Superose 6 FPLC column (10/300 GL, GE Healthcare). The mobile phase was 2.1 M NaCl, 20 mM HEPES pH 7.5, and the flow rate 0.3 ml. The proteins that were used to generate a standard curve were aprotinin (6.5 kDa), ribonuclease A (13.7 kDa), ovalbumin (44.3 kDa),  $\gamma$ -globulin (150 kDa), and thyroglobulin (670 kDa).

### Fluorimetric zinc binding assay

The zinc content of the protein was quantified using the fluorophore ZnAF-2F, which is highly specific for zinc ions and discriminates against all other bivalent metal ions [34]. HVO\_2753 was purified under native conditions as described above. The protein concentration was quantified using the BCA protein assay (<https://www.thermofischer.com>) with BSA as standard. 1  $\mu\text{M}$  HVO\_2753 was dialyzed against 25 mM NaCl and 20 mM HEPES at pH 7.5, because haloarchaeal proteins typically denature in low salt. As this treatment turned out to be insufficient for HVO\_2753, the protein was treated with 100  $\mu\text{g}\cdot\text{mL}^{-1}$  proteinase K at 37°C overnight. Measurements were performed using a microtiter plate fluorimeter with 492 nm as excitation wavelength and 517 nm for detection. 3  $\mu\text{M}$  ZnAF-2F was used, and a standard curve with different concentrations of  $\text{ZnCl}_2$  was generated alongside the HVO\_2753 samples. Four technical were used for each of the seven biological replicates, and average values and standard deviations were calculated.

### Direct infusion top-down MS analysis of HVO\_2753

Direct infusion analysis of the intact protein was performed on the LTQ Orbitrap Velos MS (Thermo Scientific, Germany) using the heated electrospray ionization source (HESI). Parameters were adjusted until a stable signal was acquired (voltage of approx. 4kV, flow rates between 5 and 15  $\mu\text{L}\cdot\text{min}^{-1}$ ). Both full scan and MS/MS analyses were performed at a target resolution of 60,000. Selective MS/MS fragmentation of various dominant multiply charged species was performed across a range of normalized collision energies (NCE) utilizing both CID (NCE: 20, 25, 30) and HCD (NCE: 17, 25, 30) activation types. MS raw data files were opened in Xcalibur Qual Browser (version 4.2.28.14) and summed regions for the various precursor ions at the specified NCE processed (deconvolution and deisotoped) using the Xtract algorithm. Peak lists for the

various summed spectra were exported, and combined lists were then assessed using the ProSight Lite software package (version 1.4).

### Bottom-up analysis by LC-MS

A 100  $\mu\text{g}$  aliquot of the sample was reduced (10 mM dithiothreitol, 56°C, 1 h) and alkylated (55 mM iodoacetamide, 20°C, 30 min). A 40  $\mu\text{g}$  aliquot was digested for 3 hours with 1  $\mu\text{g}$  of sequencing grade trypsin in 100 mM triethylammonium bicarbonate buffer (pH 8). Following digestion, the peptides were cleaned via solid phase extraction on a C18 cartridge (33 mg) (Waters, Germany), prior to a two-step elution off the cartridge with 100  $\mu\text{L}$  of 60 % acetonitrile (ACN) + 0.1% trifluoroacetic acid (TFA) and 100  $\mu\text{L}$  of 100% ACN. The peptides were dried via vacuum centrifugation (2h, 45°C) before resuspension in 50  $\mu\text{L}$  of HPLC loading buffer (3% ACN + 0.1% TFA). Aliquots of the reduced, reduced, and alkylated, and unprocessed protein were also collected during the sample preparation and diluted to a concentration of 0.1  $\mu\text{g}\cdot\mu\text{L}^{-1}$  in 10% ACN + 0.1% formic acid for direct injection analysis.

