Classification: Biological Sciences, Evolution

Displacement of the canonical single stranded DNA binding protein in the *thermoproteales*

Sonia Paytubi[1,3]*, Stephen A McMahon[1]*, Shirley Graham[1], Huanting Liu[1], Catherine H Botting[1], Kira S Makarova[2], Eugene V Koonin[2], James H Naismith[1] and Malcolm F White[1]

[1] Biomedical Sciences Research Complex, University of St Andrews, Fife KY16 9ST, UK. [2] National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894; [3] current address: *Departament de Microbiologia, Facultat de Biologia, Universitat de Barcelona, Avda. Diagonal 643, 08028 Barcelona, Spain*

*joint first authors

Corresponding authors: Prof Malcolm F White, Biomedical Sciences Research Complex, St Andrews University, North Haugh, St Andrews, Fife, KY16 9TZ. Tel. +44-1334 463432; email: mfw2@st-and.ac.uk. Prof James H Naismith, Biomedical Sciences Research Complex, St Andrews University, North Haugh, St Andrews, Fife, KY16 9TZ. Tel. +44-1334 463792; email: jhn@st-and.ac.uk

Author contributions:

SP, SAM and KM designed and performed the research and co-wrote the paper; SG, HL, CHB designed and performed the research; EVK, JHN and MFW designed research, analysed the data and wrote the paper.

Abstract

Single-stranded DNA binding proteins (SSBs) based on the OB-fold are considered ubiquitous in nature and play a central role in many DNA transactions including replication, recombination and repair. We demonstrate that the *thermoproteales*, a clade of hyperthermophilic crenarchaea, lack a canonical SSB. Instead, they encode a distinct ssDNA-binding protein that we term "ThermoDBP", exemplified by protein Ttx1576 from *Thermoproteus tenax*. ThermoDBP binds specifically to ssDNA with low sequence specificity. The crystal structure of Ttx1576 reveals a unique fold and mechanism for ssDNA binding, consisting of an extended cleft lined with hydrophobic phenylalanine residues and flanked by basic amino acids. Two ssDNA-binding domains are linked by a coiled-coil leucine zipper. ThermoDBP appears to have displaced the canonical SSB during the diversification of the *thermoproteales* – a highly unusual example where a "ubiquitous" protein has been lost in evolution.

\body

Introduction

Single-stranded DNA binding (SSB) proteins are essential for the genome maintenance of all known cellular organisms (1) and are present in many viruses (2, 3). These proteins play vital roles in DNA metabolism, sequestering and protecting transiently formed ssDNA during DNA replication and recombination (4, 5), melting double-stranded DNA (dsDNA), detecting DNA damage and recruiting repair proteins (6). The SSBs from the three domains of life share limited sequence similarity and display diverse subunit organisation. The low sequence conservation notwithstanding, all SSB family proteins contain one or more conserved Oligonucleotide Binding (OB) fold domain(s) (a five-stranded beta-sheet coiled to form a closed beta-barrel) that mediate ssDNA-binding with high affinity (7, 8). The organisation of OB folds in SSBs varies considerably. For example, *Escherichia coli* SSB is a homotetramer, with each subunit consisting of a single OB domain, in conjunction with a flexible C-terminal extension involved in protein-protein interactions (9, 10). The *Deinococcus/Thermus* SSBs, whilst still utilising the tetrameric functional binding mode, arrive at this arrangement by combining two SSB homodimers: each SSB monomer containing two OB folds linked by a conserved spacer sequence (11, 12). Moreover, the DdrB (DR0070) protein that is essential for radiation resistance in *Deinococcus radiodurans* is a highly divergent SSB homolog (13, 14).

Eukaryotes utilise a heterotrimeric SSB known as replication protein A (RPA) with six OB folds, two that mediate subunit interactions and four that are involved in

ssDNA binding (15, 16). In addition, metazoa encode one or more additional SSB proteins with a single OB-fold, exemplified by hSSB1 in *Homo sapiens*, that is implicated in the DNA damage response (17). The arrangement of euryarchaeal SSBs is similar to eukaryotic RPA: a polypeptide or polypeptides with multiple OB folds, including a characteristic OB fold interrupted by a zinc-binding domain (18-21). It appears that some euryarchaeal SSBs form heterotrimers and others heterodimers or monomers (19, 21, 22). In contrast, in most crenarchaea SSB has a bacterial-like domain structure, with a single OB fold followed by a flexible C-terminal tail that is not involved in DNA binding (23). The crystal structure of the OB fold of *Sulfolobus solfataricus* SSB demonstrated its close structural relationship with the ssDNA-binding domains of human RPA70 (24).

Structural and bioinformatic studies have identified a characteristic sequence signature for the OB-fold that allows its detection even in genomes that encode highly diverged versions of the SSB. OB-fold containing SSB proteins have been detected in all three domains of life but, as we reported previously, one group of crenarchaea, the *thermoproteales*, appear to lack an identifiable *ssb*-encoding gene (25). There are now ten fully sequenced genomes in this group (*Thermoproteus tenax, T. uzoniensis*, *T. neutrophilius*, *Caldivirga maquilingensis*, *Pyrobaculum aerophilum*, *P. arsenaticum*, *P. islandicum* and *P. calidifontis, Vulcanisaeta moutnovskia, V. distributa*) that lack any identifiable *ssb* genes. By contrast, only one sequenced genome in this clade, *Thermofilum pendens*, does encode (two) SSB proteins. We reasoned that the ten species of *thermoproteales* that apparently lack a canonical SSB must utilise an alternative single stranded DNA binding protein. By screening *T. tenax* cell extracts for ssDNA binding proteins biochemically, we identified a single candidate, Ttx1576, which was unique to the species lacking a canonical SSB. The gene encoding Ttx1576 was cloned and the protein shown to possess properties consistent with a role as a SSB. Structural characterisation of Ttx1576 has revealed an ssDNA binding domain with a distinct fold, attached to a C-terminal leucine zipper dimerization domain.

Results

Identification of *T. tenax* proteins binding to single-stranded DNA

SSB and RPA proteins have been identified previously from crude cell extracts using gel-shift experiments with labelled ssDNA (23) (26). To identify ssDNA-binding proteins from *T. tenax*, we used a related affinity purification approach. A biotinylated 45 nucleotide oligonucleotide was bound to magnetic streptavidin beads and incubated with *T. tenax* cell lysate for 90 min at 50 °C to maximise the opportunity for

a binding equilibrium to develop. The beads were harvested and washed in buffer with progressively higher NaCl concentrations. After each set of washes, the supernatants were collected and proteins precipitated using TCA/acetone prior to separation by SDS-PAGE. As shown in Figure 1A, this approach yielded a number of distinct protein bands following SDS-PAGE, which were excised and identified by mass spectrometry. The affinity purification experiment was repeated 3 times, with highly reproducible results. We expected to observe high abundance proteins known to bind to ssDNA or RNA and this was indeed the case. Prominent bands included several subunits of RNA polymerase, a DNA helicase (Ttx0530), a RadA paralog (Ttx1408), RNase E (Ttx0105) and transcription termination factor NusA (Ttx1674). In addition, we identified two proteins of unknown function, Ttx2090 and Ttx1576, which were characterised in more detail.

The Ttx2090 protein has homologs in archaeal species, such as *Aeropyrum pernix* and *Igneococcus hospitalis,* which encode a canonical SSB. The *ttx*2090 gene was cloned into an *E. coli* expression vector, but we did not succeed in obtaining soluble protein. Bioinformatic analysis suggested that it may be a member of the prefoldin family of protein chaperones (>95 % probability by HHPred (27)), and may therefore have been purified in association with partially unstructured proteins. In contrast, the Ttx1576 protein belongs to archaeal COG arCOG05578 (28), which is represented in all available genomes of *thermoproteales,* with the sole exception of *T. pendens,* and no other sequenced genomes. It therefore shows perfect complementarity with the phyletic pattern of RPA/SSB proteins (arCOG01510) (supplementary dataset 1). Thus, we focussed on Ttx1576 as the candidate for a unique ssDNA binding protein in *thermoproteales*.

Cloning and expression of Ttx1576

In order to test whether Ttx1576 had properties consistent with a ssDNA binding protein *in vitro*, we amplified the gene from *T. tenax* chromosomal DNA by PCR and cloned it into an *E. coli* expression vector (pET151/D-TOPO) allowing expression in *E. coli* with a cleavable N-terminal polyhistidine tag. The protein was purified by immobilised metal affinity chromatography and the his-tag was removed with TEV protease. Ttx1576 was then purified again by immobilised metal affinity chromatography and the flowthrough was collected, yielding essentially homogeneous protein (Figure 1B). This preparation was used for ensuing biochemical and structural studies. Polyclonal antibodies against Ttx1576 were raised in sheep and used to estimate the amount of the protein present in *T. tenax* cells (Figure 1C). Western blotting was used to quantify the levels of Ttx1576 in a

defined quantity of cell extract using the recombinant protein for calibration. This method gave an estimation of levels of Ttx1576 in the cell at between 0.07-0.13% of total soluble protein. By comparison, cellular levels of *S. solfataricus* SSB were estimated at between 0.08-0.16% of total soluble protein. Thus, Ttx1576 has a cellular concentration consistent with a putative role as a single-stranded DNA binding protein.

Ttx1576 binds ssDNA specifically

To determine the nucleic acid binding properties of Ttx1576, we carried out gel electrophoretic mobility shift experiments using a 24mer oligonucleotide of mixed sequence (5'-CTTTCAATTCTATAGTAGATTAGC) with a fluorescent label at the 5' end. Protein and DNA were mixed and incubated at either 20 °C or 80 °C for 10 min prior to loading on a polyacrylamide gel for electrophoresis and subsequent imaging (Figure 2). At both temperatures, a clear retarded species was observed in the gel corresponding to an apparent K_D of approximately 0.6 μM. Incubation at the higher temperature did not appear to influence the binding affinity significantly, although slightly more unbound DNA was observed at higher protein concentrations under these conditions. When the experiment was repeated with an RNA oligonucleotide of the same sequence the binding affinity was significantly weaker and no specific retarded species was observed (Figure 2). Binding to a DNA duplex of the same sequence was very weak, with no evidence of retarded products observed at the highest protein concentration. Together, these data suggest that Ttx1576 binds specifically to ssDNA and that this interaction is not especially sensitive to incubation temperature. Previous studies of *S. solfataricus* SSB have demonstrated that ssDNA:SSB interactions are exothermic with binding affinity higher at lower temperatures, although K_D's are also influenced strongly by the ions present in the buffer (29).

To obtain more quantitative values for DNA binding affinity, TTX1576 binding to single- and double-stranded DNA was analysed by isothermal titration calorimetry. Oligonucleotide 21T (composed of 21 deoxythymidine nucleotides) was titrated into a 10 μM solution of Ttx1576. The binding isotherm (Figure 3A) revealed exothermic binding characteristic of SSBs (24) with a calculated dissociation constant of 160 nM and a stoichiometry of 2 protein monomers (or one dimer) bound per oligonucleotide. By contrast, a 21 base pair dsDNA molecule of mixed sequence (ds21mix) was not bound by Ttx1576 under these conditions (Figure 2B), confirming the specificity for ssDNA. To examine the sequence specificity of Ttx1576, we utilised oligonucleotides

of different sequences labelled with a 5'-fluorescein dye and measured changes in anisotropy upon protein binding (Figure 3C). Binding isotherms for three oligonucleotides whose sequences are listed in the methods were determined in triplicate: 21T, 21Grich and 21C. The 21T oligonucleotide was bound with an affinity of 100 ± 8 nM, in good agreement with the dissociation constant calculated using ITC. The G-rich oligonucleotide was bound with a similar K_D, 116 ± 11 nM, whilst the 21C oligonucleotide was bound more tightly with a K_D of 22 ± 5 nM. Neither the ITC nor the anisotropy experiments revealed any evidence for cooperative binding, a known property of other SSBs. This was almost certainly due to the relatively small lengths of DNA used, which appear to accommodate only one dimer of the protein. Finally, binding to RNA was investigated using a fluorescently labelled RNA oligonucleotide 21U. The resulting binding isotherm generated a significantly higher dissociation constant of 2.1 ± 0.2 μM, in good agreement with the gel retardation data.

