FEBS PRESS
science publishing by scientists

WILEY

The FEBS Journal

www.febsjournal.org

Biological functions, genetic and biochemical characterization, and NMR structure determination of the small zinc finger protein HVO_2753 from *Haloferax volcanii*

Sebastian Zahn, Nina Kubatova, Dennis J. Pyper, Liam Cassidy, Krishna Saxena, Andreas Tholey, Harald Schwalbe and Jörg Soppa

DOI: 10.1111/febs.15559

Supplementary Material

Page Number

1. Supplementary Figures

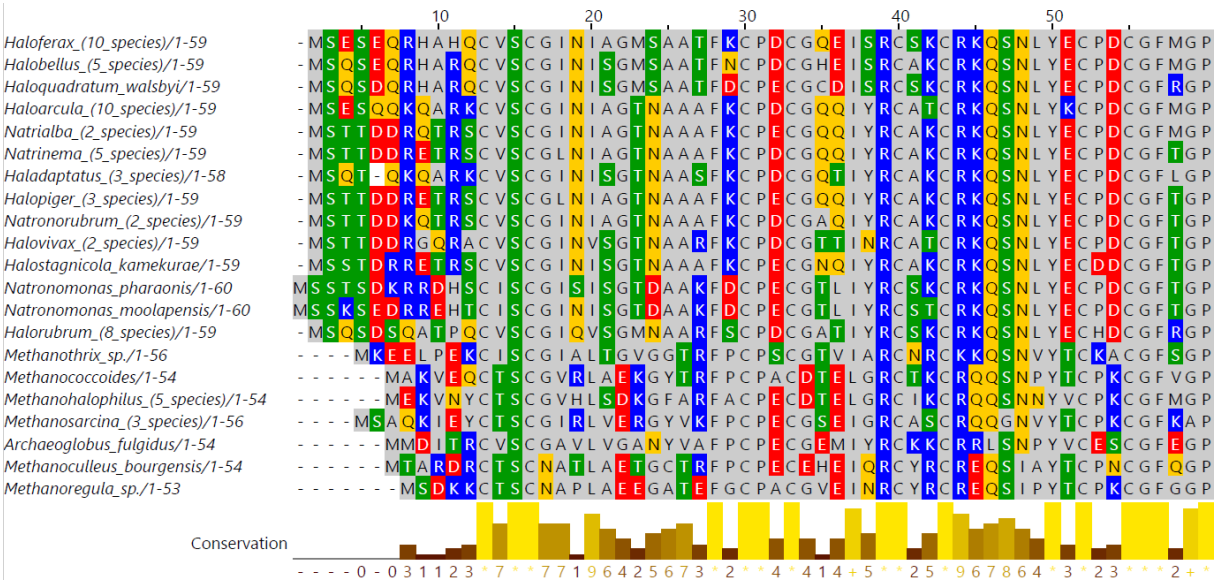

Figure S1. Multipe sequence alignment (MSA) of HVO_2753 from H. volcanii and 20 homologous proteins from selected halophilic and methanogenic (bottom seven sequences) genera and species. The degree of conservation is shown below the MSA. Color coding: red – acidic residues, blue – basic residues, yellow – amids, green – residues with hydroxyl groups.

The sequences were retrieved from the NCBI protein sequence database (https://www.ncbi.nlm.nih.gov/home/proteins/) and have the following accession numbers: *Haloferax* - WP_004042997.1, *Halobellus* - WP_081927459.1, *Haloquadratum* - WP_011572117.1, *Haloarcula* - WP_004516620.1, *Natrialba* - WP_006666601.1, *Natrinema* - WP_006184530.1, *Haladaptus* - WP_007978781.1, *Halopiger* - WP_013881574.1, *Natronorubrum* - WP_006090569.1, *Halovivax* - WP_006090569.1, *Halostagnicola* - WP_082146759.1, *Natrosomonas pharaonis* - WP_011323553.1, *Natrosomonas moolapensis* - WP_015409654.1, *Halorubrum* - WP_004047107.1, *Methanotrix* - RQW78931.1, *Methanococcoides* - WP_081955807.1, *Methanohalophilus* - WP_013038060.1, *Methanosarcina* - WP_048166236.1, *Archaeoblobus* - AAB90678.1, *Methanocelleus* - WP_082070476.1, *Methanoregula* - PKG32771.1.

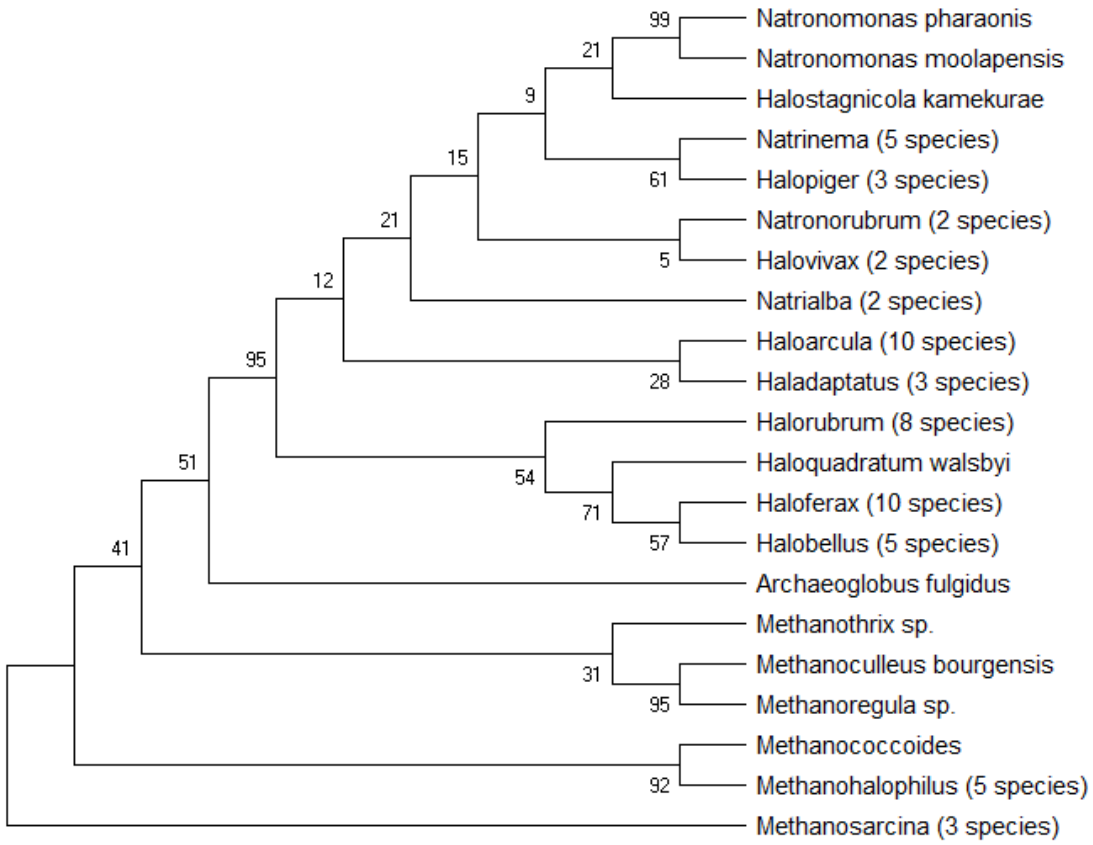

Figure S2A. Maximum parsimony phylogenetic tree of the multiple sequence alignment shown in Figure S1.

The tree was calculated with the program MEGA-X using the algorithm MUSCLE. 1000 bootstrap replications were calculated, and the results (% values) are shown at the respective nodes.

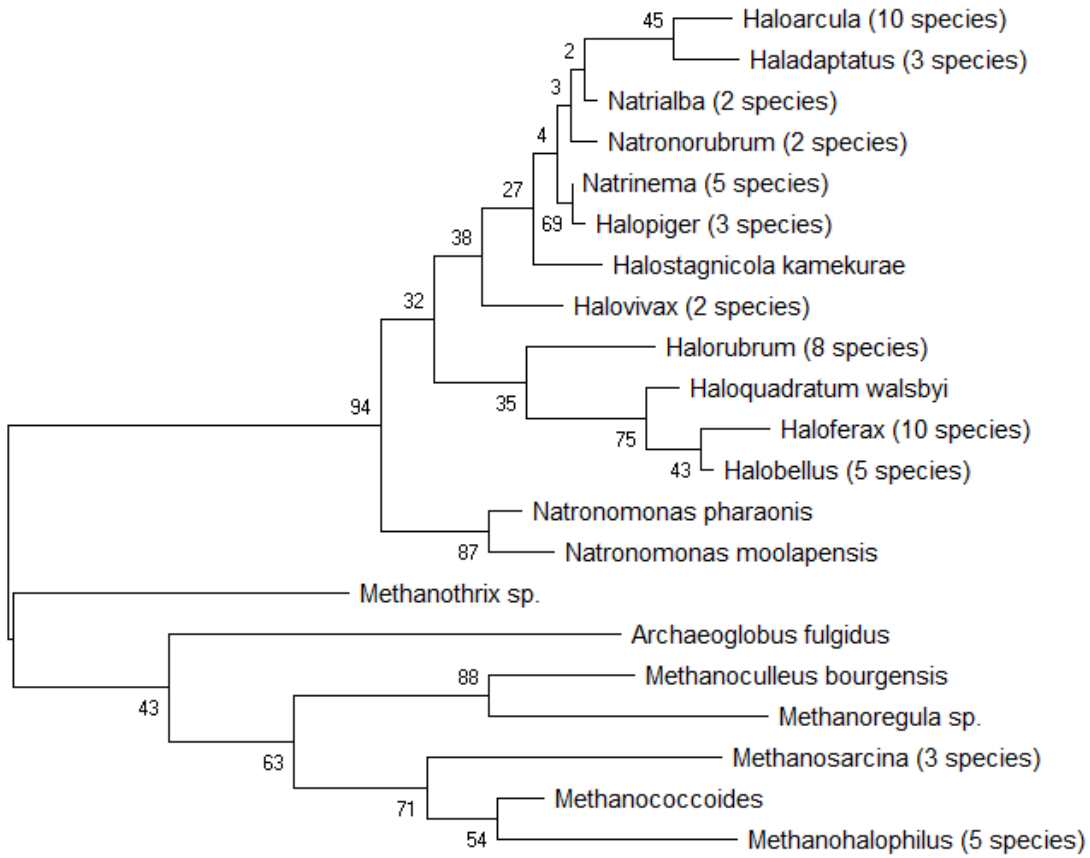

Figure S2B. Maximum likelyhood phylogenetic tree of the multiple sequence alignment shown in Figure S1.

The tree was calculated with the program MEGA-X using the algorithm MUSCLE. 1000 bootstrap replications were calculated, and the results (% values) are shown at the respective nodes.

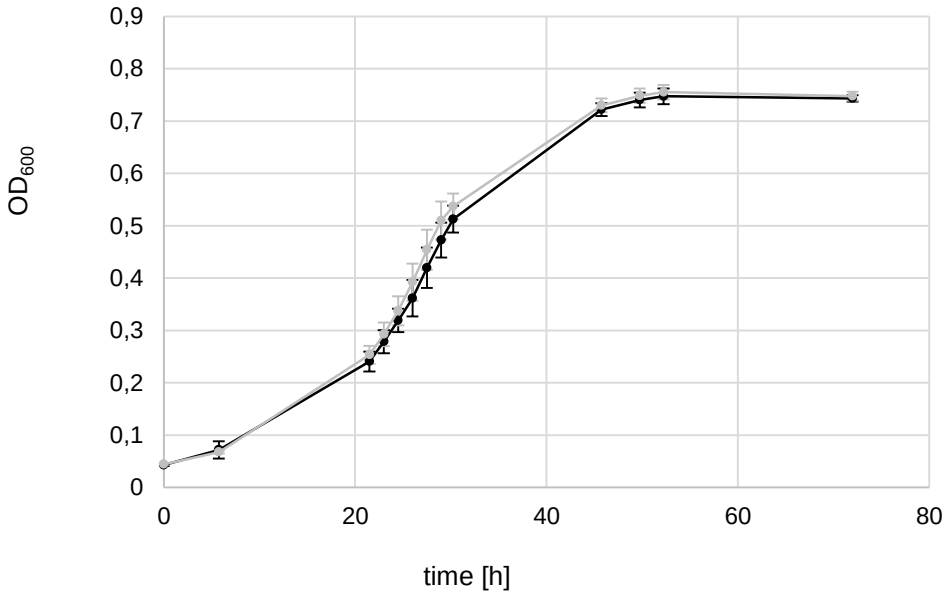

Figure S3. Growth of the wildtype (black) and the HVO_2753 deletion mutant (grey) in synthetic medium with glucose. Average values of three biological replicates and their standard deviations are shown.

