Supplementary Information

Entrapment of DNA in an intersubunit tunnel system of a single-stranded DNA-binding protein

Homa Ghalei[1,2,†,#], Holger von Moeller[1,a,#], Detlef Eppers[1], Daniel Sohmen[3], Daniel N. Wilson[3], Bernhard Loll[1], and Markus C. Wahl[1,*]

[1] Freie Universität Berlin, Fachbereich Biologie, Chemie, Pharmazie, Institut für Chemie und Biochemie, AG Strukturbiochemie, Takustr. 6, 14195 Berlin, Germany.

[2] Max-Planck-Institute for Biophysical Chemistry, Department of Cellular Biochemistry, Am Fassberg 11, 37077 Göttingen, Germany.

[3] Gene Center, Department of Biochemistry and Center of integrated Protein Science Munich (CiPSM), University of Munich, Feodor-Lynenstr. 25, 81377 Munich, Germany

[†] Current address: Department of Cancer Biology, The Scripps Research Institute, Jupiter, FL 33458, USA.

[a] Current address: moloX GmbH, Takustr. 6, 14195 Berlin, Germany

[#] These authors contributed equally to this work.

[*] Corresponding author:
Email: mwahl@zedat.fu-berlin.de (M.C.W.)

Supplementary Tables

Supplementary Table S1. DNA and RNA oligonucleotides of mixed sequence used in binding assays

Nucleic acid	Sequence
$ssDNA^+$	5'-d(ATGTGGAAAATCTCTAGCAGT)-3'
$ssDNA^-$	5'-d(ACTGCTAGAGATTTTCCACAT)-3'
$ssRNA^+$	5'-r(AUGUGGAAAAUCUCUAGCAGU)-3'
$ssRNA^-$	5'-r(ACUGCUAGAGAUUUUCCACAU)-3'
45-mer ssDNA	5'-d(CTTGCTAGGACGGATCCCTCGAGGTTTTTTTTTTTTTTTTTTTT)-3'
mixed 20-mer ssDNA	5'-d(CATGGTCAGTTAGCAGGTTC)-3'
sR2 sRNA	5'-r(GGGGGAUGAUGAGUUUUUCCCUCACUCUGA UUAGUGAUGAGGAGCCGAUGCACUGACCUC)-3'
sR12 sRNA	5'-r(GGAGGGGAUGAUGAGCGUUUACCGGUCUGAGUUGUGAUGAUA CUGGCACUGUCUGACCUUCC)-3'

Supplementary Table S2. Diffraction data collection, phasing and refinement statistics

	***pfu*ThermoDBP-RP1**				***ape*ThermoDBP-RP2**	***ape*ThermoDBP-RP2-dT$_{10}$**
Data Collection	**Form 1**			**Form 2**		
Wavelength (Å)	0.97962 (peak)	0.98003 (inflection)	0.97185 (remote)	1.0385	0.98000	0.91841
Space Group	$P6_4$	$P6_4$	$P6_4$	$C222_1$	$P2_12_12_1$	$P2_12_12_1$
Unit Cell Parameters (Å)						
a	136.6	136.6	136.6	61.3	61.2	61.2
b	136.6	136.6	136.6	199.3	108.6	108.6
c	122.1	122.1	122.1	114.9	154.9	154.9
Resolution (Å)[a]	35.00 - 3.75 (3.85 - 3.75)	35.00 - 3.75 (3.85 - 3.75)	35.00 - 3.50 (3.59 - 3.50)	35.00 - 2.43 (2.49 - 2.43)	35.00 - 2.05 (2.10 - 2.05)	30.00 - 2.90[f] (3.00 - 2.90)
Reflections						
Unique [a]	26034 (1914)	26103 (1940)	32116 (2387)	150600 (10942)	125145 (9242)	46414 (4433)
Completeness (%)[a]	99.6 (99.6)	99.6 (99.5)	99.7 (99.6)	99.3 (98.9)	99.9 (100.0)	99.8 (99.9)
Redundancy [a]	5.9 (5.9)	5.9 (5.9)	5.9 (5.9)	5.6 (5.7)	7.4 (5.9)	4.5 (4.6)
$I/\sigma(I)$ [a]	10.8 (2.9)	11.1 (2.7)	16.1 (2.9)	11.5 (4.2)	15.5 (3.8)	11.07 (1.23)
R_{meas} [a,b]	0.149 (0.665)	0.152 (0.722)	0.094 (0.659)	0.105 (0.670)	0.090 (0.614)	0.194 (1.725)
Phasing						
Heavy Atoms	24				12	
FOM[c]	0.41					
Phasing Power[c]						
(isomorphous	0.0	0.591	1.102			
anomalous)	0.792	0.098	1.190			
R_{cullis} [c]	0.792	0.976	0.792			
Overall CC[d]					0.168	
Pseudo free CC[d]					0.631	
CC_{all}/CC_{weak} [a, d]					0.278/0.165	
Refinement						
Resolution (Å) [a]	34.16 - 3.50 (3.59 - 3.50)			33.41 - 2.43 (2.52 - 2.43)	32.73 - 2.05 (2.08 - 2.05)	30.00 - 2.90[f] (3.00 - 2.90)
Reflections						
Number [a]	15568 (1148)			26694 (2612)	63473 (2581)	46426 (2541)
Completeness (%)[a]	100.0 (100.0)			99.0 (99.0)	96.7 (93.0)	99.9 (100.0)
Test Set (%)[a]	5.0			5.0	5.0	5.0
R_{work} [a, e]	20.7 (27.3)			22.3 (26.7)	17.9 (24.2)	26.7 (37.5)
R_{free} [a,e]	26.7 (35.0)			27.6 (36.5)	21.6 (29.8)	31.1 (40.6)