Taken together, the data indicate that Ttx1576 binds in preference to ssDNA and shows limited sequence specificity: properties consistent with a role as a SSB. We therefore propose the name "ThermoDBP" for this *thermoproteales*-specific protein family, to distinguish it from canonical SSB proteins.

The structure of ThermoDBP reveals a distinct fold and DNA binding surface

The full length protein having failed to give crystals with useful diffraction, we incubated Ttx1576 with chymotrypsin (30) during crystallisation. This gave crystals (denoted cTtx1576), which diffracted well. The cTtx1576 structure was solved by selenomethionine incorporation and anomalous diffraction. In the final model of cTtx1576 there was one monomer in the asymmetric unit consisting of residues 24 – 139. Because the protein was obtained in the presence of protease, it was not possible to be certain whether disorder or cleavage was the cause of the missing residues (1-23, 140-196). Based on sequence analysis described in detail later, we constructed, purified and crystallised a Ttx1576 mutant corresponding to amino acids 1-148. Crystals diffracted to 2.0 Å and the structure was solved using molecular replacement, revealing that residues 10 – 148 were well ordered (Figure 4A). The protein structure consists of a single compact domain, comprising four α helices (α1-4) and a four-stranded anti-parallel β sheet (β1-4), measuring approximately 50 Å x 20 Å x 20 Å. The N-terminus of the structure comprises β1 and β2 followed by the 4 α-helices that form three sides of a distorted quadrilateral. The fourth side is completed by the loop between α-4 and β-3. The β sheet packs against one face of

the quadrilateral, leaving an extended cleft open to solvent on the other face. Strikingly, nine phenylalanine residues are distributed along the flanking helices and form a continuous hydrophobic patch that runs along the length of the cleft (Figure 4B). The outer edge of the binding cleft has a strongly positive electrostatic surface potential due the presence of the conserved basic residues R49, K54, R65, R80, R86, R90, K97 and R112 (Figure 4B,C). Thus, the putative ssDNA binding cleft of ThermoDBP has a hydrophobic, aromatic core suitable for interaction with the nucleobases of ssDNA and a positively charged periphery for electrostatic interactions with the phosphodiester backbone.

Searching for structurally similar proteins using either PDBFold or Dali suggested that ThermoDBP has a unique fold. Allowing more tolerance than the default, it was possible to detect weak structural similarity with two RNA-binding proteins, HutP, (PDB 1wrq, an RNA binding anti-termination protein (31)) and the L31e protein from the large ribosomal subunit of *Haloarcula marismortui* (1yj9) (32). An optimum alignment of HutP and Ttx1576 matches 75 Cα with a root mean square deviation (rmsd) of 2.6 Å. Visual inspection reveals that the proteins have almost identical topology but share essentially no sequence similarity. In HutP, the secondary structural elements are displaced relative to ThermoDBP, and the cleft is absent. The RNA- binding site of HutP is on the side of the structure, parallel to α1 but remote from the presumed ThermoDBP ssDNA binding cleft. The L31e protein has the same topology as ThermoDBP, (superimposing 57 Cα atoms with an rmsd 3.1 Å) but lacks the cleft. L31e forms part of the polypeptide exit tunnel of the ribosome in close association with rRNA (32).

Domain organization of ThermoDBP

The C-terminus of ThermoDBP, which was removed by proteolysis or mutagenesis prior to crystallisation, is strongly predicted to adopt a coiled-coil, amphipathic leucine zipper structure. Circular dichroism of the full-length protein shows an overall helical content of 47 %, which is markedly higher than the 27 % helical content of the crystallised DNA binding domain, consistent with a strongly helical C-terminal domain. Analysis of the ThermoDBP sequence by Conserved Domain Database search (33) revealed a statistically significant (E-value=2.79e-03) match with profile cl02576 corresponding to the bZIP domain (Figure 4D). Multiple alignments of ThermoDBP family proteins revealed conservation of several leucines in the last alpha helical region (Figure S1). Leucine zippers are dimerization domains,

suggesting that ThermoDBP may have a dimeric structure with an N-terminal ssDNA binding domain and a C-terminal dimerization domain (modelled in Figure 4E),

To investigate the roles of different domains in the function of ThermoDBP, we made two different C-terminally truncated mutant versions and compared their DNA binding affinities by gel electrophoretic mobility shift of a fluorescent oligonucleotide. In the first mutant (1-148), the predicted leucine zipper domain was removed by introducing a stop codon at amino acid position 149. This mutant protein still bound ssDNA with an affinity only slightly lower than for the full-length wild-type protein (Figure 5A). Retarded DNA migrated more quickly than for the full-length protein, suggesting a smaller nucleoprotein complex. Deletion of a further 9 amino acids including the conserved "LIYWIRSDR" sequence (mutant 1-139) showed significantly weaker DNA binding activity, suggesting that the conserved sequence motif participates in ssDNA binding. The truncation mutant 1-139 elutes from a calibrated size exclusion column with a retention time corresponding to a molecular weight of 17 kDa, consistent with a monomeric domain molecular weight of 16 kDa (Figure 5B). In contrast the full-length protein has a retention time corresponding to a molecular weight of 57 kDa. This is slightly higher than the expected value of 46 kDa for a dimeric structure, probably due to the elongated shape of the predicted dimer, a factor that is known to influence retention times in size exclusion chromatography.

Distant archaeal homologs of ThermoDBP

An HHpred search starting from the Ttx1576 sequence (1-146aa) revealed statistically significant similarity (probability=94%) with the pfam10015 protein family (DUF2258, also known as COG4345). The reverse search using PSI-BLAST with SSO1098, a member of DUF2258 from *Sulfolobus solfataricus,* as a query, detected statistically significant similarity (E-value=0.001) with a ThermoDBP family representative, Pcal_0963 from *Pyrobaculum calidifontis*. The DUF2258 family contains archaeal proteins only, and in arCOGs the phyletic profile for the corresponding arCOG03772 includes proteins from *Thermoproteales*, *Desulfurococcales*, *Thermococci* and a few *Archaeoglobi* (see Figures S1 and S2). Multiple alignment of this protein family revealed conservation of all eight secondary structure elements of the ThermoDBP fold (Figure S2). In addition, all these proteins contain a predicted long alpha helix after the core domain suggesting that, similarly to ThermoDBP, these proteins can dimerize. No conservation of the leucine zipper signature was observed. A group of proteins within the arCOG03772 family contain an additional C-terminal domain with predicted mixed alpha and beta secondary structure, for which no similarity to any known protein could be detected (Figure S1).

We have not found any conservation of gene neighbourhood for this family, so there are no specific clues as to the function(s) of these distant homologs of ThermoDBP. Nevertheless, the sequence conservation and broad distribution of this family in Archaea suggest an important biological role. Given that the member of arCOG03772 from *T. tenax* (Ttx1840) was not detected in our ssDNA-binding assay, there seems to be a distinct possibility that these homologs of ThermoDBP are not ssDNA-binding proteins or are not highly expressed.

Discussion

Strikingly, ThermoDBP is the only protein family that is present in all ten *thermoproteales* species lacking a canonical SSB and absent in all species encoding an SSB. The complementarity of the phyletic patterns of SSB and ThermoDBP suggests the possibility that, in the *thermoproteales* lacking canonical SSB proteins, ThermoDBP supplies the essential ssDNA-binding activity, in a dramatic case of non-orthologous gene displacement (34). An alternative possibility remained that ssDNA binding in the *thermoproteales* was mediated by an extremely divergent OB-fold containing SSB variant that was undetectable by sequence analysis. This possibility appears remote because in the course of arCOG construction RPA orthologs have been detected with a high degree of confidence using sensitive sequence profiles in all available archaeal genomes except for Thermoproteales, including deep and possibly fast evolving lineages such Korarchaeota and Nanoarchaeota (28). More importantly, by affinity purifying ssDNA binding proteins from *T. tenax*, we confirmed that, even if such a protein was present in these organisms, the principal ssDNA-binding activity in the *thermoproteales* cells resides in a distinct protein, ThermoDBP (Ttx1576), which possesses functional properties characteristic of an SSB.

Although ThermoDBP shares topology with two known RNA- binding proteins, it possesses unique structural features and is unrelated to the canonical, OB fold-containing SSB family. The structure of ThermoDBP reveals that the large central phenylalanine-lined cleft is the site for DNA binding. The phenylalanine residues are spaced out along the cleft with a separation of between 3.5 to 5.5 Å, potentially allowing nucleotide bases to insert into the cleft. Stacking of aromatic and hydrophobic side chains against nucleotide bases is a well known theme in nucleic acid-binding proteins, in particular the OB fold found in SSB proteins (24). The strongly positive electrostatic charge surrounding the cleft appears optimal for binding the negatively charged phosphate backbone. The C-terminal domain has the characteristic signature of an amphipathic helical leucine zipper (Lx_6L) and is

implicated in the dimerization of ThermoDBP. The presence of a leucine zipper in a family of archaeal proteins is unusual. A common origin of leucine zippers in ThermoDBP and eukaryotic bZIP transcription factors is a provocative possibility but, given the generic coiled coil structure of this domain, convergence cannot be ruled out.

Our search of protein sequence databases for distant homologs of ThermoDBP identified a distinct family of archaeal proteins, arCOG03772. These Archaea-specific proteins have a broader phyletic distribution than ThermoDBP and seem to contain counterparts to all structural elements of the ThermoDBP fold as well as a putative coiled coil dimerization domain that is however distinct from the ThermoDBP leucine zipper. Functional characterization of arCOG03772 proteins is an interesting goal for further experiments.

Non-orthologous gene displacement is a common phenomenon in genome evolution that encompasses even central cellular functions, in particular key components of the DNA replication machinery (35). However, the SSB seemed to remain one of the few "truly universal" proteins. The present work shows that even for this fundamental function unrelated solutions have evolved.

Materials and Methods

Growth of *T. tenax* and cell lysis. One gram of *T. tenax* Kra1 cells, generously provided by Dr Bettina Siebers (University of Duisberg, Essen) was resuspended in binding buffer (BB) (20mM MES pH 6.5, 50mM NaCl, 5mM EDTA and 1mM DTT) and lysed by sonication. The lysate was centrifuged at 16,000 x g for 30 min at 4 ºC and the resulting supernatant filtered through a 0.45 μM filter.

Detection of proteins interacting with ssDNA. 0.6 mg of magnetic streptavidin-beads (Promega) were washed 3 times with 0.5 x SSC buffer (75 mM NaCl, 7.5 mM sodium citrate, pH 7.0). 1500 pmol (22 μg) of a single-stranded biotinylated oligonucleotide (Biot-45ssDNA) was bound to the beads and left at room temperature for 10 min. The beads were washed again three times with 0.1 x SSC buffer and 20 mg of cell lysate was added. The mix was incubated for 90 min at 50 ºC. After this time, the beads were washed five times with buffer BB containing 150 mM NaCl. Bound proteins were eluted progressively with buffer BB supplemented with 250, 500 and 1000 mM NaCl. All the fractions were TCA/Acetone precipitated, resuspended in

SDS-PAGE loading buffer, boiled and run on a 4-12 % Bis-Tris gel (Invitrogen). Gels were stained with SYPRO Ruby and visualised under ultraviolet light.