Chromatographic separation was performed on a Dionex U3000 nanoHPLC system (Thermo Scientific, Germany) equipped with an Acclaim PepMap 100 column (2  $\mu\text{m}$  particle size, 75  $\mu\text{m} \times 150$  mm) coupled online to a mass spectrometer. Eluent A: 0.05% formic acid (FA), eluent B: 80% ACN + 0.05% FA. Separation was performed over a programmed 60-minute run. Initial chromatographic conditions were 5% B for 2 minutes followed by linear gradients from 5% to 50% B over 30 minutes and a 5-minute increase to 95% B, and 8 minutes at 95% B. Following this, an inter-run equilibration of the column was achieved by 15 minutes at 5% B. A flow rate of 300  $\text{nl}\cdot\text{min}^{-1}$  was used, and 1  $\mu\text{L}$  of sample was injected per run.

Acquisition of data following separation was performed on the LTQ Orbitrap Velos mass spectrometer (Thermo Scientific, Germany) utilizing a combination of CID and HCD activation. A full scan MS acquisition was performed (resolution 60,000) with subsequent data-dependent MS/MS of the top 5 most intense ions fragmented using both CID (NCE 35) and HCD (NCE 35) (resolution 7,500); dynamic exclusion was enabled (30-sec duration).

MS raw data files were searched against protein fasta databases the complete UniProt *Haloferax volcanii* database (accessed from UniProt 7th November 2018) with the target HVO\_2753\_Histagged protein appended; in addition, the common laboratory contaminant database was appended. The searches were performed using the Proteome Discoverer software package (version 2.2.0.388) and the SequestHT search algorithm. Precursor mass tolerance was set to 7 ppm, with fragment mass tolerance of 0.02 Da. Searches were performed with semitrypsin specificity, maximum 3 missed cleavages, variable modifications of methionine oxidation and asparagine and glutamine

deamidation, and fixed modification of cysteine carbamidomethyl. Strict parsimony criteria have been applied (high stringency). A protein level FDR < 5% was applied, and at least two high confidence peptides were required, of which at least one is unique. Peptides were required to be of high confidence with an FDR < 1%.

### Heterologous production in *E. coli* and low-salt purification for NMR analyses

For heterologous expression in *E. coli*, the expression vector pGEX-CS was used, which leads to the production of fusion proteins with an N-terminal GST (glutathione S-transferase) tag, followed by a TEV (tobacco etch virus) protease cleavage site, followed by the respective protein of interest. The *NcoI* and *BamHI* restriction site was used to introduce the ORF of the *HVO\_2753* gene into the vector. The ORF was amplified by PCR using primers that added the respective restriction enzyme recognition motifs. The *NcoI* recognition motif, CCATGG, contains the start codon, but adds further nucleotides on both sides. This required the addition of one additional codon upstream (alanine) to stay *in frame* with the TEV recognition site and one additional codon downstream (glycine) to stay *in frame* with the *HVO\_2753* ORF, which started with the second codon. The TEV protease has the recognition site ENLYFQ\G, and it cleaves between the last two amino acids. Therefore, the four amino acid GAMG remain fused to *HVO\_2753* after TEV cleavage. They were numbered from -4 to -1, while the *HVO\_2753* sequence was numbered from +2 to +59, so that the amino acid numbering is identical for the heterologously produced and the homologically produced protein.

The resulting plasmid was used to transform T7 *Express E. coli* cells. The production strain was grown in M9 synthetic medium [58] to enable the incorporation of stable isotopes for some of the NMR experiments. 1 mM ampicillin was added to guarantee the presence of the plasmid. For labeling experiments,  $^{15}\text{N}$ -labeled  $\text{NH}_4\text{Cl}$  alone or together with  $^{13}\text{C}$ -labeled glucose (Cambridge Isotope Laboratories, Cambridge, MA, USA) replaced the unlabeled components of the M9 medium [58]. Cultures were grown at 37°C to an  $\text{OD}_{600}$  of 0.6, and expression of the fusion gene was induced by the addition of 1 mM IPTG. After a further incubation of 12 h, cells were harvested by centrifugation (5000 rpm, 15 min, 4°C). The cell pellet was resuspended in lysis buffer (25 mM Tris pH 8, 200 mM NaCl, 3 mM DTT), which was supplemented with 10% glycerol (v/v) and one protease-inhibitor tablet (cOmplete™, Roche, Germany) per liter. Cells were lysed by sonication, membranes and cell debris were removed by centrifugation, and the resulting cytoplasmic extract was used for protein purification.