Refined Residues						
Protein	590			586	881	1754
Water molecules	31			68	289	38
Ligands molecules	3 sulfate			4 sulfate	2 glycerol, 6 imidazole	1 phosphate
DNA						4 strands
Mean B-Factors (Å^2)						
Wilson	69.5			37.5	28.8	56.2
Protein	69.4			61.3	37.4	59.3
Water molecules	39.9			46.6	40.8	44.72
Ligand molecules	164.0			112.6	53.3	88.7
DNA						98.2
Ramachandran Plot[g] (%)						
Favored	97.42			98.96	99.31	97.60
Outliers	0.17			0.0	0.0	0.0
R.m.s.d. Geometry						
Bond Lengths (Å)	0.009			0.005	0.008	0.004
Bond Angles (°)	1.158			0.992	1.082	0.909

[a] Data for the highest resolution shell in parentheses

[b] $R_{meas} = \Sigma_h [n/(n-1)]^{1/2} \Sigma_i \mid I_h - I_{h,i} \mid / \Sigma_h \Sigma_i I_{h,i}$.where I_h is the mean intensity of symmetry-equivalent reflections and n is the redundancy.

[c] Figure of merit (FOM), phasing power and R_{cullis} as output by SHARP (1)

[d] Correlation coefficients as output by ShelxD/E (2,3)

[e] $R = \Sigma_{hkl} \mid |F_{obs}| - |F_{calc}| \mid / \Sigma_{hkl} |F_{obs}|$; R_{work} – hkl $\notin$ T; R_{free} – hkl $\in$ T; T – test set