Mass spectrometry: protein identification. Bands from the SDS-PAGE gel were excised into approximately 1 mm cubes. These cubes were then subjected to in-gel digestion, using a ProGest Investigator in-gel digestion robot. Briefly, the gel cubes were destained by washing with acetonitrile and subjected to reduction and alkylation before digestion with trypsin at 37 °C. The peptides were extracted with 10% formic acid and concentrated to 20 µl (SpeedVac; ThemoSavant). They were then separated using an UltiMate nanoLC (LC Packings, Amsterdam) equipped with a PepMap C18 trap and column using a 60 or 90 min elution profile, depending on the complexity (mol wt) of the sample, with a gradient of increasing acetonitrile, containing 0.1 % formic acid, to elute the peptides (5-35 % acetonitrile over 18 min or 40 min respectively, 35-50 % for a further 7 or 20 min, followed by 95 % acetonitrile to clean the column, before re-equilibration to 5 % acetonitrile). The eluent was sprayed into a Q-Star XL tandem mass spectrometer (Applied Biosystems, Foster City, CA) and analyzed in information-dependent acquisition mode, performing 1 s of MS followed by 3 s MSMS analyses of the 2 most intense peaks seen by MS. These masses were then excluded from analysis for the next 60 s. Tandem MS data for doubly and triply charged precursor ions were converted to centroid data, without smoothing, using the Analyst QS1.1 mascot.dll data import filter with default settings. The MS/MS data file generated was analysed using the Mascot 2.1 search engine (Matrix Science) searching against a database containing the protein translations of the *T. tenax* open reading frames. The *T. tenax* genome sequence was provided by Bettina Siebers prior to publication (36). The data was searched with tolerances of 0.2 Da for the precursor and fragment ions, trypsin as the cleavage enzyme, one missed cleavage, carbamidomethyl modification of cysteines as a fixed modification and methionine oxidation selected as a variable modification. The Mascot search result was accepted if a protein match with a significant score was observed that was significantly above the score for other matches and included at least one peptide with a score above the homology threshold.

Cloning, mutagenesis and expression of Ttx1576. The Ttx1576 gene was amplified by PCR using the following primers:

Ttx1576_for (5'-CACCGGAGAGGAGCTAAGAGAGGAG) and

Ttx1576_rev (5'-TTATTTCAATAAACTTGTTATC) and cloned into the pET151/D-TOPO vector (Invitrogen) following the manufacturer's instructions. pET151/D-TOPO-Ttx1576 was transformed to *E.coli* BL21-Star (DE3) cells. Cells were grown in 2 l of LB medium at 37 °C to an OD_{600} of 0.6. At this point, induction of the His_6-

tagged Ttx1576 was carried out by 1 mM IPTG for 3 h at 37 °C. Truncated mutant versions of the Ttx1576 protein were created by introducing a stop codon at position 140 or 149 to remove the C-terminal domain, as described (37). These proteins were purified as for the wild-type protein but eluted more slowly on gel filtration, consistent with the disruption of the C-terminal dimerization domain (Figure 5).

Ttx1576 protein purification. Cells were harvested and resuspended in Lysis buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.1 % Triton X-100, 1 mM $MgCl_2$ and Complete EDTA-free protease inhibitors (Roche)), lysed by sonication and clarified by centrifugation. The supernatant was heated to 70 ºC for 10 min and re-centrifuged. The resultant supernatant was diluted two-fold in Buffer A (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 30 mM NaH_2PO_4) with 30 mM imidazole and filtered through a 0.45 µM filter. The sample was then applied to a column containing Ni-NTA-Agarose (HiTrap 5ml Chelating HP; GE Healthcare) pre-equilibrated with buffer A + 30 mM imidazole. The protein was eluted with a linear gradient of 0 – 500 mM imidazole in Buffer A. Fractions containing the His-Ttx1576 protein were identified by SDS-PAGE and pooled. His-Ttx1576 was buffer exchanged against TEV cleavage buffer (20 mM Tris-HCl pH 7.0, 500 mM NaCl, 1 mM DTT, 30 mM NaH_2PO_4 and 10 % glycerol). The protein was cleaved with the TEV protease overnight at room temperature by adding a final concentration of 200 ng/µl of TEV protease. Cleaved Ttx1576 was re-purified by loading onto the same column pre-equilibrated with buffer A + 30 mM imidazole and collecting the flow through. Positive fractions were pooled and buffer exchanged extensively against storage buffer (50 mM Tris-HCl pH 7.5, 200 mM KCl, 1 mM DTT, 1 mM EDTA, 0.01 % Triton X-100 and 50 % glycerol). Mass spectrometry confirmed the expected mass for the recombinant protein following tag removal. Protein to be used for crystallization was prepared as described in (38).

Quantitative Western blotting. For the production of Ttx1576 antibodies, 1 mg of purified Ttx1576 protein was used to raise polyclonal antibodies in sheep (Scottish National Blood Transfusion Service). Western blots were performed in order to detect Ttx1576 and SSB proteins. Defined amounts of Ttx1576 and SSB were run on NuPage 4-12 % Bis-Tris SDS gels (Invitrogen) along with 25 and 75 µg of protein prepared from *T. tenax* and *S. solfataricus* lysates. Western blots were carried out following standard procedures. Protein content was measured by Bradford assay.

The concentration of endogenous Ttx1576 and SSB in the extracts was determined using a standard curve generated with the recombinant proteins.

Oligonucleotides.

Biot-45ssDNA:

5’ biotin-GTTTGAAACTACTTTTAACTATAAGTTAAAATGACTCTTAAATAG

Fluorescent 24mer DNA: 5’FAM- CTTTCAATTCTATAGTAGATTAGC

Fluorescent 24mer RNA: 5’FAM- CUUUCAAUUCUAUAGUAGAUUAGC

Fluorescent 21T: 5’FAM- TTTTTTTTTTTTTTTTTTTTT

Fluorescent 21C: 5’FAM- CCCCCCCCCCCCCCCCCCCCC

Fluorescent 21G-rich: 5’ FAM-TTCTGGGGCTGGGGCTGGGGT

Fluorescent 21U: 5’ FAM-UUUUUUUUUUUUUUUUUUUU

Fluorescent 45mer:

5’FAM-GCTTGCTAGGACGGATCCCTCGAGGTTTTTTTTTTTTTTTTTTT

21T for ITC: 5’ TTTTTTTTTTTTTTTTTTTT

The DNA duplex ds21Mix used for ITC experiments was assembled from the following oligonucleotides by mixing and slow cooling from 80 °C to 30 °C over 3 h:

21Mix-For 5’ ATTCAGTTCAACTGTTAGACT

21Mix-Rev 5’ AGTCTAACAGTTGAACTTGAAT

Gel electrophoretic mobility shift assays.

Binding of wild-type and mutant forms of Ttx1576 to fluorescent DNA and RNA oligonucleotides (200 nM) was performed in binding buffer (50 mM Tris-HCl pH7.5, 50 mM NaCl, 0.1 mg/ml BSA). Reactions were incubated for 10 min at 20 °C or 80 °C before addition of ficoll loading buffer, run on an 8 % acrylamide gel in 1 x TBE at 15 mA constant current for 60-90 min, then scanned using a Fuji FLA5000 fluorescent imager.

Isothermal titration calorimetry. Binding of Ttx1576 to ssDNA and dsDNA was assessed by isothermal titration calorimetry (ITC) using a VP-ITC unit (Microcal, GE Healthcare, Chalfont St Giles, UK). Ttx1576 protein samples were dialysed extensively against ITC buffer (50 mM Tris-HCl pH 7.5, 50 mM NaCl) and degassed in a vacuum. Oligonucleotides were also dissolved in ITC buffer. The binding experiments were performed at 25 °C. A 285 µl syringe with stirring at 300 rpm was used to titrate the 21T or the ds21Mix oligonucleotides (75 µM) into a cell containing 1.4 ml of Ttx1576 protein (10 µM). Titrations comprised 50 injections of ssDNA or dsDNA, one 2 µl injection followed by 49 x 5 µl injections. The initial data point was routinely deleted to allow for diffusion of ligand/receptor across the needle tip during the equilibration period. ITC-binding isotherms were analysed using a simple single-binding site model with ITC data analysis software (ORIGIN) provided by the manufacturer.

Fluorescence Anisotropy. The binding affinities of Ttx1576 to several oligonucleotides (21T, 21C, 21G-rich and 21U) were determined at 20 °C using a Cary Eclipse fluorimeter (Varian) with automatic polarizer (excitation, 490 nm;

emission, 535 nm). Both the excitation and emission slit width were set at 5 nm. For direct titration, 25 nM 5′-fluorescein labeled 21-nt oligonucleotide was equilibrated in 500 µl Fluorescence buffer (50 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM DTT, 1 mM EDTA). Total fluorescence intensity was measured in parallel following each protein addition and the effects of dilution were corrected. Fluorescence quenching higher than 20 % was only observed for the 21U oligonucleotide. To minimise rotational effects on fluorescence intensity, "magic angle" conditions were used. Each protein titration was repeated in triplicate. Data were fitted, using Kaleidagraph, to the following equation:

$A = A_{min} + [(D + E + K_D) - ((D + E + K_D)^2 - (4DE))^{1/2}](A_{max} - A_{min})/(2D)$

(A, measured anisotropy; E, variable protein concentration; D, total DNA concentration; A_{min}, anisotropy of free DNA; A_{max}, anisotropy of DNA-protein complex; K_D, dissociation constant) (39). The protein concentration recorded corresponded to the concentration of Ttx1576 subunits.

Structural Biology. Protein for crystallisation was concentrated to between 15 and 30 mg/ml for crystallisation trials set up as part of a previous structural proteomics effort (38). This approach failed to yield any useful crystals. We re-purified the protein and this time incubated it with chymotrypsin (1:300 v/v) prior to crystallisation as a rescue method for intractable proteins (30). The optimum native protein crystallisation conditions were refined to 15 mg/mL of chymotrypsin-treated protein equilibrated against 0.9 M sodium tartrate, 0.1 M bicine pH 8.5 and 0.05 M sodium potassium phosphate. These crystals grew in 2 weeks and were reproducible. For selenomethionine-labelled protein the best crystals grew from 28 mg/ml of chymotrypsin-treated protein against a solution of 3.05 M sodium chloride, 0.1 M bicine pH 9.5 and 0.26 M lithium chloride in 2 weeks. Phases were determined by a Se-SAD experiment performed on a single selenomethionine-labeled Ttx1576 crystal. The data have an usually high R(sym) (22.4%) as a result of both radiation damage to the crystal and the presence of ice. The crystal was exposed to the synchrotron beam without any attenuation to get data to 3.5 Å. The high resolution and much higher quality data for the native protein were then used to refine the structure. Data were collected at 100 % transmission on beamline I03 at the Diamond synchrotron light source Oxfordshire, England. A 2.9 Å native data set was collected in-house using a Rigaku Micromax™ - 007HF Cu anode with VariMax optics alongside a Rigaku Saturn 944+ CCD detector. All data were indexed and scaled with HKL2000 (40). Using both the native and derivative data the single selenium site was located with SHELXD using the SHELC/D/E (41) suite of programs

including the test version of SHELXE. Parrot and Buccaneer, part of CCP4 (42), were used for further automated model building. In contrast the mutant Ttx1576 1-148 crystallised readily overnight when equilibrated against 11.8 % isopropanol, 0.1 M sodium citrate pH 5 and 0.2 M lithium sulphate at protein concentration 15 mg/ml. High resolution data were collected on beamline I04-1 at Diamond and indexed and scaled with HKL2000 (40). Using the 2.9 Å resolution structure as a model Ttx1576 1-148 was solved by molecular replacement using Phaser in CCP4 (43). Both models were refined with REFMAC5 (44, 45) and Coot (46), with model quality assessed by MOLPROBITY (47). The final coordinates are available in the protein databank with identification code 3TEK.