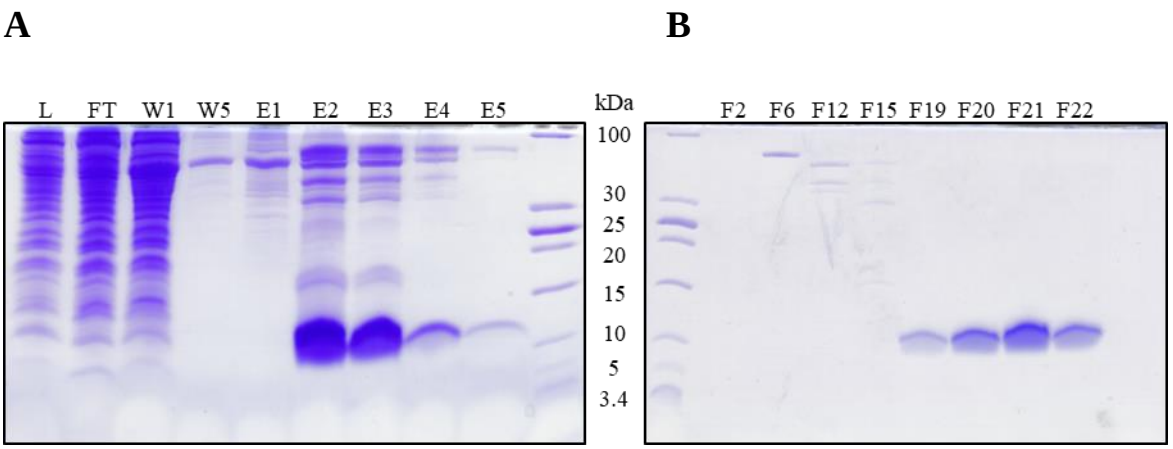

Figure S4. Native purification of HVO_2753 after homolous production in *H. volcanii*.
A. Purification gel after affinity isolation with nickel chelating sepharose. L – cytoplasmic lysate, FT – flow through, W – wash fractions 1 and 5, E – elution fractions 1 – 5, the molecular weight marker is in the rightmost lane.

B. Purification gel after size exclusion chromatography. The pooled elution fractions after the affinity isolation were applied to a superose 6 size exclusion column. Selected fractions (see numbers) were analyzed by SDS-page. Fractions 20 and 21 were pooled and the protein was used for further analyses (see text). The molecular weight marker is in the leftmost lane.

After establishment and optimization, the purification scheme was performed regularly for more than two years (> ten times). One representative example is shown.

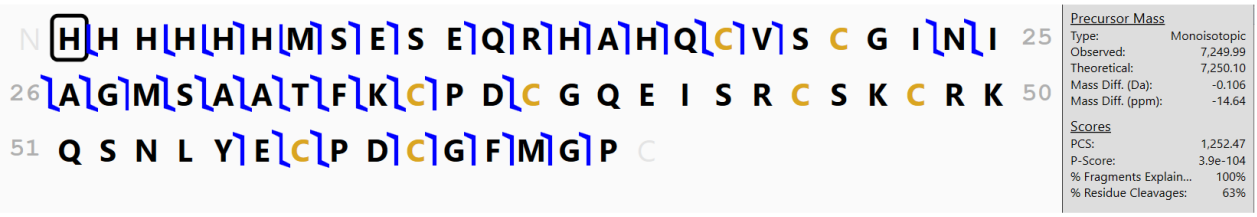

Figure S5A. Top-down MS analysis of HVO_2753 was performed utilizing a series of normalized collision energies (NCE) for both HCD (NCE 17, 25, 30) and CID (NCE 20, 25, 30) activation types. This resulted in the identification of 63% of possible inter-residue cleavage products (20 ppm tolerance of fragment ions).

```
>|HVO_2753_HIS_tagged protein
    HHHHHHMSESEQRHAHQCVSCGINIAGMSAATFKCPDCGQEISRCSKCRKQSNLYECPDCGFMGP
[-].HHHHHHMSESEQRHAHQCVSCGINIAGMSAATFK.[C]
[-].HHHHHHMSESEQR.[H]
            Q].RHAHQCVSCGINIAGMSAATFK.[C]
            [R].HAHQCVSCGINIAGMSAATFKCPDCGQEISR.[C]
            [R].HAHQCVSCGINIAGMSAATFK.[C]
            [R].HAHQCVSCGINIAGMSAA.[T]
            [R].HAHQCVSCGINIA.[G]
            [R].HAHQCVSCGIN.[I]
             [H].AHQCVSCGINIAGMSAATFK.[C]
                    [C].GINIAGMSAATFK.[C]
                       [N].IAGMSAATFK.[C]
                           [M].SAATFK.[C]
                               [T].FKCPDCGQEISR.[C]
                                [F].KCPDCGQEISR.[C]
                                 [K].CPDCGQEISRCSKCRK.[Q]
                                 [K].CPDCGQEISR.[C]
                                  [C].PDCGQEISR.[C]
                                   [P].DCGQEISR.[C]
                                                [R].KQSNLYECPDCGFMGP.[-]
                                                [R].KQSNLYECPD.[C]
                                                 [K].QSNLYECPDCGFMGP.[-]
                                                    [N].LYECPDCGFMGP.[-]
```

Figure S5B. Bottom-up MS analysis of HVO_2753 allowed for complete sequence coverage of the canonical sequence. The canonical protein sequence is given on the top line with the twenty two unique peptide sequences identified displayed below.

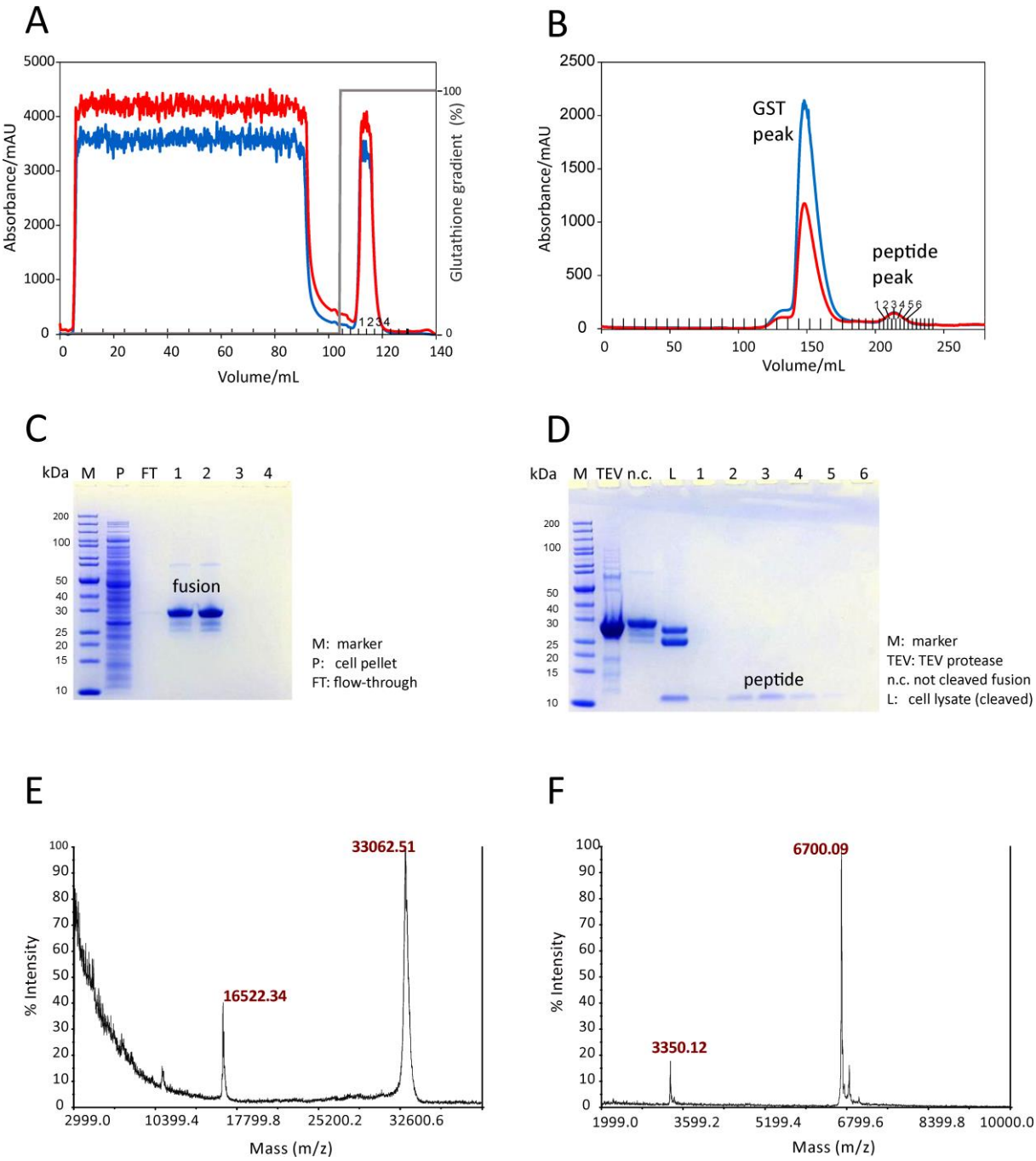

Figure S6. Purification strategy of HVO_2753 protein. A – GST affinity chromatogram of GST-HVO_2753 fusion protein. The absorption at 280 nm and 260 nm are shown in blue and red, respectively (left-handed y-axis) and are plotted as functions of the volume. The applied 100% glutathione gradient (right-handed y-axis) is shown in gray. The collected fractions are shown on the x-axes. **B** – Size-exclusion chromatogram of GST-HVO_2753 fusion protein after cleavage with TEV protease. **C, D** – The corresponding SDS-PAGE analyses of relevant fractions from GST affinity chromatogram (**C**) and size-exclusion chromatogram (**D**). Marker (PageRuler^{TM} Unstained Protein Ladder, ThermoFisher Scientific) (M) was used. Visualization was accomplished by coomassie staining. The protein bands are indicated in the picture. **E, F** – MALDI mass spectrometry spectra of GST-HVO_2753 fusion protein (**E**) and cleaved isolated HVO_2753 protein (**F**) are shown. Expected molecular weight is 34001 Da for ^{15}N labeled fusion protein and 6702 Da for ^{15}N labeled isolated HVO_2753 small protein. After establishment and optimization, the purification scheme was performed regularly for more than two years (>ten times). One representative example is shown.