The first step was affinity isolation using glutathione-sepharose beads [63]. The fusion protein was bound to the

beads via its GST tag, the beads were washed, and the protein was eluted with buffer containing 1 mM glutathione. Cleavage of the fusion protein with TEV protease was performed for 8 hours at 4°C. The liberated *HVO\_2753* was isolated from the GST tag and nonprocessed fusion protein by size exclusion chromatography. The identity and purity of the protein was confirmed using analytical SDS/PAGE and MALDI mass spectrometry (Fig. S6).

The QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies, Santa Clara, CA, USA) was used for the introduction of single amino acid replacements. Successful mutagenesis was confirmed by sequencing (Eurofins Genomics, Ebersberg, Germany). Purification of the mutant proteins was performed as described above for the wild-type protein.

### NMR spectroscopic experimental data

The protein samples were measured in NMR buffer containing 25 mM Bis-Tris pH 7, 200 mM NaCl, 3 mM DTT, 95%  $\text{H}_2\text{O}/5\%$   $\text{D}_2\text{O}$ . Spectra were recorded at 278 K and 298 K on Bruker spectrometers ranging from 600 to 950 MHz and equipped with z-axis gradient  $^1\text{H}\{^{13}\text{C},^{15}\text{N}\}$  triple-resonance cryogenic probes. The spectrometer was locked on  $\text{D}_2\text{O}$ .  $^1\text{H}$  chemical shifts were referenced to DSS at 0.00 ppm, and  $^{13}\text{C},^{15}\text{N}$  chemical shifts were calculated according to  $^1\text{H}$  frequency [64].

The chemical shift assignment was performed at 278 K and 298 K independently. For the sequential backbone resonance assignment, a set of 3D NMR experiments including BEST-TROSY-based HNCACB, HN(CO)CACB, HN(CA)CO, HNCO, and HNHA was used. NMR experiments performed at 298 K were complemented with additional HNN and HNNH 3D NMR experiments. Side chain assignment included (H)CC(CO)NH, H(C)CH-TOCSY, and  $^1\text{H}^1\text{H}^{15}\text{N}$  TOCSY-HSQC 3D NMR experiments. 3D  $^1\text{H}^1\text{H}^{15}\text{N}$  NOESY-HSQC (mixing time 120 ms, 950 MHz), 3D  $^1\text{H}^1\text{H}^{13}\text{C}$  NOESY-HSQC (mixing time 120 ms, 950 MHz), and 3D  $^1\text{H}^1\text{H}^{13}\text{C}$  NOESY-HSQC aromatic (mixing time 100 ms, 700 MHz) spectra were used for the NOE-based distance restraints.  $^3\text{J}(\text{H}^{\text{N}},\text{H}^{\alpha})$  coupling constants were calculated from 3D HNHA NMR experiment [65].

The temperature series of 2D  $^1\text{H}^{15}\text{N}$  BEST-TROSY spectra was measured from 278 to 348 K with 10 K increments on a 600 MHz spectrometer. Amide proton temperature coefficient was calculated from a linear fit of a chemical shift perturbation as a function of the temperature [36].

In order to analyze the slow exchange of the two present conformations, ZZ-exchange experiments were measured. Experiments were carried out at temperatures ranging from 298 K to 338 K with mixing times from 50 to 1000 ms. Cross-peaks could only be observed at 328 K and higher temperatures. The peak intensities were determined with Topspin 3.5, and values were normalized. Fits and exchange rates  $k_1$  and  $k_{-1}$  were calculated according to

Farrow et al. [29]. This was done for three well-resolved residues (C29, C39, and G58) at 338 K.