CD Spectroscopy

Ttx1576 was analysed by CD to access whether the missing regions of the crystal structure were, as predicted, largely α-helical. Spectra were recorded on a JASCO J-180 spectrometer with protein at 0.15 mg/ml in PBS. The recorded spectrum was analysed using the Dichroweb server (48).

Size exclusion chromatography

A superose 12 column (GE Healthcare) was calibrated using molecular weight standards (blue dextran, thyroglobulin, bovine gamma globulin, chicken ovalbumin, equine myoglobin and vitamin B12) in GF buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl) with a flow rate of 0.8 ml / min. The full length and 1-139 truncation mutant of Ttx1576 were analysed in the same conditions. The standards yielded a linear relationship for Kav to log molecular weight, which was used to calculate the native molecular weight of the Ttx1576 proteins (49).

Bioinformatic analysis

Protein sequence database searches were performed using PSI-BLAST (50) with an inclusion threshold E-value of 0.01 and no composition-based statistical correction against the non-redundant (NR) database at the NCBI. In addition, distant similarity detection approaches were applied, namely the conserved domain database (CDD) search (51) and the HHpred search (27). Multiple alignments of protein sequences were constructed by using the MUSCLE program (52), followed, when necessary, by a minimal manual correction on the basis of local alignments obtained using PSI-BLAST and HHpred programs. Protein secondary structure was predicted using the Jpred program (53).

Acknowledgements

This work was funded by the Biotechnology and Biological Sciences Research Council, grant reference BB/S/B14450. Thanks to the University of St Andrews BMS

Mass Spectrometry and Proteomics Facility, which is funded by grants from the Wellcome Trust. KM and EK are supported by intramural funds of the US Department of Health and Human Services (to National Library of Medicine, NIH). Thanks to Richard Hutton for helpful discussions and Bettina Siebers for the *T. tenax* biomass and genome sequence.

Figure legends

Figure 1. Identification and purification of ssDNA binding proteins from *T. tenax.*

A. SDS-PAGE analysis of *T. tenax* proteins binding to a biotinylated 45 nucleotide DNA oligonucleotide. Proteins identified by mass spectrometry from excised gel bands are labelled. For each protein, the Ttx gene number is indicated along with a description.

B. SDS-PAGE analysis of the purified, heterologously-expressed Ttx1576 protein.

C. Estimation of the cellular levels of *T. tenax* Ttx1576 and *S. solfataricus* SSB. The concentrations of Ttx1576 and SsoSSB in soluble *T. tenax* and *S. solfataricus* cell lysates, respectively, were analysed by SDS-PAGE and western blot. Purified recombinant proteins were used to calibrate the result.

For all figure sections, MW represents the molecular weight marker lane.

Figure 2. Interaction of Ttx1576 with different nucleic acids

Electrophoretic mobility shift assays showing binding of Ttx1576 to a 24 nucleotide DNA oligonucleotide (top), RNA oligonucleotide of the same sequence (middle) and DNA duplex (bottom). Oligonucleotides (0.2 μM) were incubated at 20 or 80 °C for 10 min with varying concentrations of Ttx1576 prior to electrophoresis at room temperature. Ttx1576 concentrations in each experiment were (left to right): 0, 0.05, 0.1, 0.2, 0.4, 0.8, 1.6, 3.2 and 10 μM.

Figure 3. Quantitative measurements of Ttx1576 binding to DNA and RNA.

(A) Quantification of Ttx1576 binding to a 21T oligonucleotide by ITC. Oligonucleotide 21T (75 μM in the syringe) was injected into a 10 μM solution of Ttx1576. The data were fitted with a simple one site binding model, yielding a K_D of 160 nM and a binding stoichiometry of 2:1 (Ttx1576 monomers:oligonucleotide). **(B)**

Quantification of Ttx1576 binding to a 21 bp DNA duplex of mixed sequence (21Mix) by ITC. 21Mix dsDNA (75 μM in syringe) was injected into a solution of 10 μM of Ttx1576 protein. No binding was observed. **(C)** A plot of the anisotropy changes resulting from binding of Ttx1576 to 21 nucleotide oligonucleotides tagged with a 5′ fluorescein reporter molecule was used to determine dissociation constants (21C, K_D 22 nM; 21T, K_D 100 nM; 21G-rich, K_D 116 nM; and the RNA oligonucleotide 21U, K_D 2.1 μM). Experiments were carried out in triplicate and means with standard errors are shown. Data were fitted as described in the Material and Methods.

Figure 4. The crystal structure of Ttx1576 reveals a distinct fold with ssDNA binding features.

A. Structure of the DNA binding domain of Ttx1576 (residues 10-148) showing the conserved aromatic and basic residues that line the proposed ssDNA binding cleft. The N- and C-termini are indicated with blue and red spheres, respectively.
B. A surface representation shown in the same orientation as in A, colored to show conserved phenylalanine residues in yellow and basic residues in blue. The hydrophobic binding cleft is lined with basic amino acids that are well positioned to interact with the phosphodiester backbone of bound ssDNA.
C. A surface representation shown in the same orientation as in A and B, indicating the electrostatic potential of the putative binding interface.
D. Sequence alignment of the C-terminal region of Ttx1576 (residues 154-192) with a Hidden Markov Model corresponding to the basic leucine zipper (BZIP) family, with conserved residues indicated.
E. A model full-length structure of Ttx1576, including the C-terminal leucine zipper motif. The model for the C-terminus was constructed using the zipper motif from the Jun BZIP homodimer (PDB code 2H7H). The binding cleft in the DNA binding domain is denoted by phenylalanine side chains in yellow and the termini defined in the crystal structure indicated by spheres as before. The model is not intended to represent the relative position of the two ssDNA binding domains accurately.

Figure 5. Truncation of the C-terminus of Ttx1576 alters quaternary structure and DNA binding affinity.

A. Gel-shift analysis of Ttx1576 binding to a fluorescent 45mer oligonucleotide. The wild-type protein bound efficiently to the DNA but deletion of the C-

terminal sequence after residue 139, which removes the leucine zipper region and the conserved sequence between 140-148, reduced binding affinity significantly. The deletion construct 1-148, which also lacks the leucine zipper but includes the conserved sequence, retained much of the binding affinity of the wild-type enzyme. Protein concentrations (left to right) were 0, 1, 5, 10, 20, 50 and 100 μM.

B. The molecular weights of the full length (WT) and C-terminally truncated 1-139 mutant Ttx1576 proteins were estimated by gel filtration using a calibrated superpose 12 column. The elution volumes were consistent with a dimeric structure of the WT protein and a monomeric structure for the mutant.

C. Schematic showing the domain organisation of Ttx1576. The crystallised portion is shaded.

References

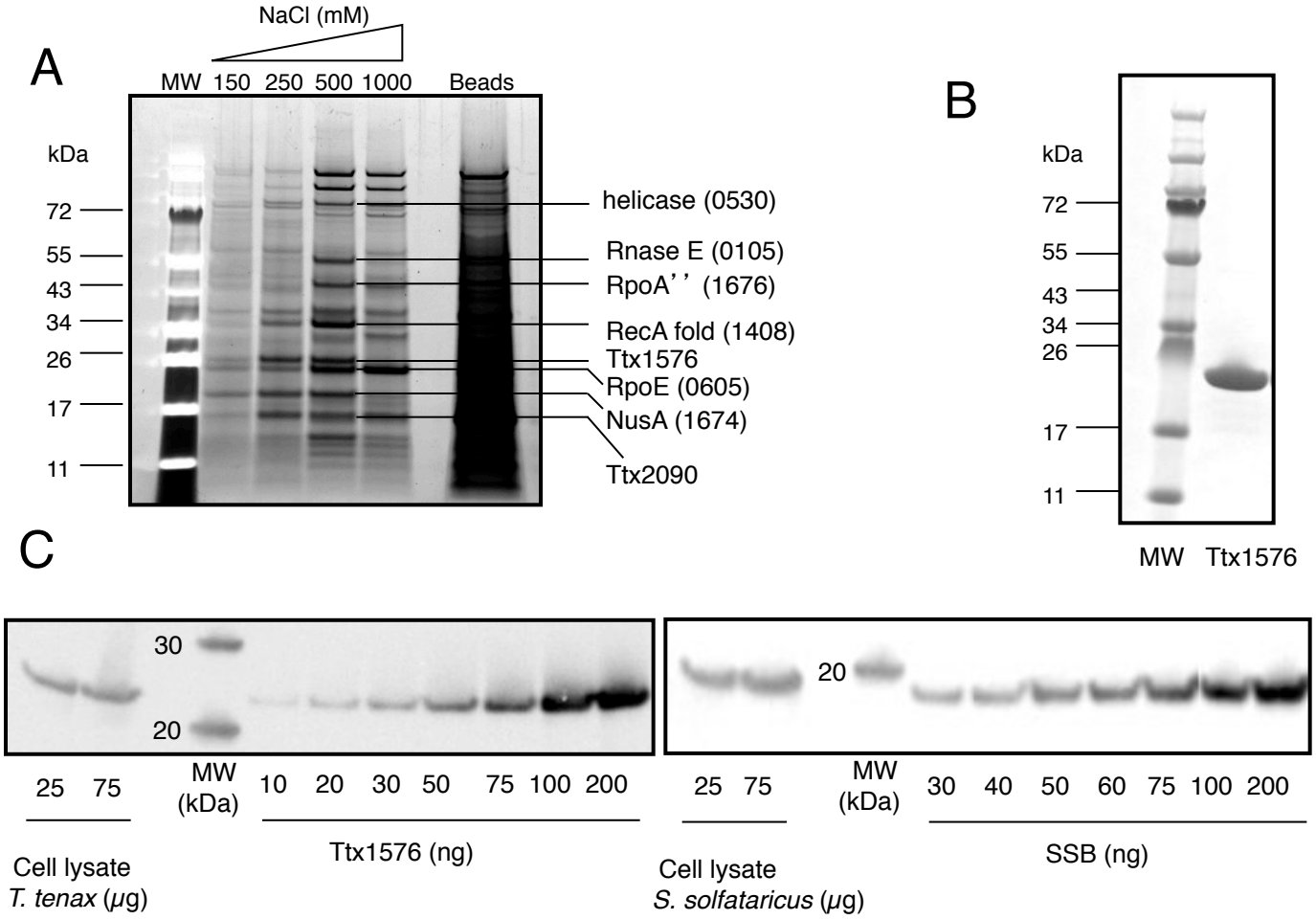
1. Mushegian AR & Koonin EV (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc. Natl. Acad. Sci. USA 93(19):10268-10273.
2. Sun S & Shamoo Y (2003) Biochemical characterization of interactions between DNA polymerase and single-stranded DNA-binding protein in bacteriophage RB69. J. Biol. Chem. 278(6):3876-3881.
3. Kowalczykowski SC, Lonberg N, Newport JW, & von Hippel PH (1981) Interactions of bacteriophage T4-coded gene 32 protein with nucleic acids. I. Characterization of the binding interactions. J. Mol. Biol. 145(1):75-104.
4. Wold MS (1997) Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. Annu. Rev. Biochem. 66:61-92.
5. Meyer RR & Laine PS (1990) The single-stranded DNA-binding protein of *Escherichia coli.* Microbiol Rev 54(4):342-380.
6. Iftode C, Daniely Y, & Borowiec JA (1999) Replication protein A (RPA): the eukaryotic SSB. Crit Rev Biochem Mol Biol 34(3):141-180.
7. Murzin AG (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. EMBO J. 12(3):861-867.
8. Suck D (1997) Common fold, common function, common origin? Nat. Struct. Biol. 4(3):161-165.