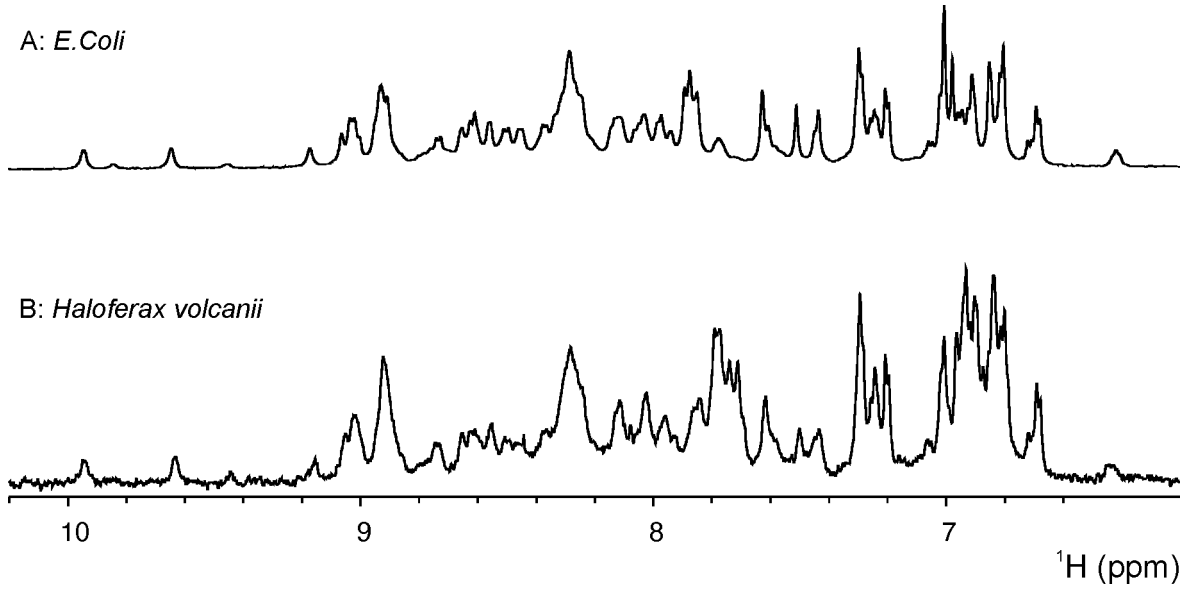

Figure S7 A. 1D ^{1}H spectra of HVO_2753 protein expressed in *E. coli* and *H. volcanii* recorded at 298 K. The samples contain 25 mM Bis-Tris buffer pH 7, 200 mM NaCl, 200 µM $ZnCl_2$, 5% D_2O, 0.5 mM DSS. **A** – HVO_2753 protein expressed in E. coli using uniformly labeled M9 medium. 1.2 mM sample; 1D ^{1}H NMR experiment with ^{13}C, ^{15}N decoupling; 16 scans; 298K, 600 MHz. **B** – HVO_2753 protein isolated from Haloferax volcanii. 0.34 mM sample; 1D ^{1}H NMR experiment; 64 scans; 298 K, 600 MHz.

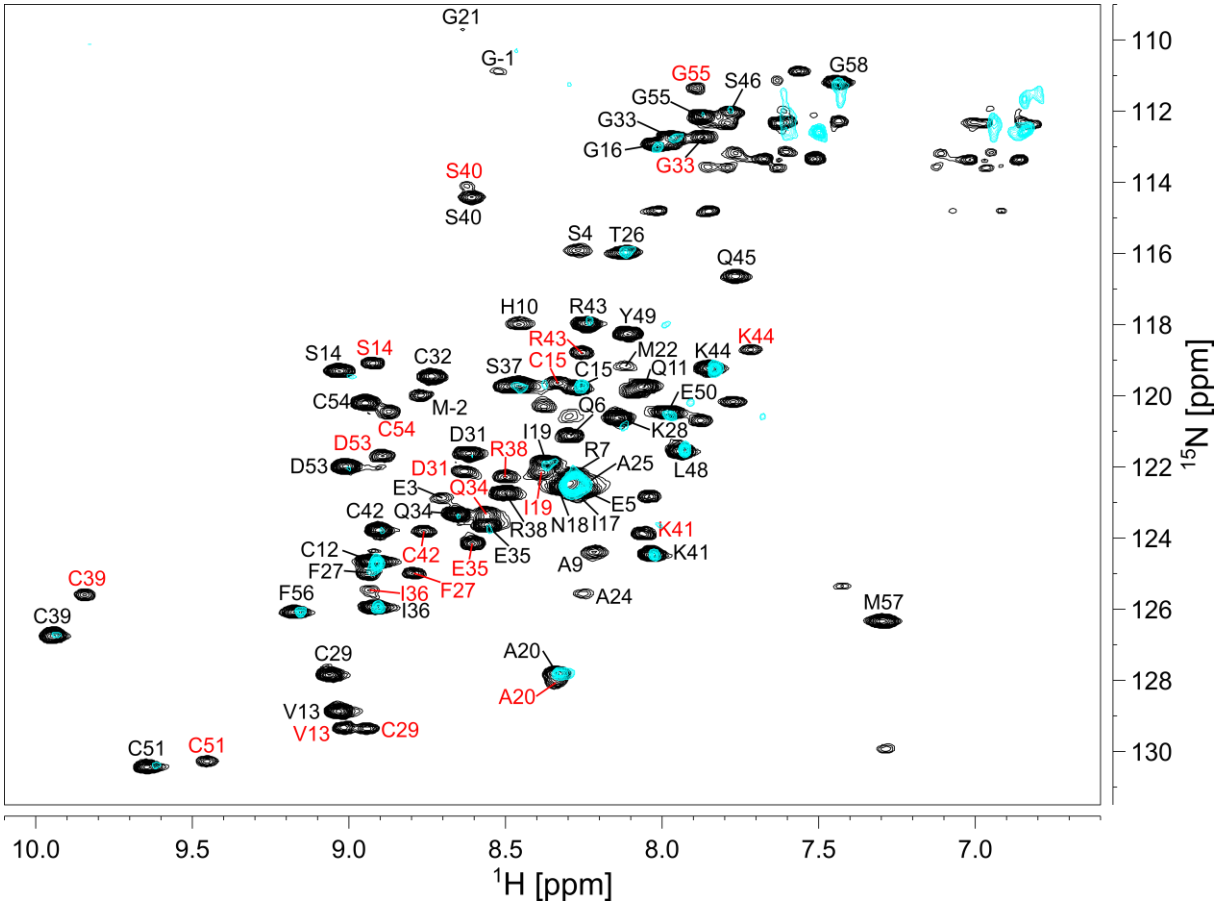

Figure S7 B. 2D $^{1}H^{15}N$ HSQC spectra of HVO_2753 protein expressed in *E. coli* and *H. volcanii* recorded at 298 K, 600 MHz. The samples contain 25 mM Bis-Tris buffer pH 7, 200 mM NaCl, 5% D_2O, 0.5 mM DSS. Black – 2D $^{1}H^{15}N$ Best-TROSY spectrum of HVO_2753 protein expressed in E. coli using uniformly labeled M9 medium. 1.2 mM sample; 32 scans. Cyan – Natural abundance sofast HMQC spectrum of HVO_2753 isolated from Haloferax volcanii. 0.34 mM sample; 16k scans.

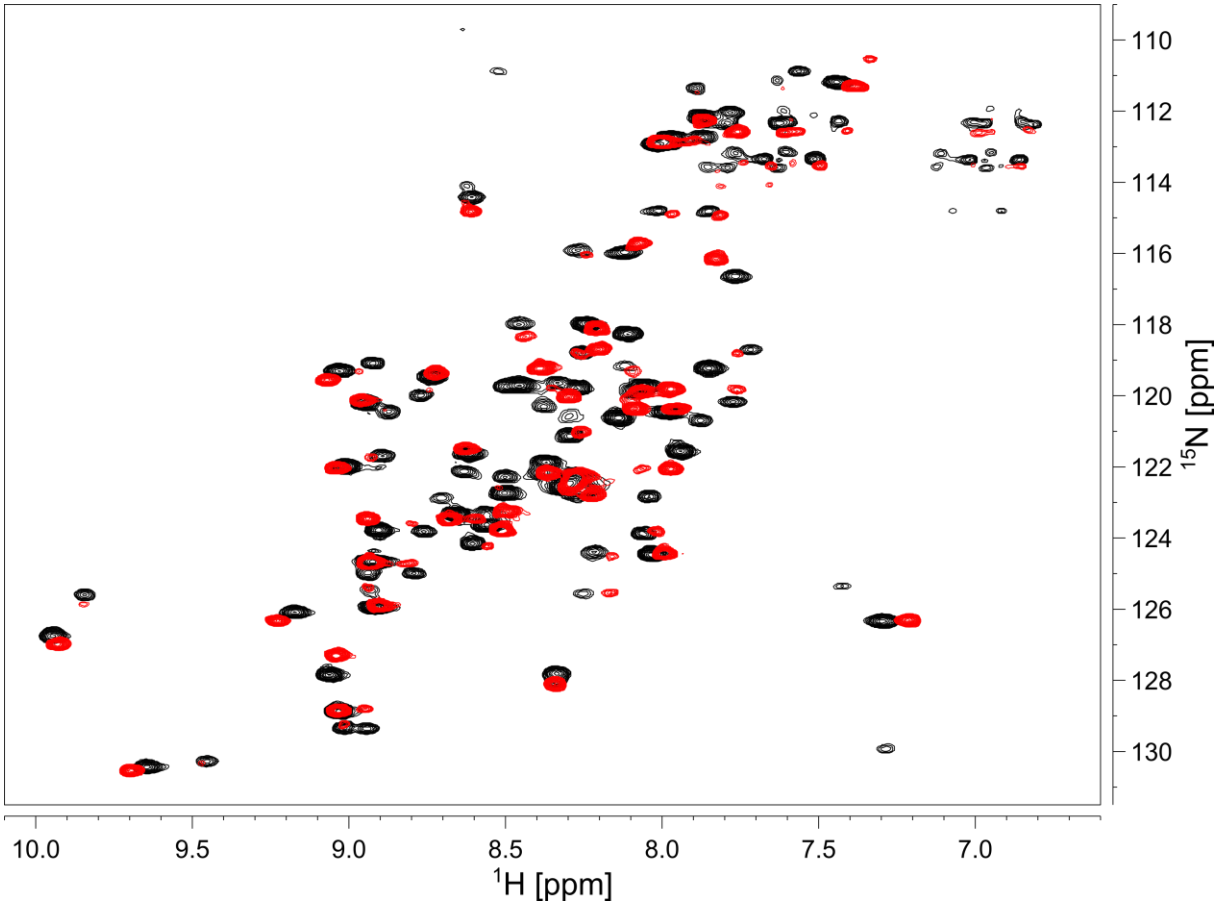

Figure S8. 2D $^{1}H^{15}N$ Best-TROSY spectrum of HVO_2753 protein recorded with 0.2 mM (black) and 1 M NaCl (red). The sample contains 25 mM Bis-Tris buffer pH 7, 5% D_2O, 0.5 mM DSS; recorded at 600 MHz, 298K.