Heteronuclear  $^{15}\text{N}$  relaxation experiments ( $\{^1\text{H}\}$ - $^{15}\text{N}$  het-NOE,  $T_1$ , and  $T_2$ ) [12,23] were performed at 278 K on a 600 MHz spectrometer. The  $T_1$  longitudinal  $^{15}\text{N}$  relaxation time was calculated based on a series of spectra with the different relaxation delays: 20, 60, 100, 200, 400, 600, 800, 1200, 1500, and 1800 ms. The  $T_2$  transverse  $^{15}\text{N}$  relaxation time was recorded with the following delays: 16.96, 33.92, 67.84, 135.68, 169.60, 203.52, 237.44, and 271.36 ms. The  $\{^1\text{H}\}$ - $^{15}\text{N}$  het-NOEs were calculated from the signal intensity ratio ( $I_{\text{on}}/I_{\text{off}}$ ) obtained from the spectra recorded with and without saturation of amide protons with recovery ( $d_1$ ) and saturation delay of 3.5 seconds. The  $S^2$  order parameter was calculated using the TENSOR2 [12] software and the average of the first three best NMR structure models calculated with CYANA 3.97.

All NMR spectra were processed by using TopSpin version 3.5 (Bruker Biospin) and analyzed and visualized with SPARKY version 3.114 [31].

## Structure calculation

Structure calculations were performed by CYANA 3.97. The chemical shift assignment of the HVO\_2753 protein was performed manually and together with unassigned NOESY peak lists used as input for fully automated NOE cross-peak assignment. Therefore, three 3D NOESY NMR experiments were used: i) the aliphatic 3D  $^1\text{H}^1\text{H}^{13}\text{C}$ -NOESY-HSQC and  $^1\text{H}^1\text{H}^{15}\text{N}$ -NOESY-HSQC with a mixing time of 120 ms and ii) the aromatic 3D  $^1\text{H}^1\text{H}^{13}\text{C}$ -NOESY-HSQC with a mixing time of 100 ms. Peak picking of the spectra was performed by using the restricted peak picking function in the Sparky 3.114 program [31]. The NOESY peak lists were refined by manual inspection of the NOE cross-peaks.

The chemical shift tolerances were set to 0.02 for the bound protons and other protons and to 0.20 ppm for the heavy atoms. The final structure calculation included the NOE data, hydrogen bond distances, dihedral angles derived from the TALOS-N [9] predictions, and  $^3\text{J}(\text{H}^{\text{N}}, \text{H}^{\alpha})$  coupling constants (Karplus relation).

To imitate the coordination of the  $\text{Zn}^{2+}$  ion in the binding pocket, the upper distance restraints between four coordinating sulfur atoms from the cysteine residues were manually defined to 4 Å.

The structure calculation was performed using a standard protocol with 100 initial steps, 15000 refinement steps per iteration and 20 final structures. The final bundle of 20 top-ranked (lowest energy) structures was analyzed and validated, and the atomic coordinates were subsequently deposited in the protein data bank (PDB ID: 6YDH). The chemical shift assignment was deposited in the biological magnetic resonance bank (BMRB ID: 34500).

Structure validation was performed by calculating the root mean square error (RMSE) and mean square error (MSE) according to formulae 1 and 2, respectively:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (P_i - E_i)^2}{n}}, \quad (1)$$

where  $n$  is number of resonances,  $P$  is predicted value, and  $E$  is experimental value,

$$\text{MSE} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}, \quad (2)$$

where  $n$  is number of resonances,  $X$  is the difference between predicted and experimental value, and  $\bar{X}$  is the mean value.

Chemical shift perturbations (CSPs) between the minor and major conformations were calculated using the following formula:

$$\Delta\delta = \sqrt{(\Delta\delta_{\text{H}})^2 + (\alpha \cdot \Delta\delta_{\text{N}})^2}, \quad (3)$$

where  $\Delta\delta$  is the combination of  $^1\text{H}$ - and  $^{15}\text{N}$ -CSPs of the backbone amides,  $\Delta\delta_{\text{H}}$  and  $\Delta\delta_{\text{N}}$  are the chemical shift differences between major and minor conformations of  $^1\text{H}$ - and  $^{15}\text{N}$ -atoms, respectively, and  $\alpha$  is a correction factor for the gyromagnetic ratio ( $\alpha = 0.1$ ) [36].