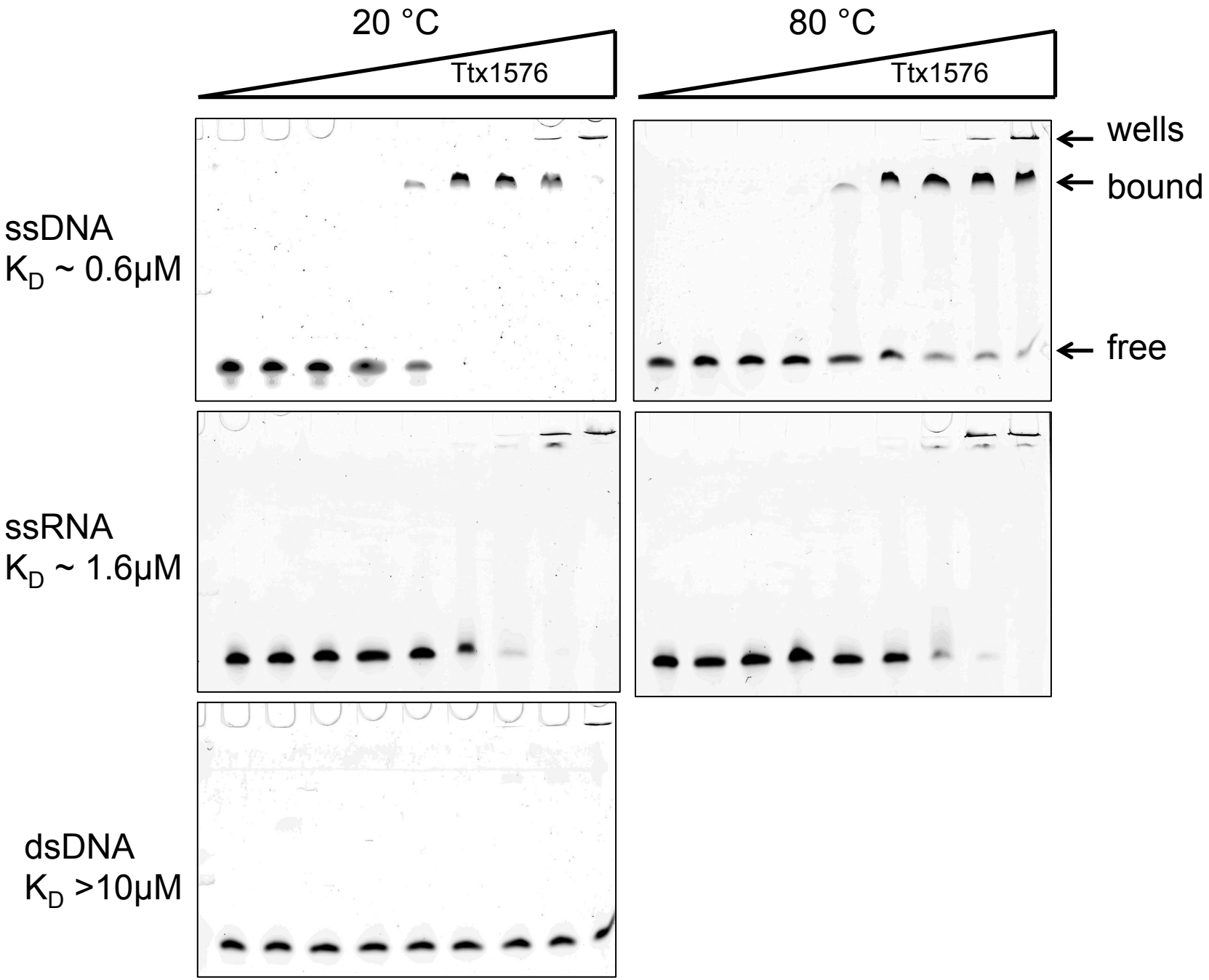
9. Lohman TM & Ferrari ME (1994) Escherichia coli single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities. Annual Review Of Biochemistry 63:527-570.
10. Raghunathan S, Kozlov AG, Lohman TM, & Waksman G (2000) Structure of the DNA binding domain of E. coli SSB bound to ssDNA. Nat. Struct. Biol. 7(8):648-652.
11. Bernstein DA*, et al.* (2004) Crystal structure of the Deinococcus radiodurans single-stranded DNA-binding protein suggests a mechanism for coping with DNA damage. Proc. Natl. Acad. Sci. USA 101(23):8575-8580.
12. Dabrowski S*, et al.* (2002) Identification and characterization of single-stranded-DNA-binding proteins from Thermus thermophilus and Thermus aquaticus - new arrangement of binding domains. Microbiology 148(Pt 10):3307-3315.
13. Sugiman-Marangos S & Junop MS (2010) The structure of DdrB from Deinococcus: a new fold for single-stranded DNA binding proteins. Nuc. Acids Res. 38(10):3432-3440.
14. Norais CA, Chitteni-Pattu S, Wood EA, Inman RB, & Cox MM (2009) DdrB protein, an alternative Deinococcus radiodurans SSB induced by ionizing radiation. J. Biol. Chem. 284(32):21402-21411.
15. Bochkarev A, Bochkareva E, Frappier L, & Edwards AM (1999) The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding. EMBO J. 18(16):4498-4504.
16. Bochkarev A, Pfuetzner RA, Edwards AM, & Frappier L (1997) Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. Nature 385(6612):176-181.
17. Richard DJ*, et al.* (2008) Single-stranded DNA-binding protein hSSB1 is critical for genomic stability. Nature 453:677-682.
18. White MF (2003) Archaeal DNA repair: paradigms and puzzles. Biochem Soc Trans 31(3):690-693.
19. Kelly TJ, Simancek P, & Brush GS (1998) Identification and characterization of a single-stranded DNA-binding protein from the archaeon Methanococcus jannaschii. Proc. Natl. Acad. Sci. U S A 95(25):14634-14639.
20. Chedin F, Seitz EM, & Kowalczykowski SC (1998) Novel homologs of replication protein A in archaea: implications for the evolution of ssDNA-binding proteins. Trends Biochem. Sci. 23(8):273-277.

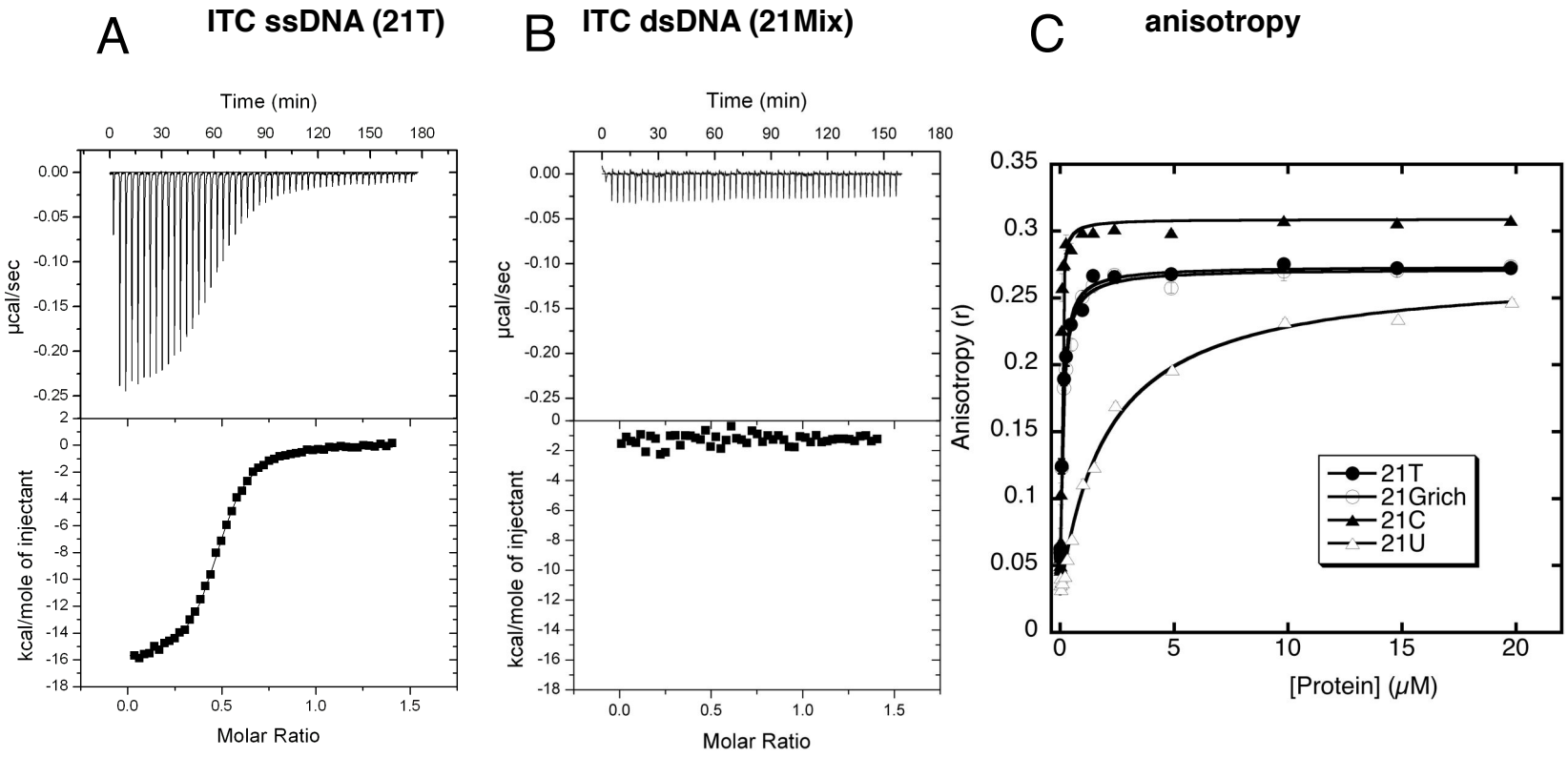
21. Komori K & Ishino Y (2001) Replication protein A in Pyrococcus furiosus is involved in homologous DNA recombination. J. Biol. Chem. 7:7.
22. Kelman Z, Pietrokovski S, & Hurwitz J (1999) Isolation and characterization of a split B-type DNA polymerase from the archaeon Methanobacterium thermoautotrophicum deltaH. J. Biol. Chem. 274(40):28751-28761.
23. Wadsworth RI & White MF (2001) Identification and properties of the crenarchaeal single-stranded DNA binding protein from Sulfolobus solfataricus. Nuc. Acids Res. 29(4):914-920.
24. Kerr ID*, et al.* (2003) Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of Sulfolobus SSB. EMBO J. 22(11):2561-2570.
25. Luo X*, et al.* (2007) CC1, a novel crenarchaeal DNA binding protein. J Bacteriol 189(2):403-409.
26. Seroussi E & Lavi S (1993) Replication protein A is the major single-stranded DNA binding protein detected in mammalian cell extracts by gel retardation assays and UV cross-linking of long and short single-stranded DNA molecules. J. Biol. Chem. 268(10):7147-7154.
27. Soding J, Biegert A, & Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nuc. Acids Res. 33(Web Server issue):W244-248.
28. Makarova KS, Sorokin AV, Novichkov PS, Wolf YI, & Koonin EV (2007) Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. Biology Direct 2:33.
29. Kernchen U & Lipps G (2006) Thermodynamic analysis of the single-stranded DNA binding activity of the archaeal replication protein A (RPA) from Sulfolobus solfataricus. Biochemistry 45(2):594-603.
30. Wernimont A & Edwards A (2009) In situ proteolysis to generate crystals for structure determination: an update. PLoS ONE 4(4):e5094.
31. Kumarevel T*, et al.* (2004) Crystal structure of activated HutP; an RNA binding protein that regulates transcription of the hut operon in Bacillus subtilis. Structure 12(7):1269-1280.
32. Nissen P, Hansen J, Ban N, Moore PB, & Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis. Science 289(5481):920-930.
33. Marchler-Bauer A*, et al.* (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nuc. Acids Res. 39(Database issue):D225-229.