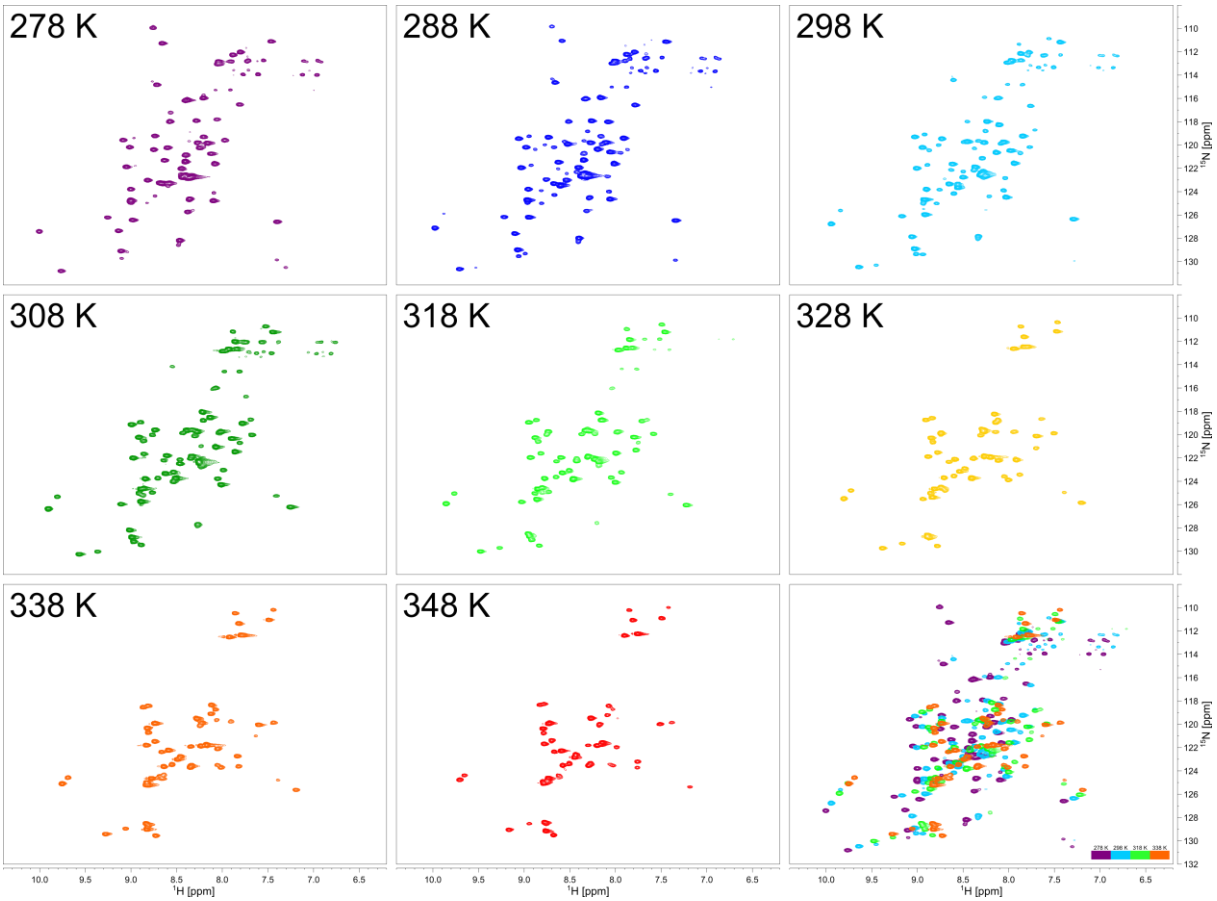

Figure S9. Amide proton temperature-dependency of HVO_2753 protein. 2D $^{1}H^{15}N$ BEST-TROSY spectra recorded at 600 MHz, at different temperatures (color code is shown on the legend) using 25 mM Bis-Tris buffer pH 7, 200 mM NaCl, 5% D_2O, 0.5 mM DSS.

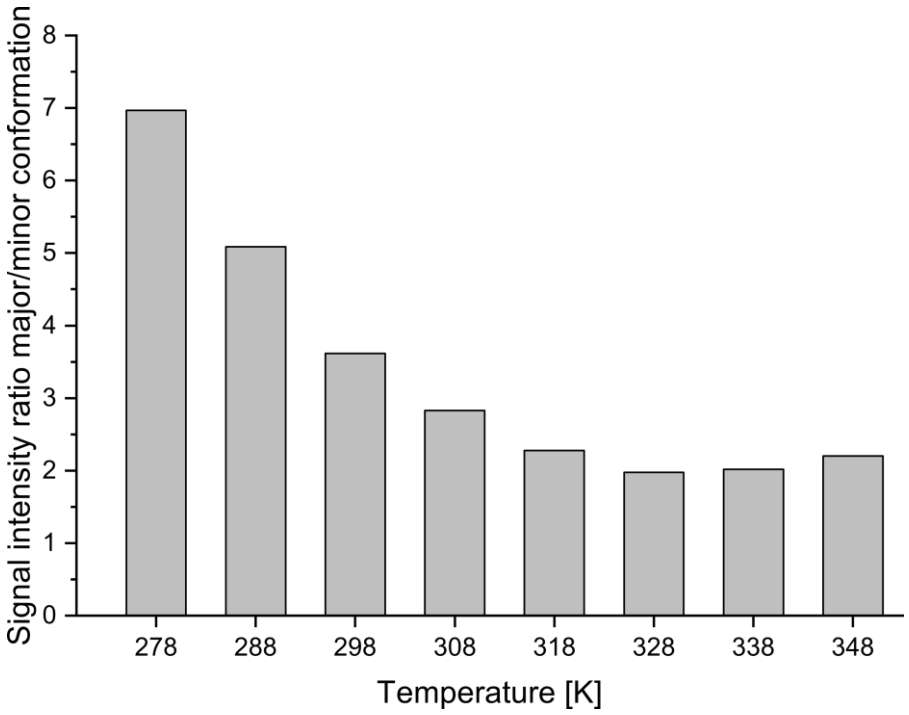

Figure S10. Temperature dependent population ratio between major and minor conformations.

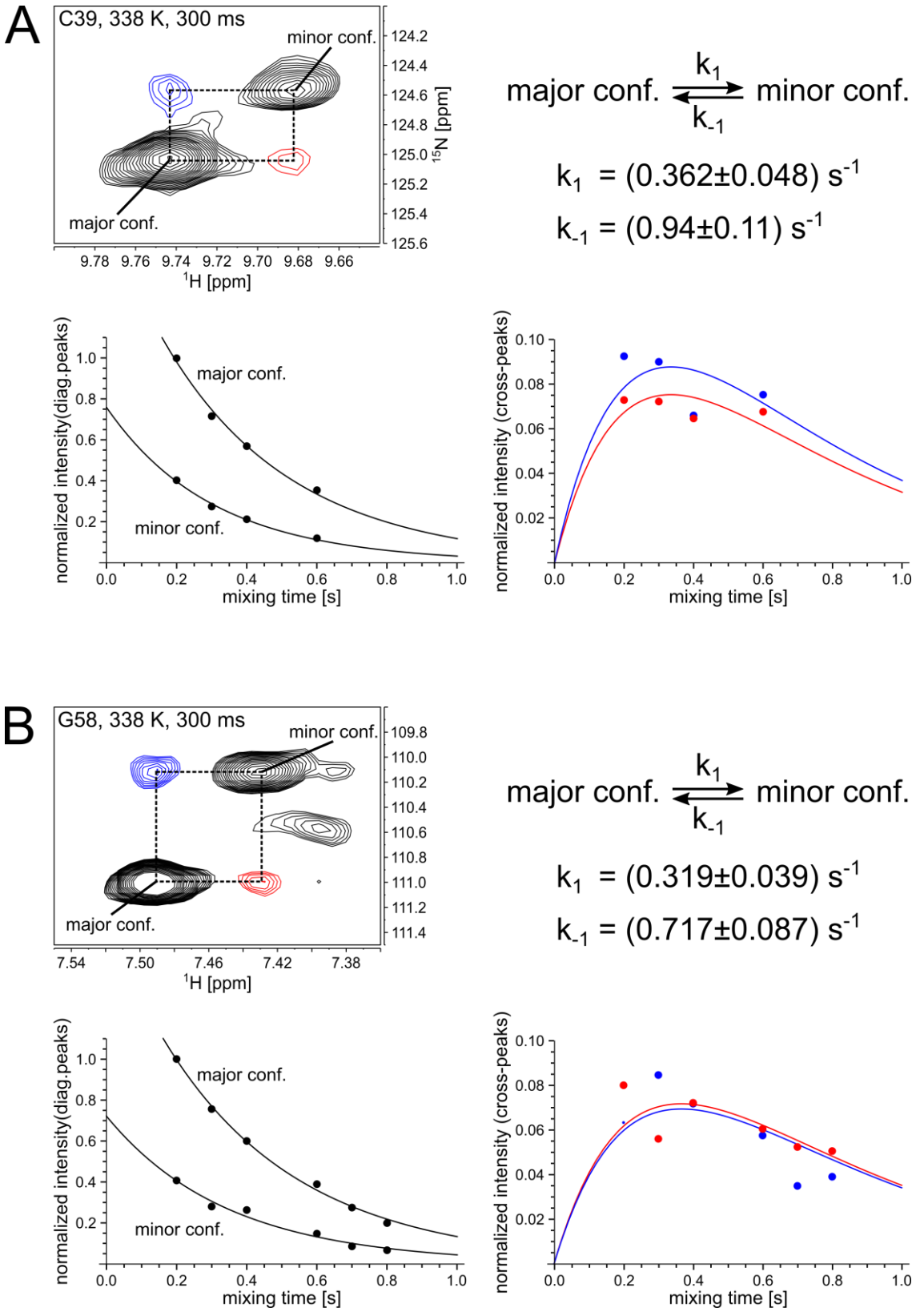

Figure S11. ZZ-exchange NMR experiments. Expansion of the ZZ-exchange spectra of the residue C39 (**A**) and G58 (**B**) at mixing times 300 ms at 338 K. On the left is the reduction of the diagonal peaks and on the right is the buildup of the exchange cross-peaks between the major and minor conformations.

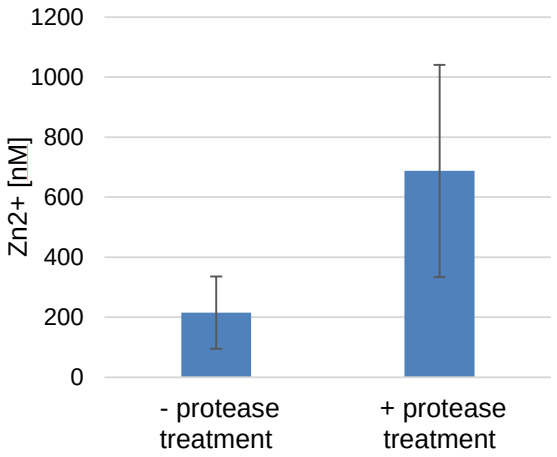

Figure S12. Zinc ion quantification with a fluorimetic assay. HVO_2753 was produced homologously in H. volcanii and purified as shown in Figure S4. It was dialyzed against low salt buffer (left) and additionally treated with proteinase K to destroy the protein (right). 1 µM protein was used for zinc ion quantification with the zinc specific fluorophore ZnAF-2f. Average results of seven biological replicates and their standard deviations are shown.