## Acknowledgments

We thank Friedhelm Pfeiffer for generating the statistics of the amino acid composition of the proteome of *H. volcanii*. We also thank Hendrik R.A. Jonker for helpful scientific discussions on NMR experiments and structure calculation.

The collaborative project was funded by the German Research Council (Deutsche Forschungsgemeinschaft, DFG) in the framework of the Priority Program SPP2002 'Small proteins in prokaryotes' (www.spp2002.uni-kiel.de). The projects of the groups of JS and HS and the Z1 project headed by AT were involved, and the project numbers are JS264/26-1, SCHW 701/21-1, and TH872/10-1. Open access funding enabled and organized by ProjektDEAL.

## Conflict of interest

The authors declare no conflict of interest.

## Author contributions

All authors participated at conceptualization of the study and data analyses, SZ performed genetic and biochemical experiments, DP and NK performed NMR analyses, NK performed cyana structure calculation, LC performed top-down and bottom-up ESI and LC-ESI MS, LC, and AT performed MS data analysis, NK, AT, HS, and JS wrote the initial draft

of the analysis, all authors contributed to writing the manuscript.

## References

- Hartford CCR & Lal A (2020) When long noncoding becomes protein coding. *Mol Cell Biol* **40**, e00528–e619.
- Delcourt V, Staskevicius A, Salzet M, Fournier I & Roucou X (2018) Small proteins encoded by unannotated ORFs are rising stars of the proteome, confirming shortcomings in genome annotations and current vision of an mRNA. *Proteomics* **18**, e1700058.
- Mumtaz MAS & Couso JP (2015) Ribosomal profiling adds new coding sequences to the proteome. *Biochem Soc Trans* **43**, 1271–1276.
- Karunratanakul K, Tang H-Y, Speicher DW, Chuangsuwanich E & Sriswasdi S (2019) Uncovering thousands of new peptides with sequence-mask-search hybrid de novo peptide sequencing framework. *Mol Cell Proteomics* **18**, 2478–2491.
- Schrader M (2018) Origins, technological development, and applications of peptidomics. *Methods Mol Biol* **1719**, 3–39.
- Plaza S, Menschaert G & Payre F (2017) In search of lost small peptides. *Annu Rev Cell Dev Biol* **33**, 391–416.
- Cabrera-Quio LE, Herberg S & Pauli A (2016) Decoding sORF translation – from small proteins to gene regulation. *RNA Biol* **13**, 1051–1059.
- Duval M & Cossart P (2017) Small bacterial and phagic proteins. *Curr Opin Microbiol* **39**, 81–88.
- Baumgartner D, Kopf M, Klähn S, Steglich C & Hess WR (2016) Small proteins in cyanobacteria provide a paradigm for the functional analysis of the bacterial micro-proteome. *BMC Microbiol* **16**, 285.
- Storz G, Wolf YI & Ramamurthi KS (2014) Small proteins can no longer be ignored. *Annu Rev Biochem* **83**, 753–777.
- Klein C, Aivaliotis M, Olsen JV, Falb M, Besir H, Scheffer B, Bisle B, Tebbe A, Konstantinidis K, Siedler F et al. (2007) The low molecular weight proteome of *Halobacterium salinarum*. *J Proteome Res* **6**, 1510–1518.