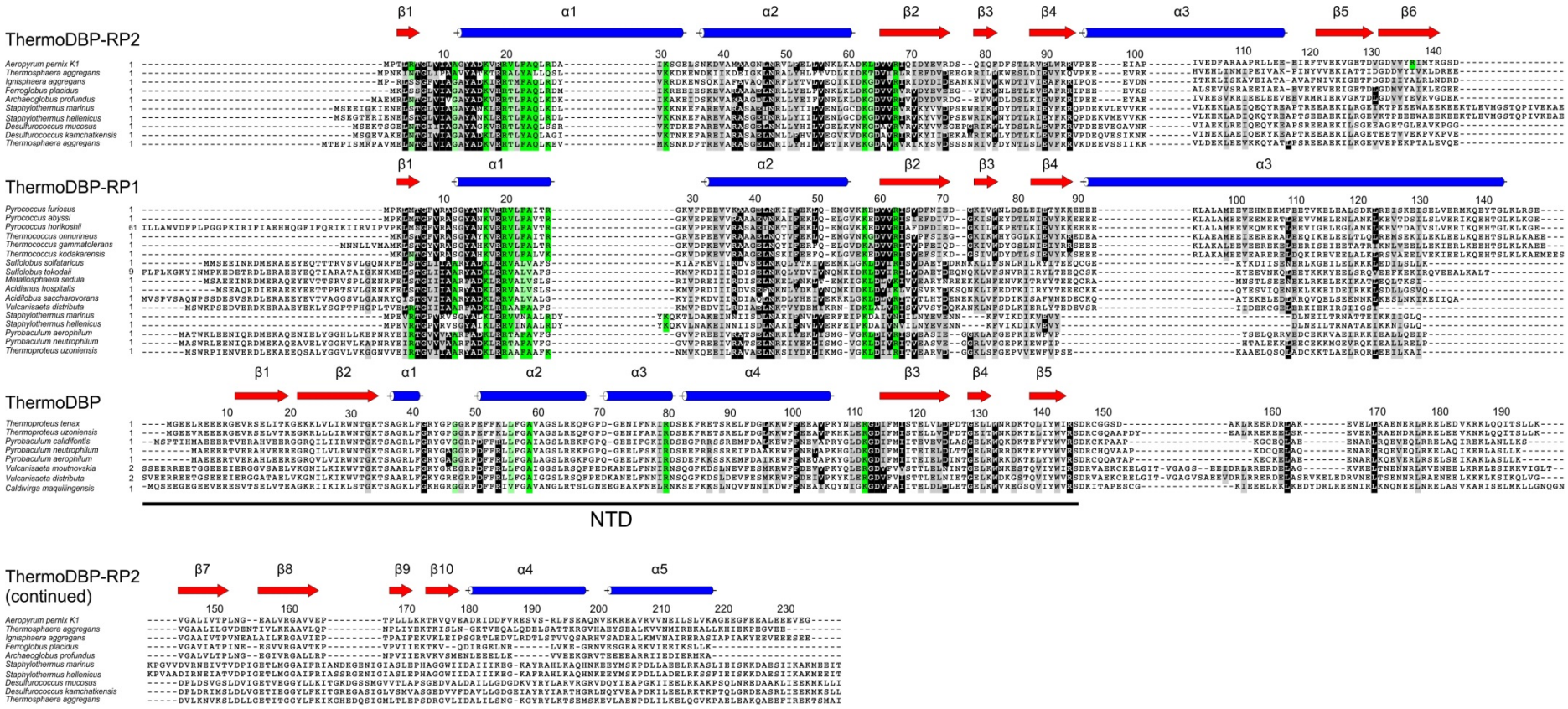
[f] observed anisotropy with resolution limits along a* = 2.6 Å, b* = 2.9 Å and c* = 3.1 Å

[g] Calculated with MolProbity (4)

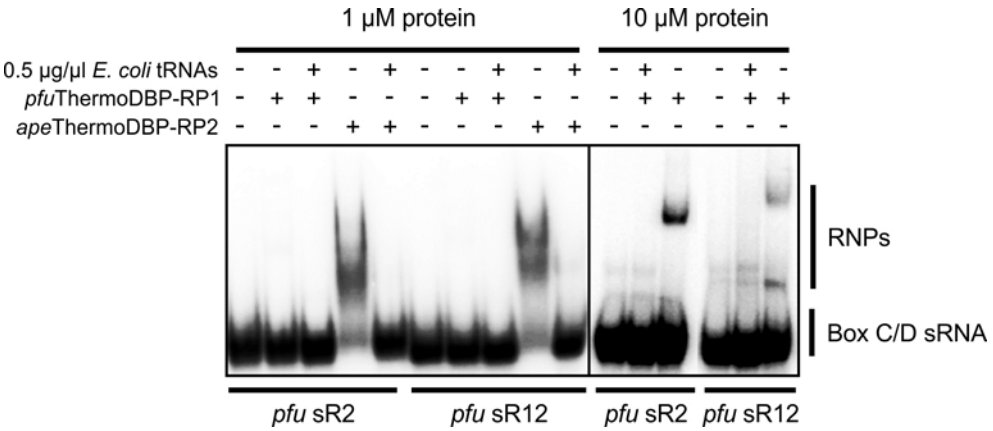
Supplementary Table S3. List of protein-DNA contacts