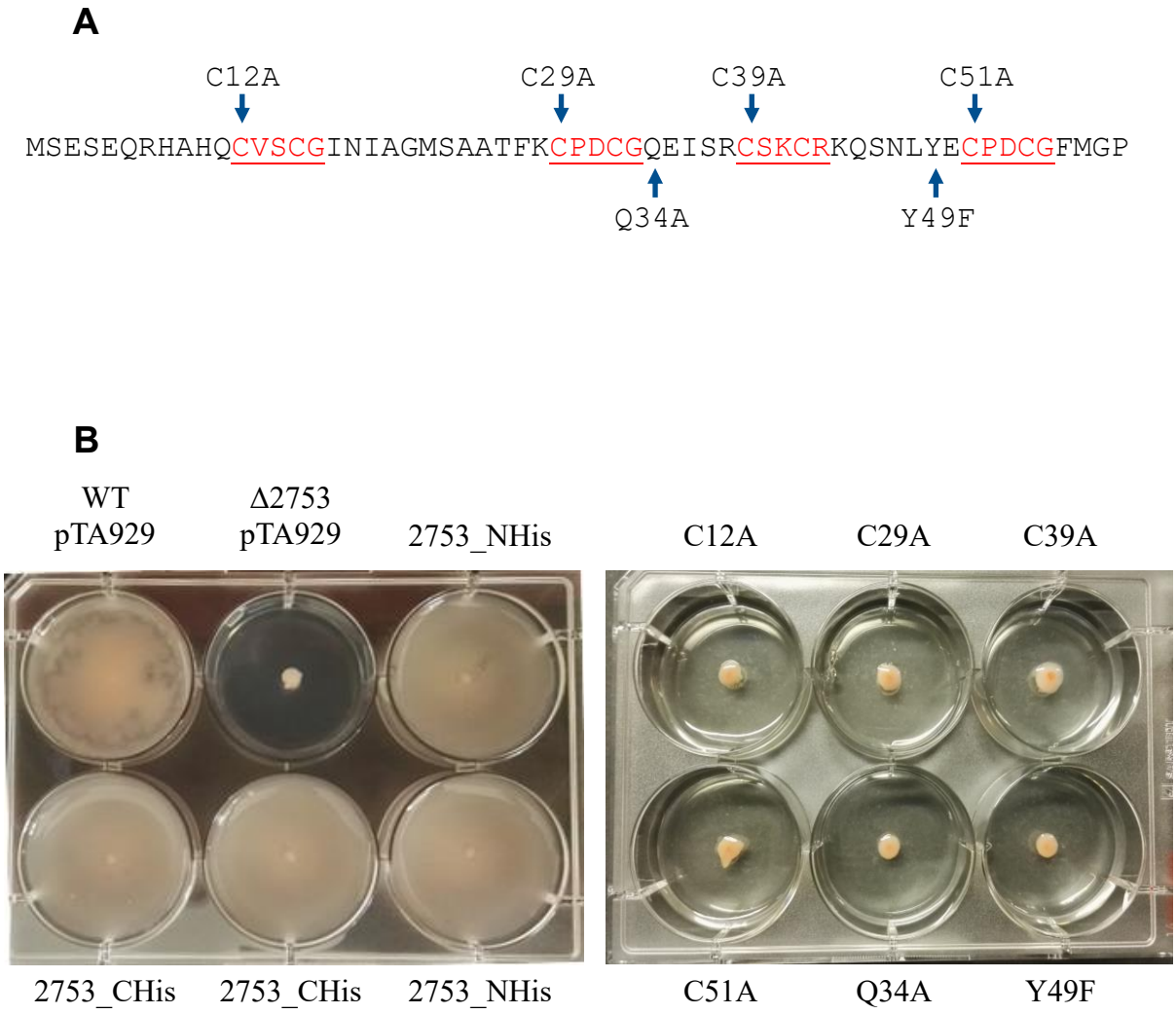

Figure S13. Swarm assay of single amino acid point mutants of HVO_2753.

A. Sequence of HVO_2753. The four C(P)XCG motifs are highlighted in red. The six single amino acid point mutations are indicated above and below the sequence.

B. Swarm assay. For comparison, swarming of the wildtype, the HVO_2753 deletion mutants, and strains complemented with the wildtype sequence are shown to the left. Swarming of six strains complemented with point mutated versions of HVO_2753 are shown to the right. The identity of the strains is indicated above and below the swarm plates. Typical plates of three biological replicates are shown.

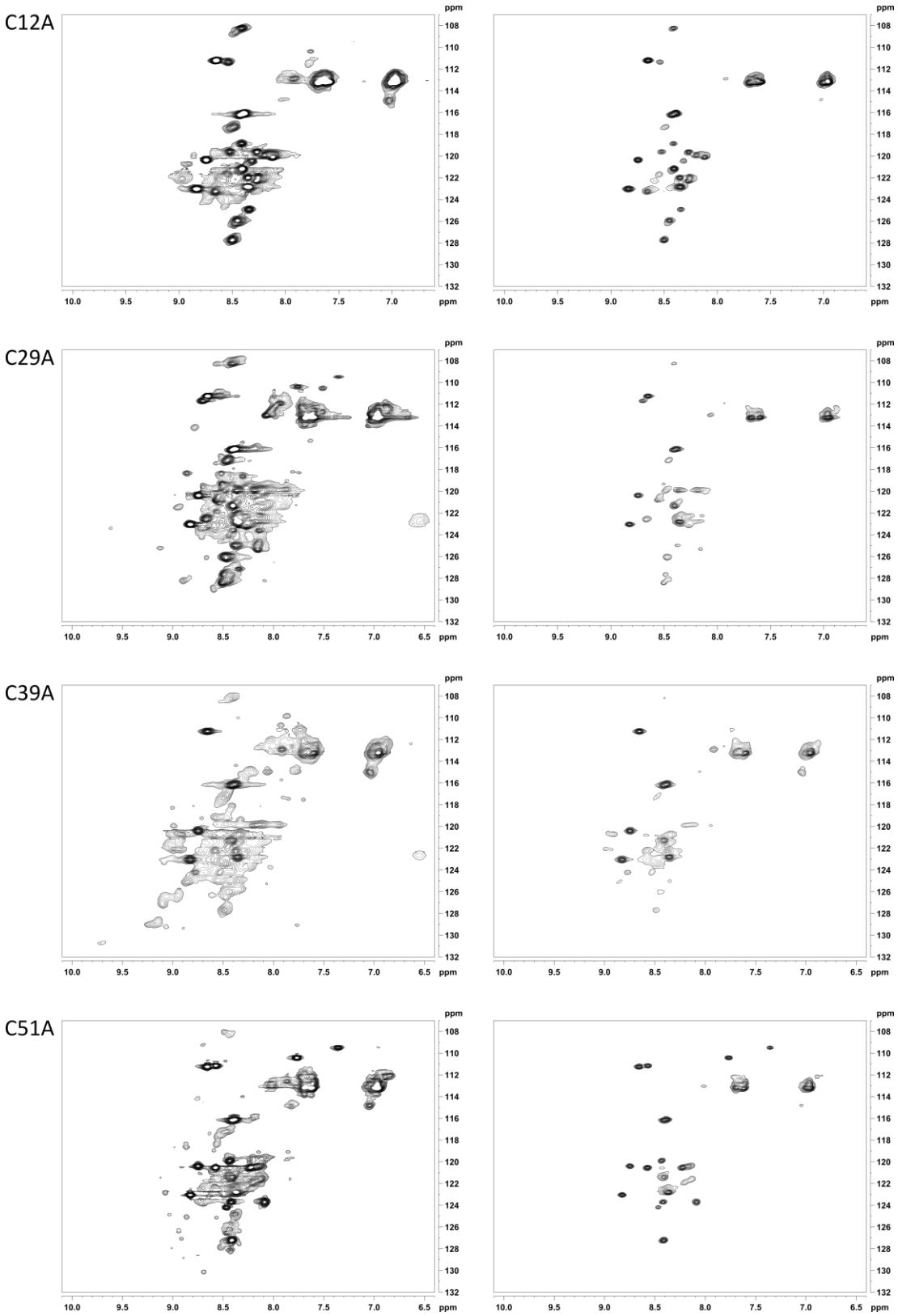

Figure S14. Single point mutations. 2D $^{1}H^{15}N$ BEST-TROSY spectra of four single point cysteine to alanine mutations recorded at 600 MHz, 298 K, using 25 mM Bis-Tris buffer pH 7, 200 mM NaCl, 5% D_2O, 0.5 mM DSS. In order to show the molten globule state formation the counter level of the spectra are increased on the left side of the picture.

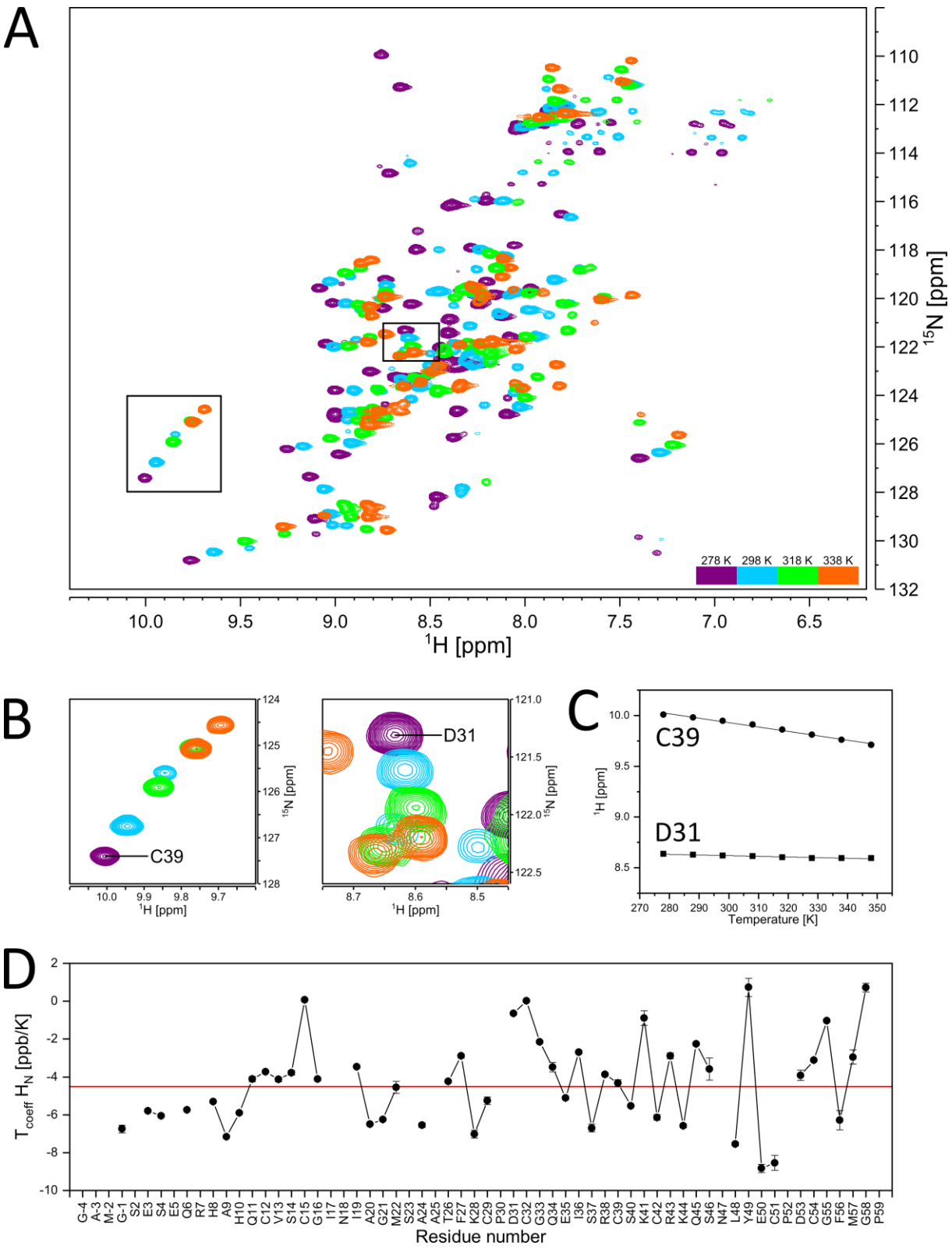

Figure S15. Amide proton temperature-dependency of HVO_2753 protein. **A** – 2D $^{1}H^{15}N$ Best-TROSY spectra recorded at 600 MHz, at different temperatures (color code is shown on the legend) using 25 mM Bis-Tris buffer pH 7, 200 mM NaCl. **B** – Zoom in the region of C39 and D31 residues, representing different hydrogen bond formation. **C** – Plot of the amide proton chemical shift (in ppm) as a function of temperature (in K).