- Eom KS, Cheong JS & Lee SJ (2016) Structural analyses of zinc finger domains for specific interactions with DNA. *J Microbiol Biotechnol* **26**, 2019–2029.
- Matthews JM & Sunde M (2002) Zinc fingers – folds for many occasions. *IUBMB Life* **54**, 351–355.
- Cassandri M, Smirnov A, Novelli F, Pitolli C, Agostini M, Malewicz M, Melino G & Raschella G (2017) Zinc-finger proteins in health and disease. *Cell Death Discovery* **3**, 17071.
- Krishna SS, Majumdar I & Grishin NV (2003) Structural classification of zinc fingers. *Nucleic Acids Res* **31**, 532–550.
- Maret W (2013) Zinc biochemistry. *Advances Nut* **4**, 82–91.
- Tarasov VY, Besir H, Schwaiger R, Klee K, Furtwängler K, Pfeiffer F & Oesterheld D (2008) A small protein from the *bop-brp* intergenic region of *Halobacterium salinarum* contains a zinc finger motif and regulates *bop* and *crtBI* transcription. *Mol Microbiol* **67**, 772–780.
- Nagel C, Machulla A, Zahn S & Soppa J (2019) Several one-domain zinc finger  $\mu$ -proteins of *Haloferax volcanii* are important for stress adaptation, biofilm formation, and swarming. *Genes* **10**, 361.
- 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.
- Wüthrich K (2003) NMR studies of structure and function of biological macromolecules (Nobel lecture). *Angew Chem Int Ed Engl* **42**, 3340–3363.
- Kay LE. (2005). NMR studies of protein structure and dynamics. *J Mag Res (San Diego, Calif: 1997)* **173**, 193–207.
- Kubatova N, Pyper DJ, Jonker HRA, Saxena K, Rimmel L, Richter C, Brantl S, Evgenieva-Hackenberg E, Hess WR, Klug G et al. (2020) Rapid biophysical characterization and NMR spectroscopy structural analysis of small proteins from bacteria and Archaea. *ChemBioChem* **21**, 1178–1187.
- Laass S, Monzon VA, Kliemt J, Hammelmann M, Pfeiffer F, Förstner KU & Soppa J (2019) Characterization of the transcriptome of *Haloferax volcanii*, grown under four different conditions, with mixed RNA-Seq. *PLoS One* **14**, e0215986.
- Babski J, Haas KA, Näther-Schindler D, Pfeiffer F, Förstner KU, Hammelmann M, Hilker R, Becker A, Sharma CM, Marchfelder A & et al. (2016) Genome-wide identification of transcriptional start sites in the haloarchaeon *Haloferax volcanii* based on differential RNA-Seq (dRNA-Seq). *BMC Genom* **17**, 629.
- Danner S & Soppa J (1996) Characterization of the distal promoter element of halobacteria *in vivo* using saturation mutagenesis and selection. *Mol Microbiol* **19**, 1265–1276.
- Schulte-Herbrüggen T & Sorensen OW (2000) Clean TROSY. *J Magn Reson* **144**, 123–128.
- Lescop E, Schanda P & Brutscher B (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J Magn Reson* **187**, 163–169.
- Montelione GT & Wagner G (1989) 2D Chemical exchange NMR spectroscopy by proton-detected heteronuclear correlation. *J Am Chem Soc* **111**, 3096–3098.