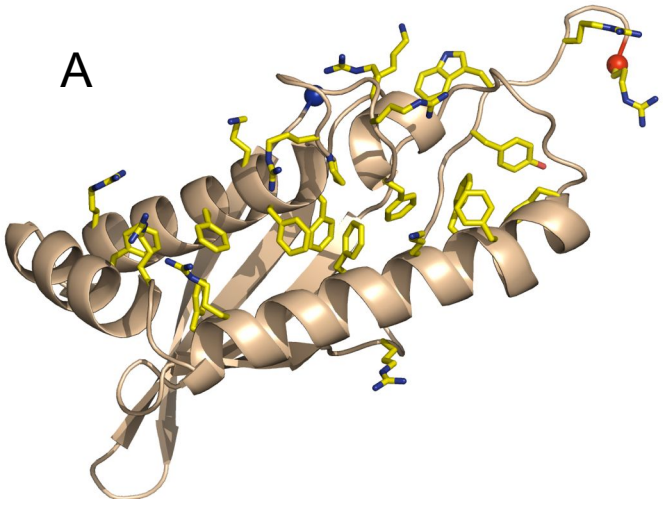
34. Koonin EV, Mushegian AR, & Bork P (1996) Non-orthologous gene displacement. Trends in Genetics 12(9):334-336.
35. Leipe DD, Aravind L, & Koonin EV (1999) Did DNA replication evolve twice independently? Nuc. Acids Res. 27(17):3389-3401.
36. Siebers B*, et al.* (2011) The complete genome sequence of Thermoproteus tenax: a physiologically versatile member of the crenarchaeota. PLoS ONE 6(10):e24222.
37. Liu H & Naismith JH (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol 8:91.
38. Oke M*, et al.* (2010) The Scottish Structural Proteomics Facility: targets, methods and outputs. J Struct Funct Genomics 11:167-180.
39. Reid SL, Parry D, Liu HH, & Connolly BA (2001) Binding and recognition of GATATC target sequences by the EcoRV restriction endonuclease: a study using fluorescent oligonucleotides and fluorescence polarization. Biochemistry 40(8):2484-2494.
40. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Macromolecular Crystallography, part A, Methods in Enzymology, eds Carter CWJ & Sweet RM (Academic Press, New York), Vol 276, pp 307-326.
41. Sheldrick GM (2008) A short history of SHELX. Acta Crystallogr A 64(Pt 1):112-122.
42. Bailey S (1994) The CCP4 suite: Programs for Protein Crystallography. Acta Crystallographica Section D 50:760-763.
43. McCoy AJ*, et al.* (2007) Phaser crystallographic software. Journal of Applied Crystallography 40(Pt 4):658-674.
44. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallographica Section D-Biological Crystallography 53:240-255.
45. Winn MD, Isupov MN, & Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallographica Section D-Biological Crystallography 57:122-133.
46. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallographica Section D-Biological Crystallography 60:2126-2132.

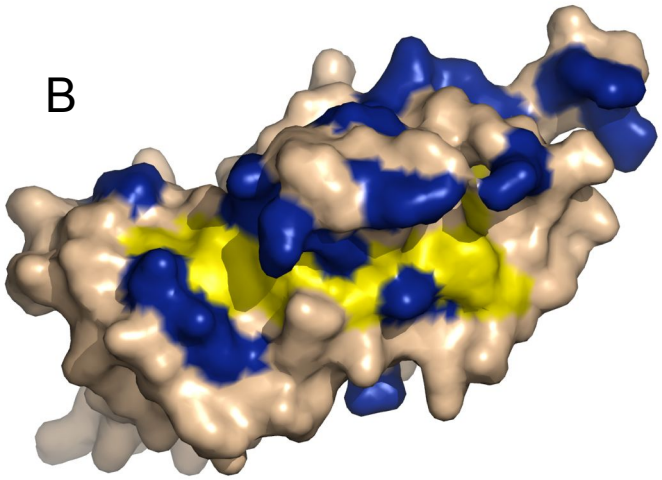
47. Davis IW*, et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucl. Acids Res. 35(Web Server issue):W375-383.
48. Whitmore L & Wallace BA (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucl. Acids Res. 32(Web Server issue):W668-673.
49. Andrews P (1965) The gel-filtration behaviour of proteins related to their molecular weights over a wide range. Biochem. J. 96(3):595-606.
50. Altschul SF*, et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nuc. Acids Res. 25(17):3389-3402.
51. Marchler-Bauer A*, et al.* (2009) CDD: specific functional annotation with the Conserved Domain Database. Nuc. Acids Res. 37(Database issue):D205-210.
52. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nuc. Acids Res. 32(5):1792-1797.
53. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, & Barton GJ (1998) JPred: a consensus secondary structure prediction server. Bioinformatics 14(10):892-893.

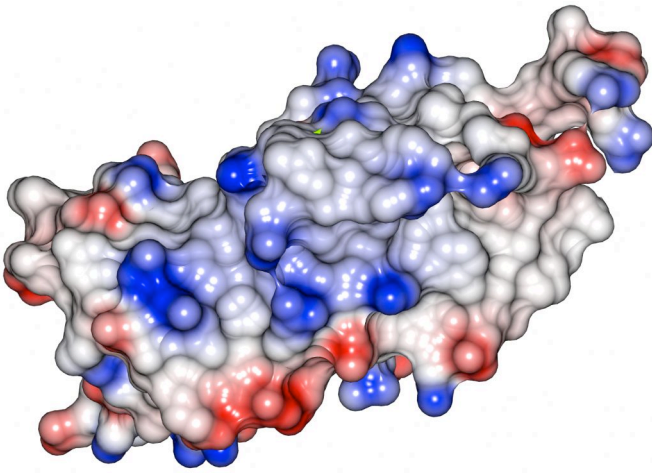

A
NaCl (mM)
MW 150 250 500 1000 Beads
kDa
72
55
43
34
26
17
11
helicase (0530)
Rnase E (0105)
RpoA'' (1676)
RecA fold (1408)
Ttx1576
RpoE (0605)
NusA (1674)
Ttx2090
B
kDa
72
55
43
34
26
17
11
MW Ttx1576
C
30
20
25 75 MW (kDa) 10 20 30 50 75 100 200
Cell lysate
T. tenax (μg)
Ttx1576 (ng)
20
25 75 MW (kDa) 30 40 50 60 75 100 200
Cell lysate
S. solfataricus (μg)
SSB (ng)

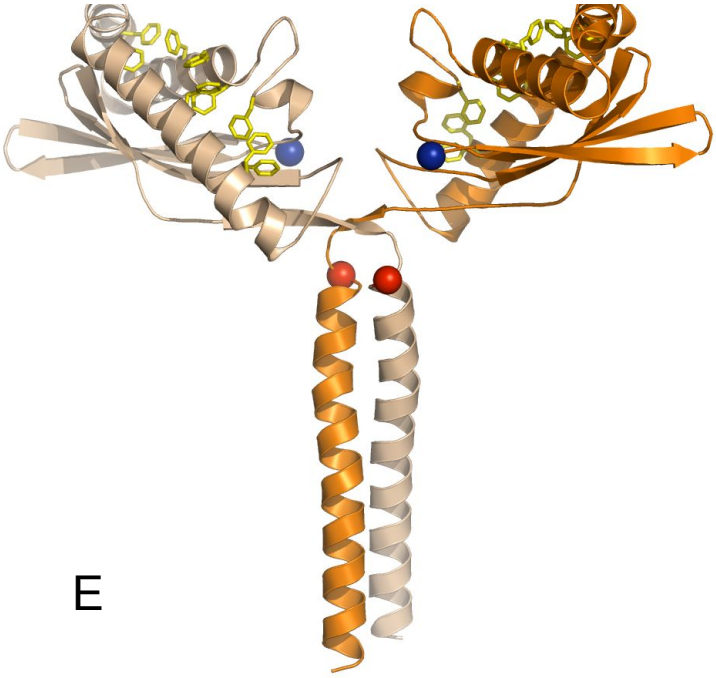

20 °C
Ttx1576
80 °C
Ttx1576
wells
bound
free
ssDNA
K_D ~ 0.6μM
ssRNA
K_D ~ 1.6μM
dsDNA
K_D >10μM


A
ITC ssDNA (21T)
B
ITC dsDNA (21Mix)
C
anisotropy
Time (min)
µcal/sec
kcal/mole of injectant
Molar Ratio
Anisotropy (r)
[Protein] (µM)
21T
21Grich
21C
21U



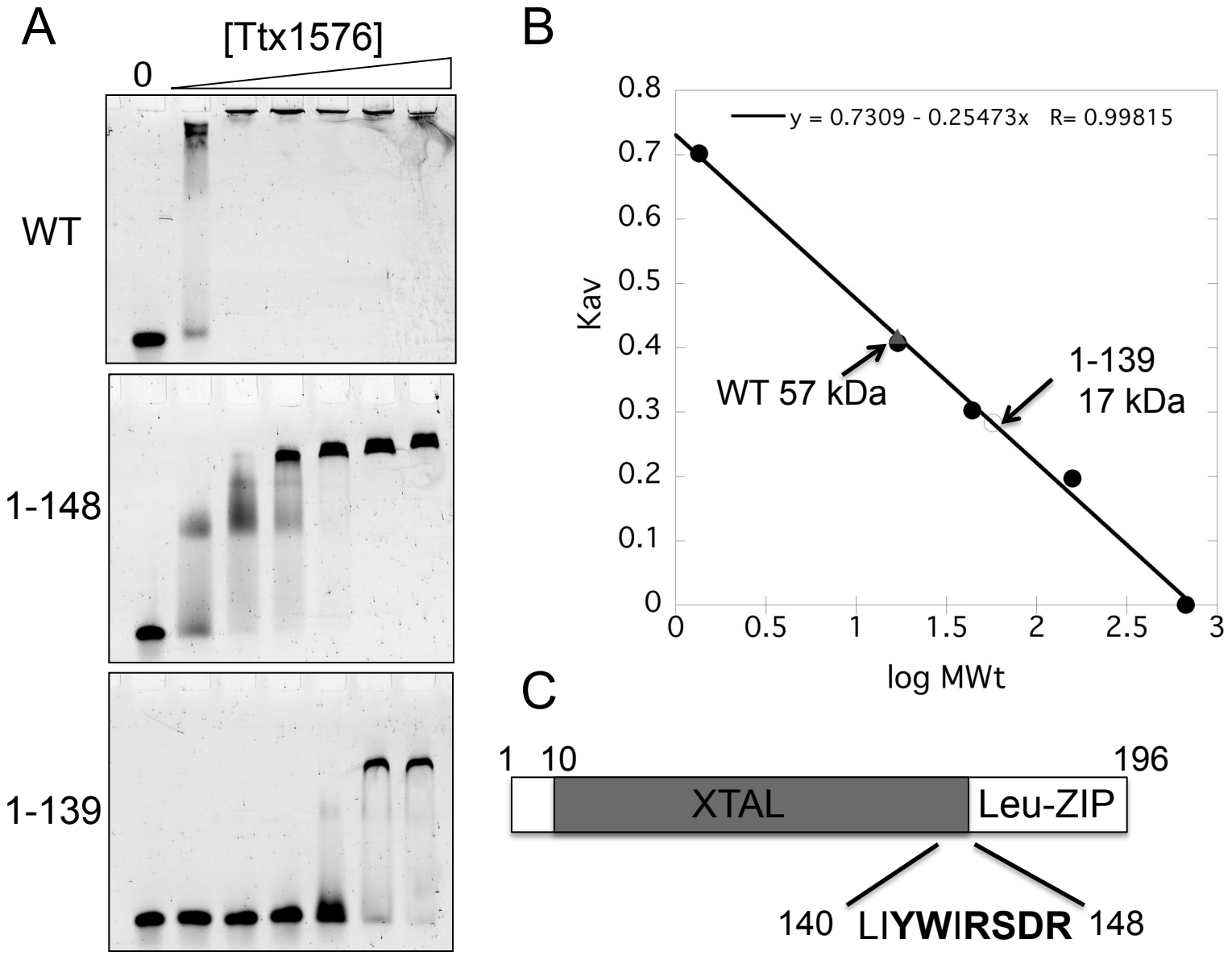






D

```
BZIP HMM     EELEKEKSELAQEVEKLKEEVARLARERDALKEKLEKLA
MATCH         +L +E+ +LA+EVE LK+E+ RL+RE +  K+KL++++
TTX1576 154 AKLREERDRLAKEVELLKAENERLRRELEDVKRKLQQIT 192
```