	DNA1	DNA2
Sugars:	Asn211^{B} - $^{O4'}$dT1	Arg27$^{B'}$ - $^{O3'}$dT11′
	Asn211^{B} - ^{sugar}dT1	Val21$^{A'}$ - ^{sugar}dT7′
	Val207^{B} - ^{sugar}dT2	Arg204$^{A'}$ - ^{O4}dT10′
	Leu64$^{A'}$ - ^{sugar}dT4	
	Ala12$^{A''}$ - ^{sugar}dT4	
	Arg27^{A} – $^{O3'}$dT5	
	Lys31^{A} – ^{sugar}dT5	
	Arg20^{A} – ^{sugar}dT6	
	Val21^{A} – ^{sugar}dT7	
	Ala24^{B} - ^{sugar}dT9	
	Arg20$^{B'}$ – $^{O3'}$dT10	
Phosphates:	Arg137^{B} - ^{O1}dT2	Arg68$^{B'}$ - ^{O1}dT6′
	Lys203^{B} - ^{O1}dT3	Arg27$^{A'}$ – ^{O1}dT6′
	Arg68^{B} - ^{O1}dT5	Lys17$^{A'}$ - ^{O1}dT8′
	Arg27^{A} - ^{O2}dT6	
	Lys17^{A} - ^{O1}dT8	
Base:	Asn211^{B} - ^{base}dT1	Arg204$^{A'}$ - ^{O2}dT10′
	Arg204^{B} - ^{O2}dT6	Phe23$^{A'}$ - ^{base}dT6′
	Phe23^{A} - ^{base}dT6	Ala24^{A} - ^{base}dT6
	Ala24^{A} - ^{base}dT7	Gln25$^{B'}$ - $^{O2, N3}$dT8′
	Gln25^{A} - ^{base}dT8	Arg5$^{B'}$ -$^{O4, base}$dT6′
	Arg5^{B} - ^{O2}dT8	Lys17$^{B'}$ - $^{O2, N3}$dT9′
	Lys17^{B} - $^{N3, O2}$dT9	Val21$^{B'}$ - ^{base}dT7′
	Val21^{B} - ^{base}dT9	