- 29 Farrow NA, Zhang O, Forman-Kay JD & Kay LE (1994) A heteronuclear correlation experiment for simultaneous determination of  $^{15}\text{N}$  longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J Biomol NMR* **4**, 727–734.
- 30 Palmer AG (2004) NMR characterization of the dynamics of biomacromolecules. *Chem Rev* **104**, 3623–3640.
- 31 Goddard TD & Kneller DB. SPARKY 3. University of California, San Francisco, USA.
- 32 Shen Y & Bax A (2013) Protein backbone and side chain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* **56**, 227–241.
- 33 Nyborg JK & Peersen OB (2004) That zinging feeling. *Biochem J* **381**, e3–4.
- 34 Hirano T, Kikuchi K, Urano Y & Nagano T (2002) Improvement and biological applications of fluorescent probes for zinc, ZnAFs. *J Am Chem Soc* **124**, 6555–6562.
- 35 Dosset P, Hus JC, Blackledge M & Marion D (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J Biomol NMR* **16**, 23–28.
- 36 Baxter NJ & Williamson MP (1997) Temperature dependence of  $^1\text{H}$  chemical shifts in proteins. *J Biomol NMR* **9**, 359–369.
- 37 Güntert P & Buchner L (2015) Combined automated NOE assignment and structure calculation with CYANA. *J Biomol NMR* **62**, 453–471.
- 38 Schmidt E & Güntert P (2015) Automated structure determination from NMR spectra. *Methods Mol Biol* **1261**, 303–329.
- 39 Güntert P (2003) Automated NMR protein structure calculation. *Prog Nucl Magn Reson Spectrosc* **43**, 105–125.
- 40 Schubert M, Labudde D, Oschkinat H & Schmieder P (2002) A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on  $^{13}\text{C}$  chemical shift statistics. *J Biomol NMR* **24**, 149–154.
- 41 Shen Y & Bax A (2010) Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts. *J Biomol NMR* **46**, 199–204.
- 42 Kornhaber GJ, Snyder D, Moseley HNB & Montelione GT (2006) Identification of zinc-ligated cysteine residues based on  $^{13}\text{C}$ alpha and  $^{13}\text{C}$ beta chemical shift data. *J Biomol NMR* **34**, 259–269.
- 43 Han B, Liu Y, Ginzinger SW & Wishart DS (2011) SHIFTX2. *J Biomol NMR* **50**, 43–57.
- 44 Tarasov V, Schwaiger R, Furtwängler K, Dyll-Smith M & Oesterhelt D (2011) A small basic protein from the *brz-brb* operon is involved in regulation of *bop* transcription in *Halobacterium salinarum*. *BMC Mol Biol* **12**, 42.
- 45 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.
- 46 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 Gö1. *Biochimie* **117**, 138–148.
- 47 Eisenberg H, Mevarech M & Zaccari G (1992) Biochemical, structural, and molecular genetic aspects of halophilism. *Adv Protein Chem* **43**, 1–62.
- 48 Hölzle A, Fischer S, Heyer R, Schütz S, Zacharias M, Walther P, Allers T & Marchfelder A (2008) Maturation of the 5S rRNA 5' end is catalyzed in vitro by the endonuclease tRNase Z in the archaeon *H. volcanii*. *RNA* **14**, 928–937.
- 49 Jevtić Ž, Stoll B, Pfeiffer F, Sharma K, Urlaub H, Marchfelder A & Lenz C (2019) The response of *Haloferax volcanii* to salt and temperature stress. *Proteomics* **19**, e1800491.
- 50 Marg B-L, Schweimer K, Sticht H & Oesterhelt D (2005) A two-alpha-helix extra domain mediates the halophilic character of a plant-type ferredoxin from halophilic archaea. *Biochemistry* **44**, 29–39.
- 51 Kumar S, Stecher G, Li M, Knyaz C & Tamura K (2018) *MEGA X. Molecular Biology and Evolution* **35**, 1547–1549.
- 52 Freese NH, Norris DC & Loraine AE (2016) Integrated genome browser. *Bioinformatics* **32**, 2089–2095.
- 53 Pfeiffer F & Oesterhelt D (2015) A manual curation strategy to improve genome annotation. *Life (Basel, Switzerland)* **5**, 1427–1444.
- 54 Pfeiffer F, Broicher A, Gillich T, Klee K, Mejía J, Rampp M & Oesterhelt D (2008) Genome information management and integrated data analysis with HaloLex. *Arch Microbiol* **190**, 281–299.
- 55 Allers T, Ngo H-P, Mevarech M & Lloyd RG (2004) Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Appl Environ Microbiol* **70**, 943–953.
- 56 Dambeck M & Soppa J (2008) Characterization of a *Haloferax volcanii* member of the enolase superfamily. *Arch Microbiol* **190**, 341–353.
- 57 Jantzer K, Zerulla K & Soppa J (2011) Phenotyping in the archaea. *FEMS Microbiol Lett* **322**, 123–130.
- 58 Green MR & Sambrook K (2012) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.
- 59 Kliemt J, Jaschinski K & Soppa J (2019) A haloarchaeal Small Regulatory RNA (sRNA) is essential for rapid adaptation to phosphate starvation conditions. *Front Microbiol* **10**, 1219.