A
[Ttx1576]
0
WT
1-148
1-139
B
y = 0.7309 - 0.25473x R= 0.99815
Kav
log MWt
WT 57 kDa
1-139
17 kDa
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.5
1
1.5
2
2.5
3
C
1 10
196
XTAL
Leu-ZIP
140 LIYWIRSDR 148

Supporting information - Text

Supplementary Figure 1:

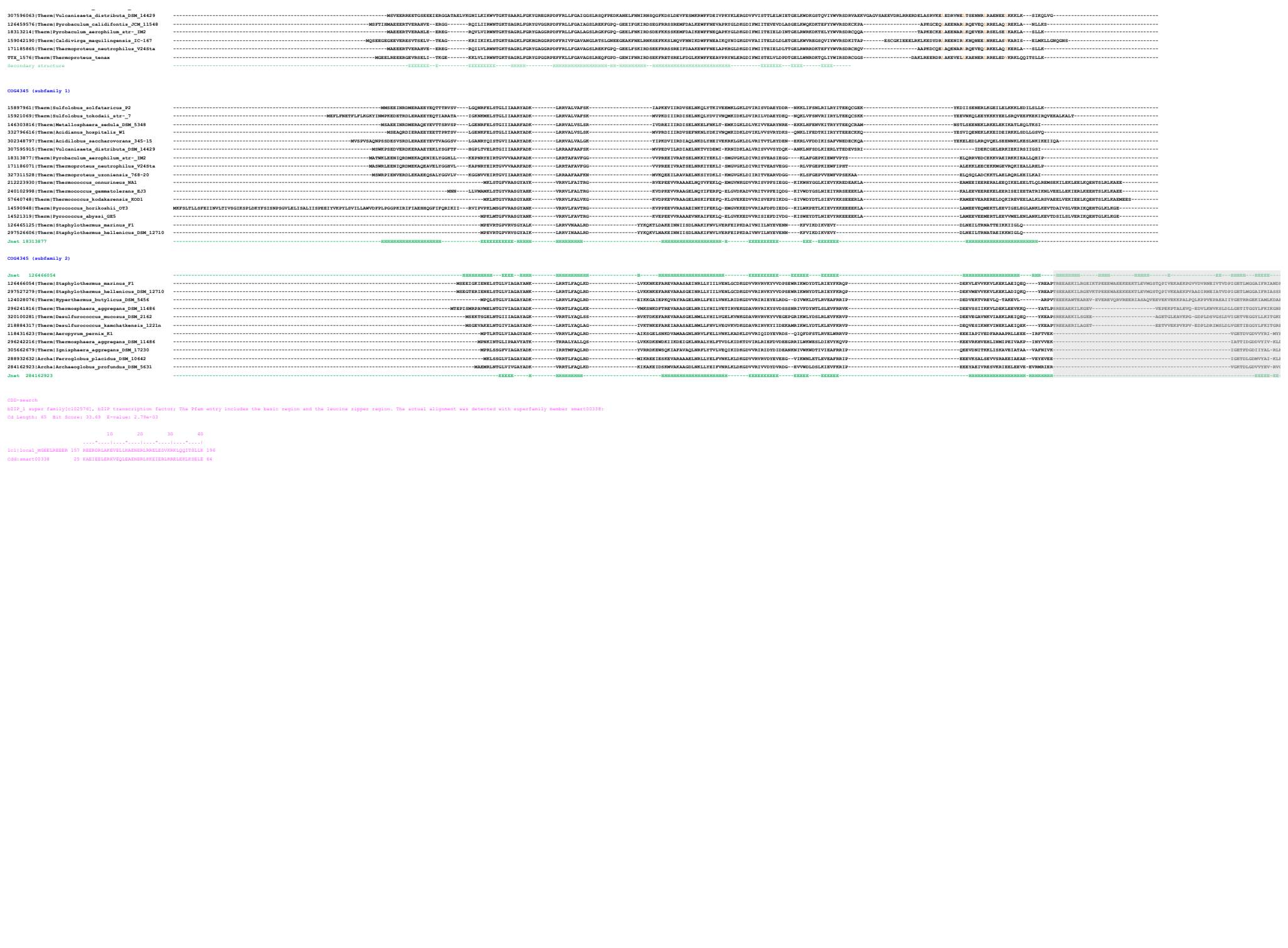
Multiple sequence alignment of full length sequences of TTX_1576 (arCOG05578) and arCOG03772 family proteins. Secondary structure prediction (or actual in case of TTX_1576 protein) is colored green as follows: 'H' indicates α-helix and 'E' indicates extended conformation (β-strand). The sequences are denoted by their GI numbers and species names. For TTX_1576 family multiple alignment with several sequences from bZIP superfamily (Maf family) is also shown; the conservation of leucines in the domain is shown by # symbol and the corresponding positions colored orange. Unaligned regions, specific for the corresponding subfamily are shaded.

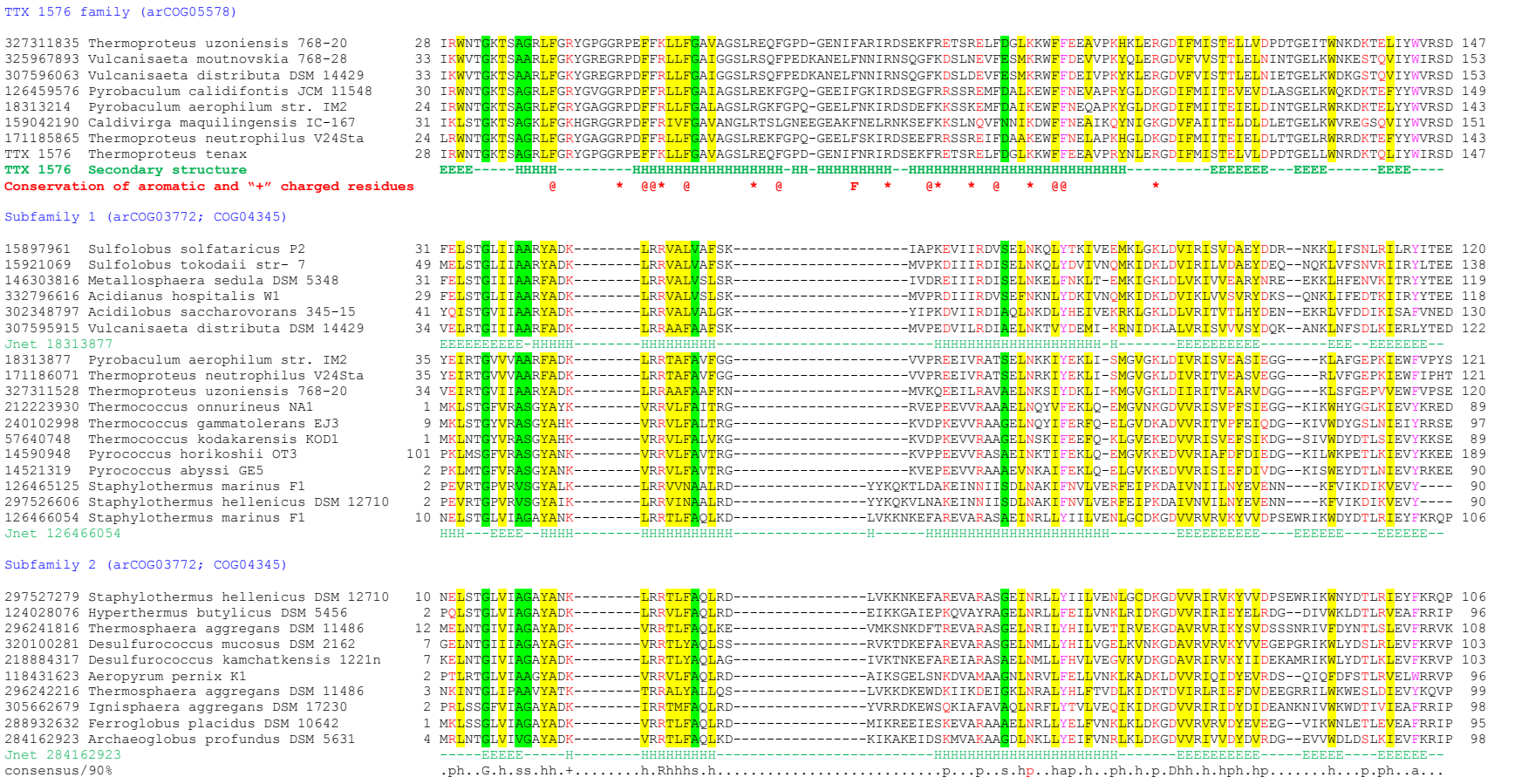
Supplementary Figure 2:

Multiple sequence alignment of the common core domain of TTX_1576 (arCOG05578) and arCOG03772 family proteins. Secondary structure prediction (or actual in case of TTX_1576 protein) is colored green as follows: 'H' indicates α-helix and 'E' indicates extended conformation (β-strand). The sequences are denoted by their GI numbers and species names. The coloring is based on the consensus shown underneath the alignment; 'h' indicates hydrophobic residues (WFYMLIVACTH), 'p' indicates polar residues (EDKRNQHTS),'s' indicates small residues (ACDGNPSTV), 'a' – aromatic residues (YWF). Location of aromatic core (@) for interaction with the nucleobases of ssDNA and a positively charged periphery (*) identified in TTX_1576 structure are shown by red beneath TTX 1576 family alignment.

Supplementary Dataset 1

Distribution of Ttx1576-like and canonical SSB protein-encoding genes in archaeal genomes. The Ttx1576 family (arCOG 05578) is present only in the thermoproteales highlighted in yellow. These organisms lack a canonical SSB-family gene, which is present in all other genomes. The Ttx1576 related protein (arCOG 03772) described in the accompanying paper is present in a wider range of archaea. Standard abbreviations are used for the archaeal genomes.