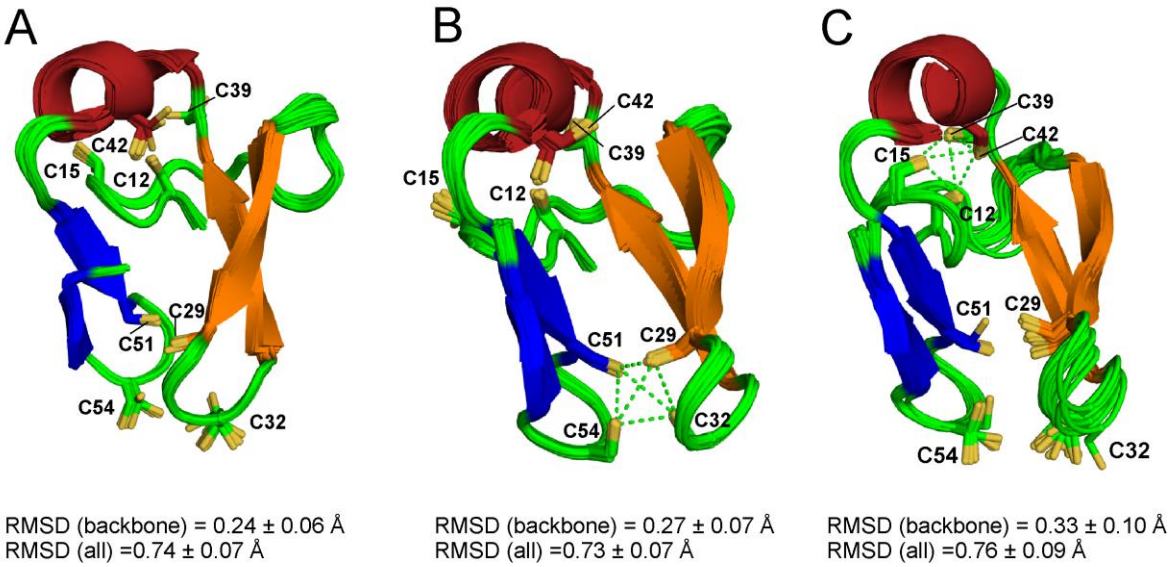

Figure S16. NMR solution structure of HVO_2753 protein. Ribbon representation of the ensemble of the best 20 structures of the rigid core of the protein. Color code: β-sheets are colored in blue and orange, α-helix is shown in red, sulfur atoms from corresponding cysteine residues are represented as yellow sticks and manually defined binding pocket is shown as greed dashed lines. **A** – No binding pocket is defined. **B** – The binding pocket 2, which was manually defined by the cyana structure calculation, is formed by the binding motifs 2 and 4 and includes the residues C29, C32, C51, and C54. **C** – The binding pocket 1, which was manually defined by the cyana structure calculation, is formed by the binding motifs 1 and 3 and includes residues C12, C15, C39, C42 and C42. The figure was generated by PyMOL.

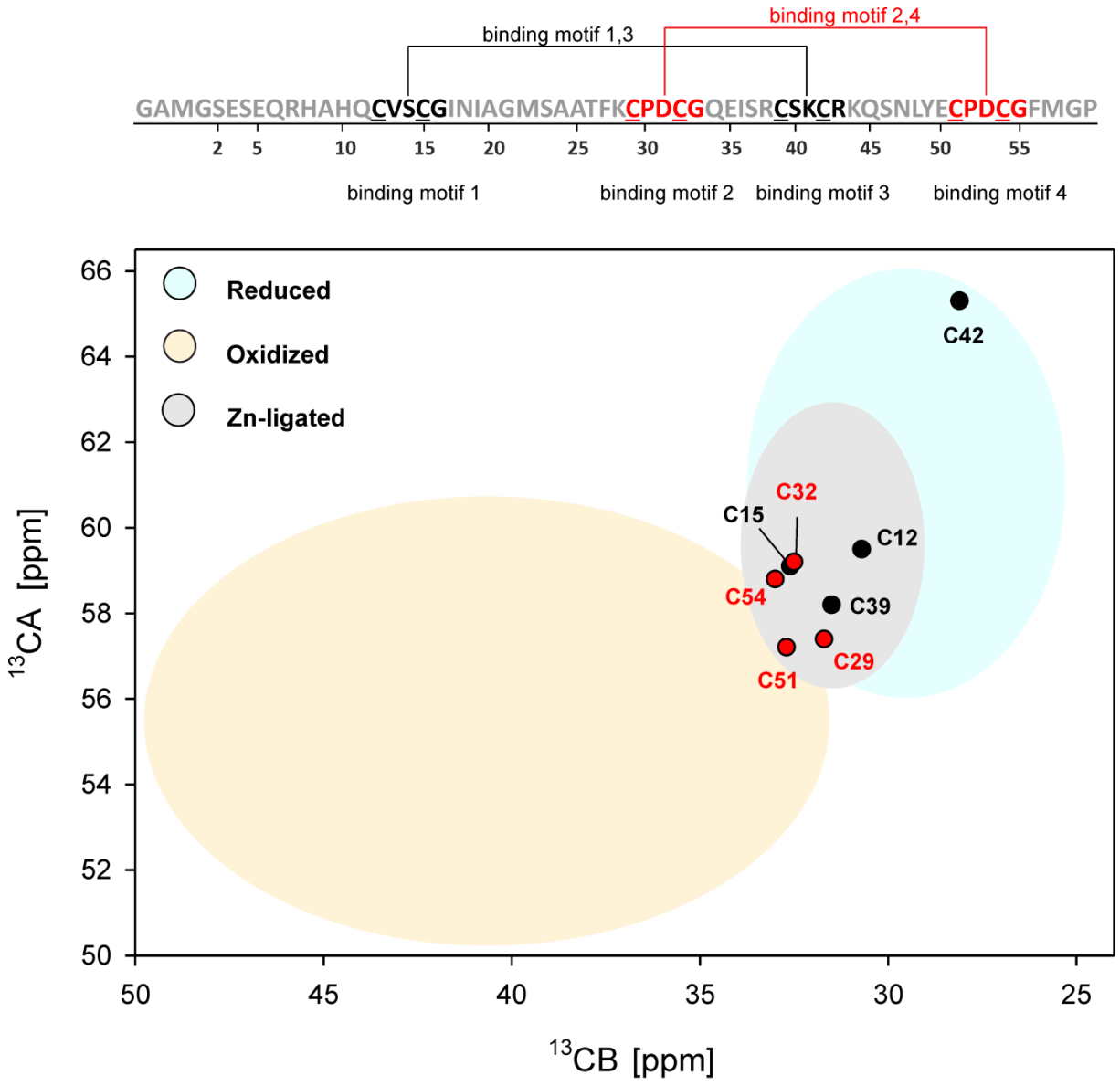

Figure S17. Mapping of the C_α/C_β chemical shift pairs from eight cysteines of the HVO_2753 on the schematic distribution plot of three thiol states. The distribution plot is modified after [Kornhaber et al. 2006]. Region for oxidized cysteines is highlighted in yellow, reduced non-metal-ligated cysteines are shown in blue and Zn-ligated cysteines are pictured in gray. The black and red dots correspond to the cysteines forming ZBP1 and ZBP2 respectively.

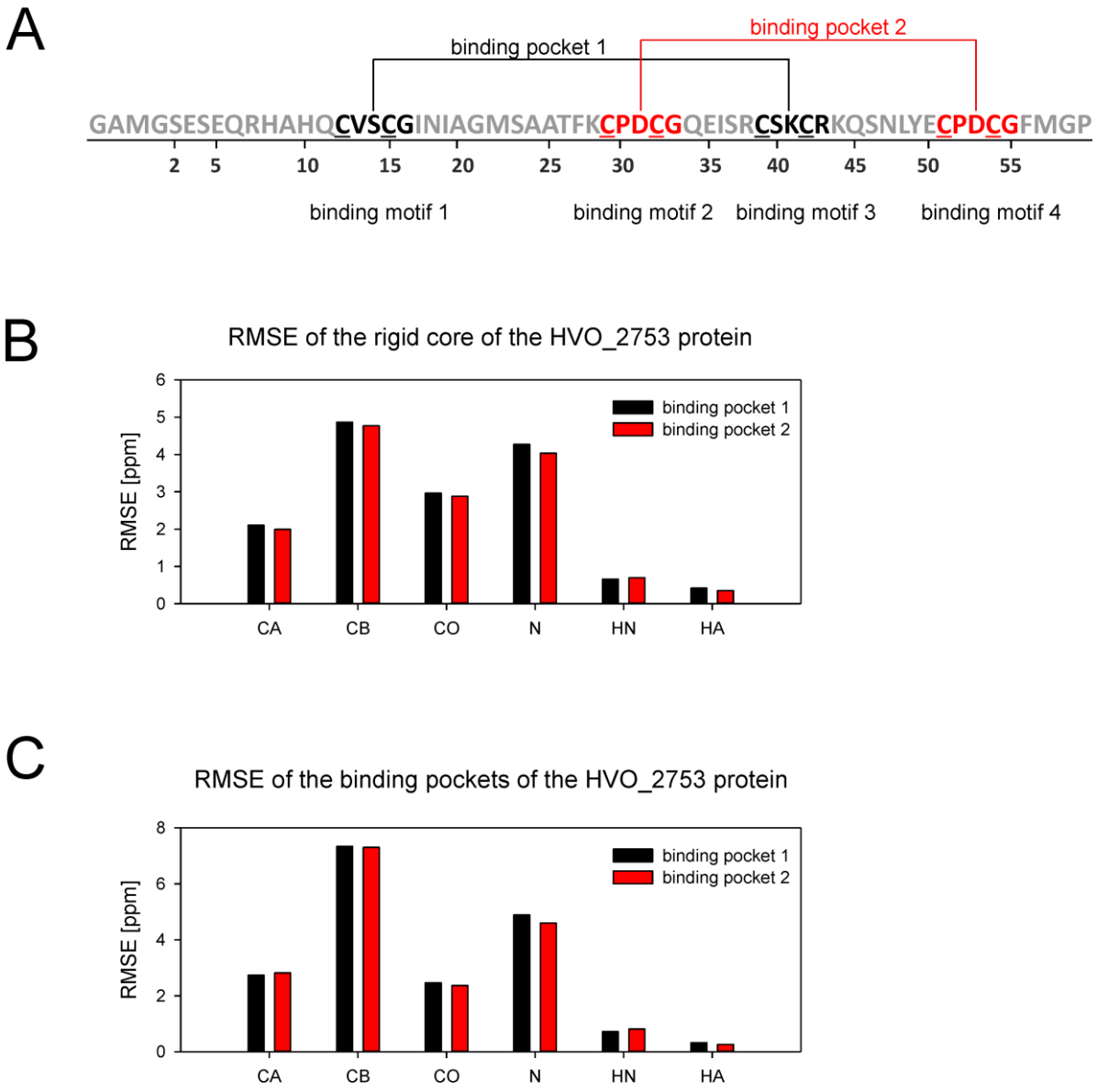

Figure S18. Root mean square error (RMSE) structure validation of HVO_2753 protein. **A** – Schematic representation of the binding motifs based on the cyana structure calculation. Binding pockets consisting of binding motifs 1, 3 and 2, 4 are shown in black and red respectively. **B** – RMSE values for the rigid core, which excludes the first N-terminal flexible 13 amino acids. **C** – RMSE values for the binding pockets, which include the four binding motifs (in total 20 residues). RMSE values were calculated according to formula 1.