- 60 Jaschinski K, Babski J, Lehr M, Burmester A, Benz J, Heyer R, Dörr M, Marchfelder A & Soppa J (2014) Generation and phenotyping of a collection of sRNA gene deletion mutants of the haloarchaeon *Haloferax volcanii*. *PLoS One* **9**, e90763.
- 61 Maurer S, Ludt K & Soppa J (2018) Characterization of copy number control of two *Haloferax volcanii* replication origins using deletion mutants and haloarchaeal artificial chromosomes. *J Bacteriol* **200**, e00517–17.
- 62 Allers T, Barak S, Liddell S, Wardell K & Mevarech M (2010) Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in *Haloferax volcanii*. *Appl Environ Microbiol* **76**, 1759–1769.
- 63 Harper S & Speicher DW (2008) Expression and purification of GST fusion proteins. *Curr Prot Protein Sci Chapter 6*, Unit 6.6.
- 64 Wishart DS (2011) Interpreting protein chemical shift data. *Prog Nucl Magn Reson Spectrosc* **58**, 62–87.
- 65 Vuister GW & Bax A (1993) Quantitative J correlation. *J Am Chem Soc* **115**, 7772–7777.

## Supporting information

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

**Figure S1.** Multiple sequence alignment (MSA) of HVO\_2753 from *H. volcanii* and 20 homologous proteins from selected halophilic and methanogenic genera and species.

**Figure S2.** Maximum parsimony (A) and maximum likelihood (B) phylogenetic trees of the multiple sequence alignment shown in Figure S1.

**Figure S3.** Growth of the wildtype and the HVO\_2753 deletion mutant.

**Figure S4.** Native purification of HVO\_2753 after homologous production in *H. volcanii*.

**Figure S5.** Top down and bottomup MS analysis of HVO\_2753.

**Figure S6.** Purification of HVO\_2753 after heterologous production in *E. coli*.

**Figure S7.** (A) 1D  $^1\text{H}$  spectra of HVO\_2753 protein expressed in *E. coli* and *H. volcanii* recorded at 298 K. (B) 2D  $^1\text{H}^{15}\text{N}$  HSQC spectra of HVO\_2753 protein expressed in *E. coli* and *H. volcanii* recorded at 298 K.

**Figure S8.** 2D  $^1\text{H}^{15}\text{N}$  Best-TROSY spectrum of HVO\_2753 protein recorded with 0.2 mM and 1 M NaCl.

**Figure S9.** Amide proton temperature-dependency of HVO\_2753 protein.

**Figure S10.** Temperature dependent population ratio between major and minor conformations.

**Figure S11.** ZZ-exchange NMR experiments.

**Figure S12.** Zinc ion quantification with a fluorimetric assay.

**Figure S13.** Swarm assay of single amino acid point mutants of HVO\_2753.

**Figure S14.** 2D  $^1\text{H}^{15}\text{N}$  BEST-TROSY spectra of four cysteine to alanine point mutations.

**Figure S15.** Amide proton temperature-dependency of HVO\_2753 protein.

**Figure S16.** NMR solution structure of HVO\_2753 protein.

**Figure S17.** Mapping of the  $C_\alpha/C_\beta$  chemical shift pairs from eight cysteines of the HVO\_2753 on the schematic distribution plot of three thiol states.

**Figure S18.** Root mean square error (RMSE) structure validation of HVO\_2753 protein.

**Figure S19.** Mean square error (MSE) structure validation of HVO\_2753 protein.

**Figure S20.** CSP analysis of the HVO\_2753 minor conformation.

**Figure S21.** Analysis of protein dynamics measured by NMR relaxation parameters based on the transverse relaxation rates.

**Figure S22.** Representative long-distance NOEs of cysteines.

**Figure S23.** Analysis of NOEs of the cysteines.

**Table S1.** Amino acid composition of HVO\_2753 in comparison to that the *H. volc.* proteome.

**Table S2.**  $^{13}\text{C}$  chemical shift assignment of cysteine residues of the HVO\_2753 protein.

**Table S3.**  $^{13}\text{C}$  chemical shift assignment of proline residues of the HVO\_2753 protein.

**Table S4.**  $^3\text{J}(\text{H}^{\text{N}}, \text{H}^\alpha)$  coupling constants determined from a 3D HNHA NMR experiment.

**Table S5.** List of oligonucleotides used for the generation of the deletion mutant and for probes for Southern and Northern analyses.