COG4345 (subfamily 1)
COG4345 (subfamily 2)
CDD-search
bZIP_1 super family(cl02576), bZIP transcription factor; The Pfam entry includes the basic region and the leucine zipper region. The actual alignment was detected with superfamily member smart00338:
Cd Length: 65 Bit Score: 33.69 E-value: 2.79e-03
10 20 30 40
lcl|local_MGEELREEER 157 REERDRLAKEVELLKAENERLRRELEDVKRKLQQITSLLK 196
Cdd:smart00338 25 KAETEELERKVEQLEAENERLKKEIERLRRELEKLKSELE 64


TTX 1576 family (arCOG05578)
327311835 Thermoproteus uzoniensis 768-20 28 IRWNTGKTSAGRLFGRYGPGGRPEFFKLLFGAVAGSLREQFGPD-GENIFARIRDSEKFRETSRELFDGLKKWFFEEAVPKHKLERGDIFMISTELLVDPDTGEITWNKDKTELIYWVRSD 147
325967893 Vulcanisaeta moutnovskia 768-28 33 IKWVTGKTSAARLFGKYGREGRPDFFRLLFGAIGGSLRSQFPEDKANELFNNIRNSQGFKDSLNEVFESMKRWFFDEVVPKYQLERGDVFVVSTTLELNINTGELKWNKESTQVIYWIRSD 153
307596063 Vulcanisaeta distributa DSM 14429 33 IKWVTGKTSAARLFGKYGREGRPDFFRLLFGAIGGSLRSQFPEDKANELFNNIRNSQGFKDSLDEVFESMKRWFFDEIVPKYKLERGDVFVISTTLELNIETGELKWDKGSTQVIYWVRSD 153
126459576 Pyrobaculum calidifontis JCM 11548 30 IRWNTGKTSAGRLFGRYGVGGRPDFFRLLFGAIAGSLREKFGPQ-GEEIFGKIRDSEGFRRSSREMFDALKEWFFNEVAPRYGLDKGDIFMIITEVEVDLASGELKWQKDKTEFYYWVRSD 149
18313214 Pyrobaculum aerophilum str. IM2 24 IRWNTGKTSAGRLFGRYGAGGRPDFFRLLFGALAGSLRGKFGPQ-GEELFNKIRDSDEFKKSSKEMFDAIKEWFFNEQAPKYGLDKGDIFMIITEIELDINTGELRWRKDKTELYYWVRSD 143
159042190 Caldivirga maquilingensis IC-167 31 IKLSTGKTSAGKLFGKHGRGGRPDFFRIVFGAVANGLRTSLGNEEGEAKFNELRNKSEFKKSLNQVFNNIKDWFFNEAIKQYNIGKGDVFAIITELDLDLETGELKWVREGSQVIYWVRSD 151
171185865 Thermoproteus neutrophilus V24Sta 24 LRWNTGKTSAGRLFGRYGAGGRPDFFRLLFGAVAGSLREKFGPQ-GEELFSKIRDSEEFRRSSREIFDAAKEWFFNELAPKHGLDKGDIFMIITEIELDLTTGELRWRRDKTEFYYWVRSD 143
TTX 1576 Thermoproteus tenax 28 IRWNTGKTSAGRLFGRYGPGGRPEFFKLLFGAVAGSLREQFGPD-GENIFNRIRDSEKFRETSRELFDGLKKWFFEEAVPRYNLERGDIFMISTELVLDPDTGELLWNRDKTQLIYWIRSD 147
TTX 1576 Secondary structure EEEE-----HHHHH---------HHHHHHHHHHHHHHHHHH-HH-HHHHHHHHH--HHHHHHHHHHHHHHHHHHHHHHHHH----------EEEEEEE---EEEE------EEEE----
Conservation of aromatic and "+" charged residues @ * @@* @ * @ F * @* * @ * @@ *
Subfamily 1 (arCOG03772; COG04345)
15897961 Sulfolobus solfataricus P2 31 FELSTGLIIAARYADK--------LRRVALVAFSK--------------------IAPKEVIIRDVSELNKQLYTKIVEEMKLGKLDVIRISVDAEYDDR--NKKLIFSNLRILRYITEE 120
15921069 Sulfolobus tokodaii str- 7 49 MELSTGLIIAARYADK--------LRRVALVAFSK--------------------MVPKDIIIRDISELNKQLYDVIVNQMKIDKLDVIRILVDAEYDEQ--NQKLVFSNVRIIRYLTEE 138
146303816 Metallosphaera sedula DSM 5348 31 FELSTGIIIAARFADK--------LRRVALVSLSR--------------------IVDREIIIRDISELNKELFNKLT-EMKIGKLDLVKIVVEARYNRE--EKKLHFENVKITRYYTEE 119
332796616 Acidianus hospitalis W1 29 FELSTGLIIAARYADK--------LRRVALVSLSK--------------------MVPRDIIIRDVSEFNKNLYDKIVNQMKIDKLDVIKLVVSVRYDKS--QNKLIFEDTKIIRYYTEE 118
302348797 Acidilobus saccharovorans 345-15 41 YQISTGVIIAARYADK--------LRRVALVALGK--------------------YIPKDVIIRDIAQLNKDLYHEIVEKRKLGKLDLVRITVTLHYDEN--EKRLVFDDIKISAFVNED 130
307595915 Vulcanisaeta distributa DSM 14429 34 VELRTGIIIAARFADK--------LRRAAFAAFSK--------------------MVPEDVILRDIAELNKTVYDEMI-KRNIDKLALVRISVVVSYDQK--ANKLNFSDLKIERLYTED 122
Jnet 18313877 EEEEEEEEEE-HHHHH--------HHHHHHHHH-------------------------HHHHHHHHHHHHHHHHHHH-H--------EEEEEEEEE--------EEE--EEEEEEE--
18313877 Pyrobaculum aerophilum str. IM2 35 YEIRTGVVVAARFADK--------LRRTAFAVFGG--------------------VVPREEIVRATSELNKKIYEKLI-SMGVGKLDIVRISVEASIEGG----KLAFGEPKIEWFVPYS 121
171186071 Thermoproteus neutrophilus V24Sta 35 YEIRTGVVVAARFADK--------LRRTAFAVFGG--------------------VVPREEIVRATSELNRKIYEKLI-SMGVGKLDIVRITVEASVEGG----RLVFGEPKIEWFIPHT 121
327311528 Thermoproteus uzoniensis 768-20 34 VEIRTGVIIAARYADK--------LRRAAFAAFKN--------------------MVKQEEILRAVAELNKSIYDKLI-KMGVGKLDIIRITVEARVDGG----KLSFGEPVVEWFVPSE 120
212223930 Thermococcus onnurineus NA1 1 MKLSTGFVRASGYAYK--------VRRVLFAITRG--------------------RVEPEEVVRAAAELNQYVFEKLQ-EMGVNKGDVVRISVPFSIEGG--KIKWHYGGLKIEVYKRED 89
240102998 Thermococcus gammatolerans EJ3 9 MKLSTGYVRASGYAHK--------VRRVLFALTRG--------------------KVDPKEVVRAAGELNQYIFERFQ-ELGVDKADVVRITVPFEIQDG--KIVWDYGSLNIEIYRRSE 97
57640748 Thermococcus kodakarensis KOD1 1 MKLNTGYVRASGYAHK--------VRRVLFALVKG--------------------KVDPKEVVRAAGELNSKIFEEFQ-KLGVEKEDVVRISVEFSIKDG--SIVWDYDTLSIEVYKKSE 89
14590948 Pyrococcus horikoshii OT3 101 PKLMSGFVRASGYANK--------VRRVLFAVTRG--------------------KVPPEEVVRASAELNKTIFEKLQ-EMGVKKEDVVRIAFDFDIEDG--KILWKPETLKIEVYKKEE 189
14521319 Pyrococcus abyssi GE5 2 PKLMTGFVRASGYANK--------VRRVLFAVTRG--------------------KVEPEEVVRAAAEVNKAIFEKLQ-ELGVKKEDVVRISIEFDIVDG--KISWEYDTLNIEVYRKEE 90
126465125 Staphylothermus marinus F1 2 PEVRTGFVRVSGYALK--------LRRVVNAALRD----------------YYKQKTLDAKEINNIISDLNAKIFNVLVERFEIPKDAIVNIILNYEVENN----KFVIKDIKVEVY---- 90
297526606 Staphylothermus hellenicus DSM 12710 2 PEVRTGPVRVSGYAIK--------LRRVINAALRD----------------YYKQKVLNAKEINNIISDLNAKIFNVLVERFEIPKDAIVNILNYEVENN----KFVIKDIKVEVY---- 90
126466054 Staphylothermus marinus F1 10 NELSTGLVIAGAYANK--------LRRTLFAQLKD----------------LVKKNKEFAREVARASAEINRLLYIILVENLGCDKGDVVRVRVKYVVDPSEWRIKWDYDTLRIEYFKRQP 106
Jnet 126466054 HHH---EEEE--HHHH-------HHHHHHHHHH----------------H------HHHHHHHHHHHHHHHHHHHHH-------EEEEEEEEE----EEEEE----EEEEE--
Subfamily 2 (arCOG03772; COG04345)
297527279 Staphylothermus hellenicus DSM 12710 10 NELSTGLVIAGAYANK--------LRRTLFAQLRD----------------LVKKNKEFAREVARASGEINRLLYIILVENLGCDKGDVVRIRVKYVVDPSEWRIKWNYDTLRIEYFKRQP 106
124028076 Hyperthermus butylicus DSM 5456 2 PQLSTGLVIAGAYADK--------LRRVLFAQLRD----------------EIKKGAIEPKQVAYRAGELNRLLFEILVNKLRIDKGDVVRIRIEYELRDG--DIVWKLDTLRVEAFRRIP 96
296241816 Thermosphaera aggregans DSM 11486 12 MELNTGIVIAGAYADK--------VRRTLFAQLKE----------------VMKSNKDFTREVARASGELNRILYHILVETIRVEKGDAVRVRIKYSVDSSSNRIVFDYNTLSLEVFRRVK 108
320100281 Desulfurococcus mucosus DSM 2162 7 GELNTGIIIAGAYAGK--------VRRTLYAQLSS----------------RVKTDKEFAREVARASGELNMLLYHILVGELKVNKGDAVRVRVKYVVEGEPGRIKWLYDSLRLEVFKRVP 103
218884317 Desulfurococcus kamchatkensis 1221n 7 KELNTGIVIAGAYADK--------LRRTLYAQLAG----------------IVKTNKEFAREIARASAELNMLLFHVLVEGVKVDKGDAVRIRVKYIIDEKAMRIKWLYDTLKLEVFKRVP 103
118431623 Aeropyrum pernix K1 2 PTLRTGLVIAAGYADK--------VRRVLFAQLRD----------------AIKSGELSNKDVAMAAGNLNRVLFELLVNKLKADKLDVVRIQIDYEVRDS--QIQFDFSTLRVELWRRVP 96
296242216 Thermosphaera aggregans DSM 11486 3 NKINTGLIPAAVYATK--------TRRALYALLQS----------------LVKKDKEWDKIIKDEIGKLNRALYHLFTVDLKIDKTDVIRLRIEFDVDEEGRRILWKWESLDIEVYKQVP 99
305662679 Ignisphaera aggregans DSM 17230 2 PRLSSGFVIAGAYADK--------IRRTMFAQLRD----------------YVRRDKEWSQKIAFAVAQLNRFLYTVLVEQIKIDKGDVVRIRIDYDIDEANKNIVWKWDTIVIEAFRRIP 98
288932632 Ferroglobus placidus DSM 10642 1 MKLSSGLVIAGAYADK--------VRRTLFAQLRD----------------MIKREEIESKEVARAAAELNRLLYELFVNKLKLDKGDVVRVRVDYEVEEG--VIKWNLETLEVEAFRRIP 95
284162923 Archaeoglobus profundus DSM 5631 4 MRLNTGLVIVGAYADK--------VRRTLFAQLKD----------------KIKAKEIDSKMVAKAAGDLNKLLYEIFVNRLKLDKGDVVRIVVDYDVRDG--EVVWDLDSLKIEVFKRIP 98
Jnet 284162923 -----EEEEE-----H--------HHHHHHHHH-------------------------HHHHHHHHHHHHHHHHHHHH-------EEEEEEEEE-----EEEEE----EEEEE--
consensus/90% .ph..G.h.ss.hh.+........h.Rhhhs.h.........................p...p..s.hp..hap.h..ph.h.p.Dhh.h.hph.hp.......h...p.ph..a...