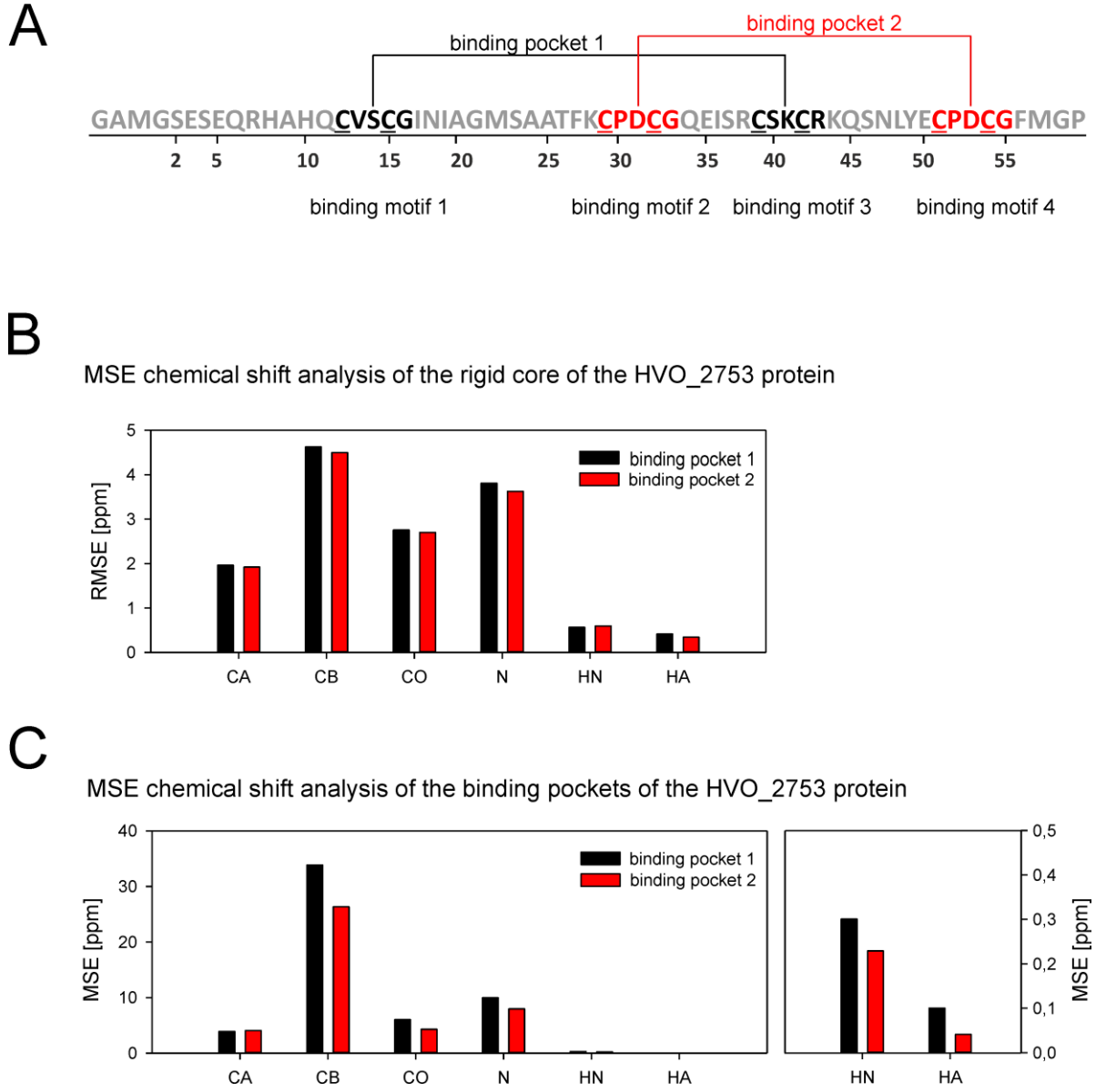

Figure S19. Mean square error (MSE) structure validation of HVO_2753 protein. A – Schematic representation of the binding motifs based on the cyana structure calculation. Binding pockets consisting of binding motifs 1, 3 and 2, 4 are shown in black and red respectively. **B** – MSE values for the rigid core, which excludes the first N-terminal flexible 13 amino acids. **C** – MSE values for the binding pockets, which include the four binding motifs (in total 20 residues). The expanded region of HN and HA atoms is shown on the right. MSE values were calculated according to formula 2.

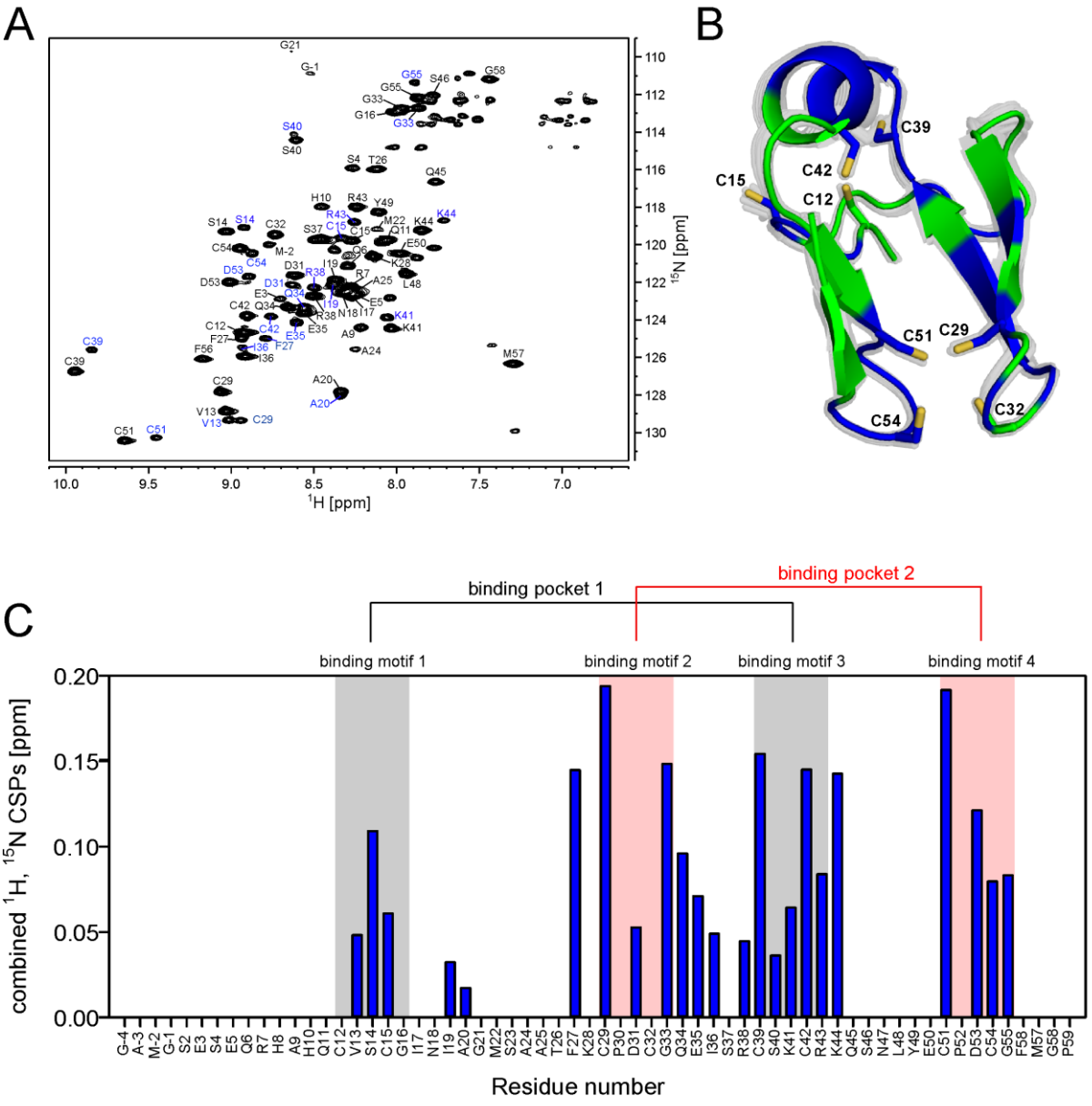

Figure S20. CSP analysis of the HVO_2753 minor conformation. A – 2D $^{1}H^{15}N$ HSQC spectrum recorded at 298 K, 800 MHz. Assignment of the major conformation is highlighted in black and the minor conformation is colored in blue. The sample contains 25 mM Bis-Tris buffer pH 7, 200 mM NaCl, 5% D_2O, 0.5 mM DSS. **B** – Mapping of CSPs of the minor conformation on the HVO_2753 solution structure of the major conformation. Residues, which show amide signal from the minor conformation, are colored in blue. **C** – Combined ^{1}H and ^{15}N CSPs of the minor conformation of the backbone amides as a function of the protein residue number. Residues involved in the binding pocket 1 and 2 are highlighted in grey and pink boxes respectively.

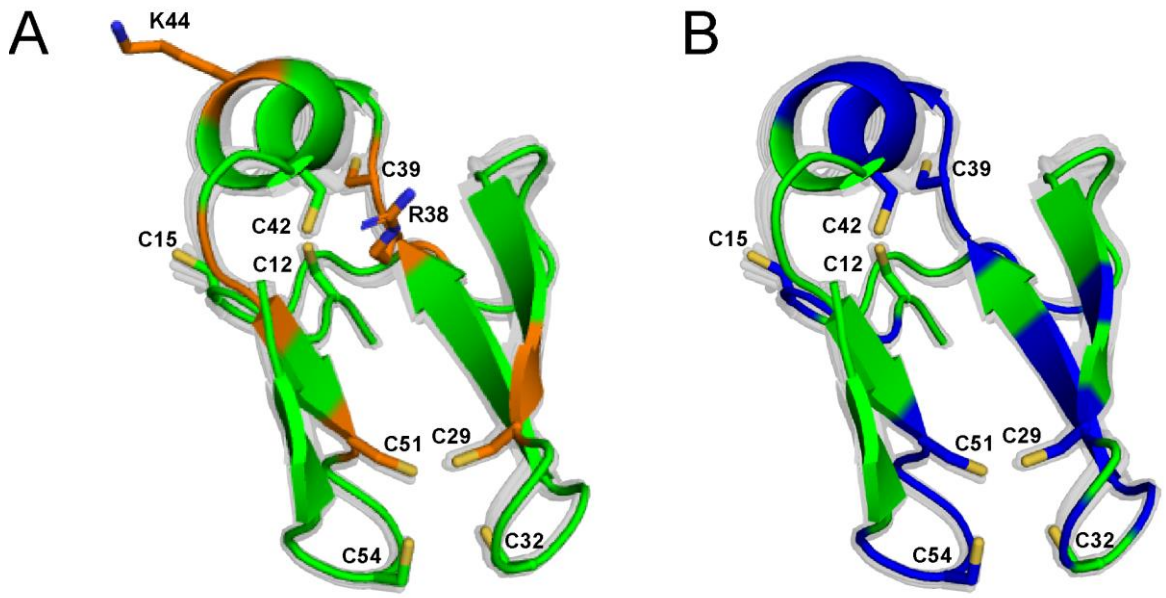

Figure S21. Analysis of protein dynamics measured by NMR relaxation parameters based on the transverse relaxation rates. A – Residues that exhibit transversal relaxation rate R_2 higher than 13 s^{-1} are highlighted in orange (average of the rigid core of the protein is 9.5 ± 0.2 s^{-1}). **B** – Residues that contribute to the formation of the minor conformation are highlighted in blue.

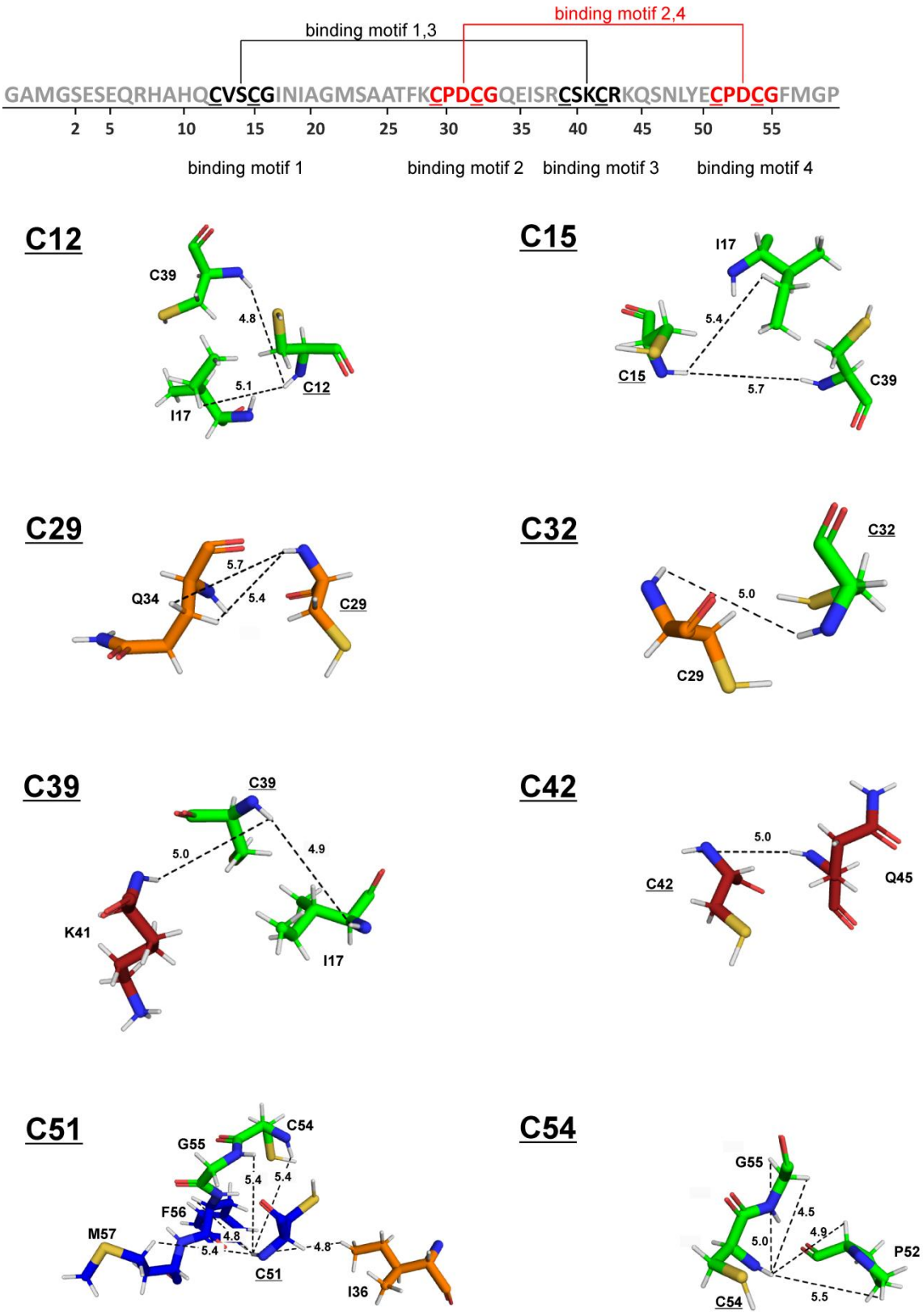

Figure S22. Representative long-distance NOEs of cysteines. Intra-molecular long range distances (≤ 4.5Å) that correspond to the cross peaks in 3D $^{1}H^{1}H^{15}N$ NOESY-HSQC experiment are highlighted as black dashes. Cysteines that form ZBP1 and ZBP2 are colored in black and red respectively.

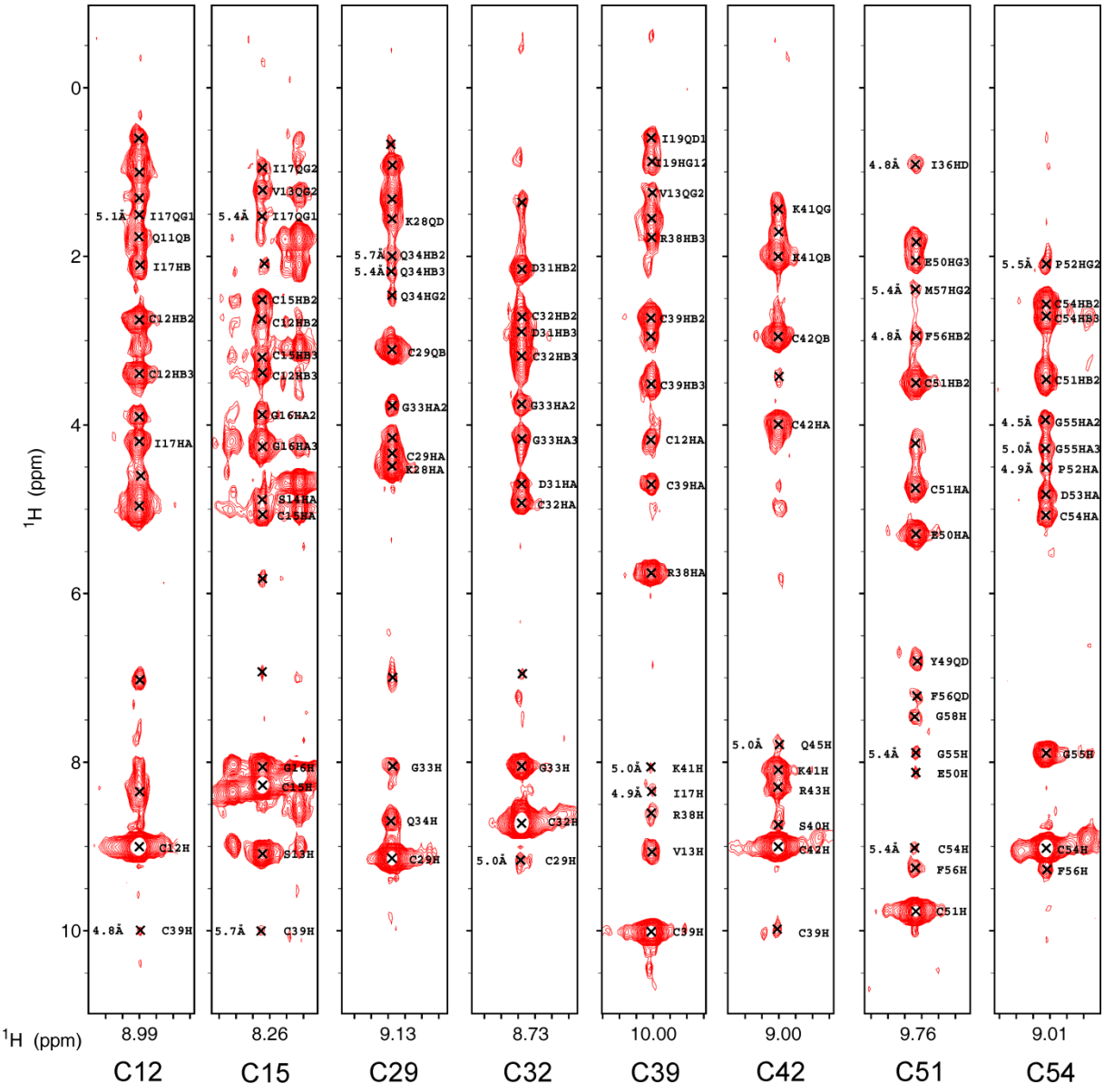

Figure S23. Analysis of NOEs of the cysteines. Strip plots of cysteine residues from a 3D ^{1}H^{1}H^{15}N NOESY-HSQC spectrum recorded at 278 K, 950 MHz with mixing time of 120 ms. Cross peaks that were assigned and used for the structure calculation are highlighted with cross.

2. Supplementary Tables

Table S1. Amino acid composition of HVO_2753 in comparison to that the H. volc. proteome. The fractions of HVO_2753 that are more than twofold higher or lower than in the proteome are highlighted in red and blue, respectively.

Amino acid	H. volcanii Proteome	HVO_2753
	Fraction [%]	Fraction [%]
A	11.0	**5**
C	0.7	**14**
D	8.4	**3**
E	8.0	7
F	3.5	3
G	8.5	8
H	2.0	3
I	3.8	5
K	2.0	5
L	9.1	**2**
M	1.8	**5**
N	2.4	3
P	4.6	5
Q	2.4	**7**
R	6.7	5
S	5.9	**12**
T	6.2	**2**
V	9.2	**2**
W	1.1	**0**
Y	2.7	2
No. aa	1.168.832	59

Table S2. ^{13}C chemical shift assignment of cysteine residues of the HVO_2753 protein.

	CA [ppm]	CB [ppm]
C12	59.5	30.7
C15	59.1	32.6
C29	57.4	31.7
C32	59.2	32.5
C39	58.2	31.5
C42	65.3	28.1
C51	57.2	32.7
C54	58.8	33.0

Table S3. ^{13}C chemical shift assignment of proline residues of the HVO_2753 protein.

	CB [ppm]	CG [ppm]	Δ(CB-CG) [ppm]
P30	32.6	27.4	5.2
P52	32.4	26.7	5.7
P59	34.9	25.1	9.8

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

Residue	Coupling constant [Hz]
M-2	6,95
E3	6,27
E5	6,39
Q6	6,99
H8	7,81
A9	5,61
H10	8,14
Q11	8,74
V13	3,57
S14	9,46
C15	9,58
N18	4,47
I19	8,58
A20	4,88
G21	5,15
M22	7,55
S23	6,79
A24	6,27
A25	6,03
T26	7,52
C29	2,45

Residue	Coupling constant [Hz]
D31	10,18
C32	9,20
Q34	4,69
E35	5,17
I36	9,28
S37	8,20
R38	7,84
C39	5,82
S40	2,55
K41	5,35
C42	4,76
R43	4,51
K44	4,76
Q45	7,24
S46	6,30
L48	4,94
Y49	8,59
E50	9,24
C51	4,32
D53	9,85
C54	9,61
M57	8,95

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

Name	Sequence 5'-3'	Purpose
HVO_2753_PstI_NdeI_NHis_fw	GACTAGCTGCAGGACTAGGACCATATGCACCACCACCACCACCACAGCGAGTCCGAACAGCGACAC	Construction of overexpression plasmid for HVO_2753 in *H. volcanii*
HVO_2753_KpnI_rev	GACTAGGGTACCTTATGGACCCATGAAGCCGCAGTCG	
HVO_2753_P1	CCTACGTCACCGCGACCAGC	Construction of pMH101_Δ2753 for the construction of the deletion mutant
HVO_2753_P2	TGGACCCATGAAGCCCTCGCTCATACGCTCATAAACCGGC	
HVO_2753_P3	GAGCGTATGAGCGAGGGCTTCATGGGTCCATAATCATGGG	
HVO_2753_P4	GCCGAGCAGTTCGGTATCGGAC	
HVO_2753_C1A_for	GACACGCGCACCAGGCTGTGTCCTGTGGCA	Construction of the respective point mutant complementation vector
HVO_2753_C1A_rev	TGCCACAGGACACAGCCTGGTGCGCGTGTC	
HVO_2753_C3A_for	GGCGACGTTCAAGGCCCCCGACTGCGGC	Construction of the respective point mutant complementation vector
HVO_2753_C3A_rev	GCCGCAGTCGGGGGCCTTGAACGTCGCC	
HVO_2753_C5A_for	CCAGGAGATTTCGCGTGCTTCCAAGTGCCGCAAG	Construction of the respective point mutant complementation vector
HVO_2753_C5A_rev	CTTGCGGCACTTGGAAGCACGCGAAATCTCCTGG	
HVO_2753_C7A_for	GCAACCTCTACGAGGCTCCCGACTGCGGCT	Construction of the respective point mutant

HVO_2753_C7A_rev	AGCCGCAGTCGGGAGCCTCGTAGAGGTTGC	complementation vector
HVO_2753_Q34A_for	CCCCGACTGCGGCGCGGAGATTTCGCGT	Construction of the respective point mutant complementation vector
HVO_2753_Q34A_rev	ACGCGAAATCTCCGCGCCGCAGTCGGGG	
HVO_2753_Y49F_for	CAAGCAGAGCAACCTCTTCGAGTGTCCCG	Construction of the respective point mutant complementation vector
HVO_2753_Y49F_rev	CGGGACACTCGAAGAGGTTGCTCTGCTTG	
HVO_2753_for_NHis	GACTAGCATATGCACCACCACCACCACCACAGCGAGTCCGAACAGCGACAC	Construction of the respective complementation vector
HVO_2753_rev_NHis	GACTAGAAGCTTTTATGGACCCATGAAGCCGCAGTCG	
HVO_2753_for_CHis	GACTAGCATATGAGCGAGTCCGAACAGCGACAC	Construction of the respective complementation vector
HVO_2753_rev_CHis	GACTAGAAGCTTTTAGTGGTGGTGGTGGTGGTGTGGACCCATGAAGCCGCAGTCG	
HVO_2753_fwd_nb	ATGAGCGAGTCCGAACAGCGACAC	Construction of the respective oligonucleotide probe
HVO_2753_rev_nb	TTATGGACCCATGAAGCCGCAGT	
HVO_2752_for_nb	AATCAAGGTCATGCCGAACAGCCC	Construction of the respective oligonucleotide probe
HVO_2752_rev_nb	CAGATACGGCCGACGTTCTCGAC	