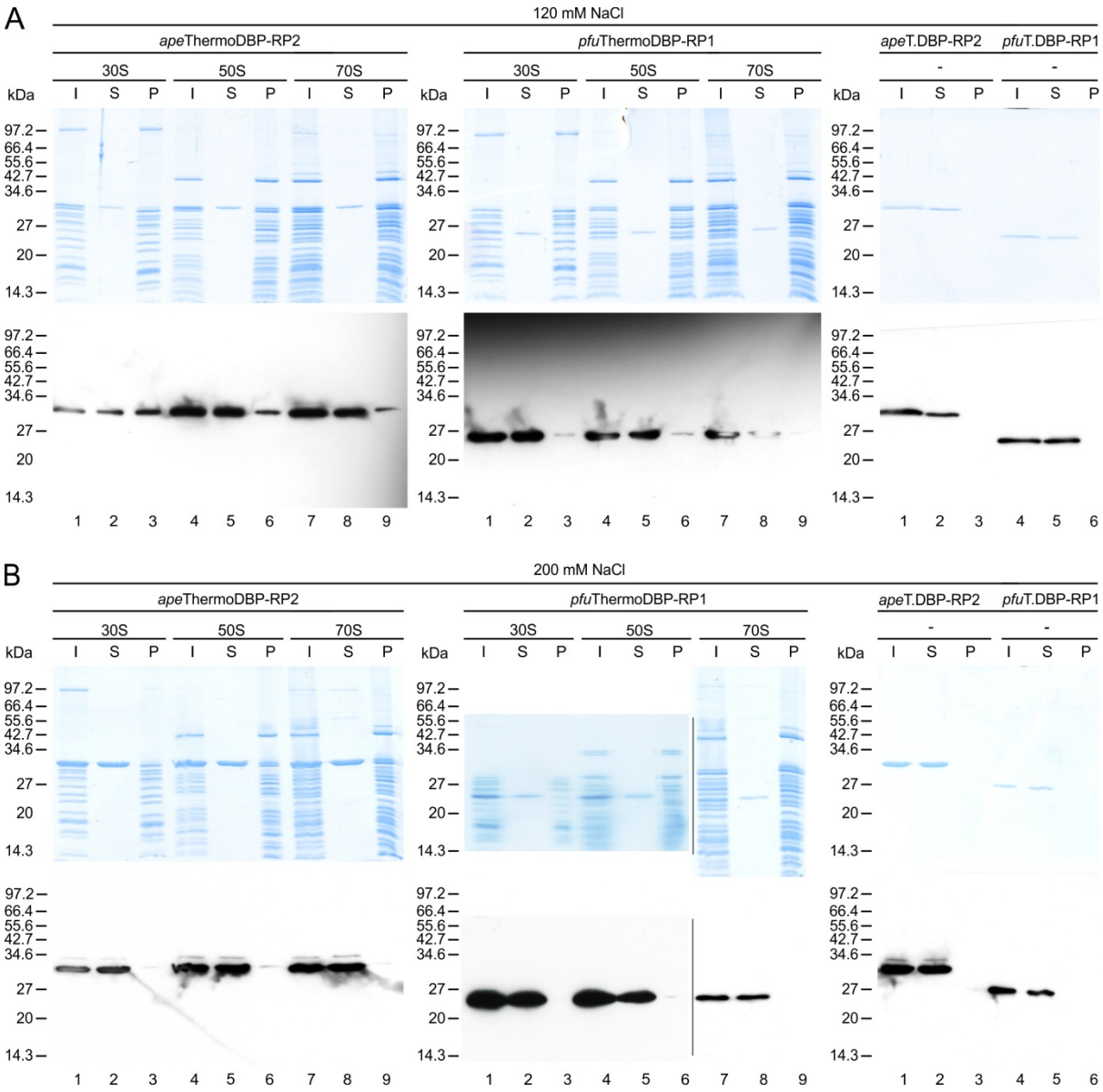
Supplementary Figures



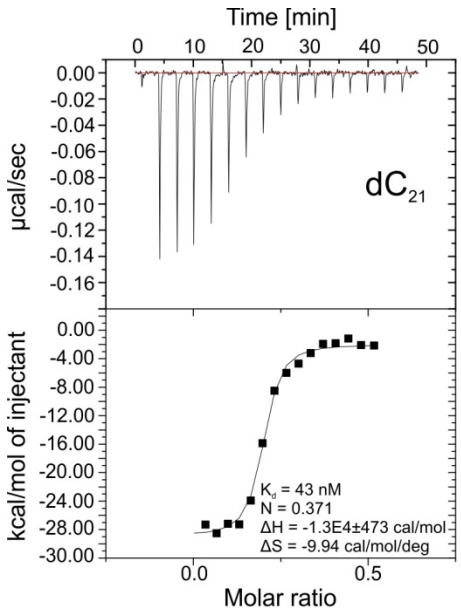
Supplementary Figure S1. Multiple sequence alignment. Multiple structure-based sequence alignment of the proteins containing the DUF2258 domain. Secondary structure elements of *ape*ThermoDBP-RP2, *pfu*ThermoDBP-RP1 and *tte*ThermoDBP are shown (α-helices – blue; β strands – red). Darker background indicates higher conservation with respect to *ape*ThermoDBP-RP2. Residue with green background contact DNA in the *ape*ThermoDBP-RP2–dT_{10} structure. Regions corresponding to the NTDs are indicated. Proteins listed: *Aeropyrum pernix* - APE_1866.1, *Ignisphaera aggregans* - IGAG-0238, *Thermosphaera aggregans* - TAGG_0476, *Desulfurococcus fermentans* - DESFE_1124, *Desulfurococcus kamchatkensis* - DKAM_1006, *Desulfurococcus mucosus* - DESMU_0072, *Thermosphaera aggregans* - TAGG_0069, *Staphylothermus marinus* - SMAR_1160, SMAR_0213, *Staphylothermus hellenicus* - SHELL_1311, SHELL_0606, *Thermogladius cellulolyticus* - TCELL_0577, *Hyperthermus butylicus* - HBUT_1217, *Pyrolobus fumarii* - PYRFU_0856, *Acidilobus saccharovorans* - ASAC_0853, ASAC_0999, *Vulcanisaeta moutnovskia* - VMUT_0250, *Pyrobaculum aerophilum* - PAE3173, *Pyrobaculum sp. 1860* - P186_1371, *Pyrobaculum arsenaticum* - PARS_1656, *Pyrobaculum islandicum* - PISL_0538, *Thermoproteus neutrophilus* - TNEU_1621, *Pyrobaculum calidifontis* - PCAL_1811, *Thermoproteus uzoniensis* - TUZN_1647, *Vulcanisaeta distributa* - VDIS_1809, *Sulfolobus islandicus* - LS215_1226, *Sulfolobus islandicus* - M1425_1129, *Sulfolobus solfataricus* - SSO1098, SSOL_2070, Ssol98_010100004655, *Sulfolobus tokodaii* - ST0853, *Sulfolobus acidocaldarius* - SACI_1216, *Acidianus hospitalis* - Ahos_0932, *Metallosphaera sedula* - MSED_1043, *Metallosphaera yellowstonensis* - METMK1DRAFT_00029730, *Archaeoglobus profundus* - ARCPR_1830, *Ferroglobus placidus* - FERP_2293, *Thermococcus sp. AM4* - TAM4_671, *Thermococcus gammatolerans* - TGAM_0941, *Thermococcus zilligii* - TZILA_05110, *Thermococcus kodakarensis* - TK0813, *Thermococcus onnurineus* - TON_0781, *Thermococcus sibiricus* - TSIB_0836, *Thermococcus litoralis* - OCC_01989, *Thermococcus barophilus* - TERMP_00659, *Pyrococcus yayanosii* - PYCH_06080, *Pyrococcus horikoshii* - PH1118, *Pyrococcus abyssi* - PAB1631, *Pyrococcus furiosus* - PF1044

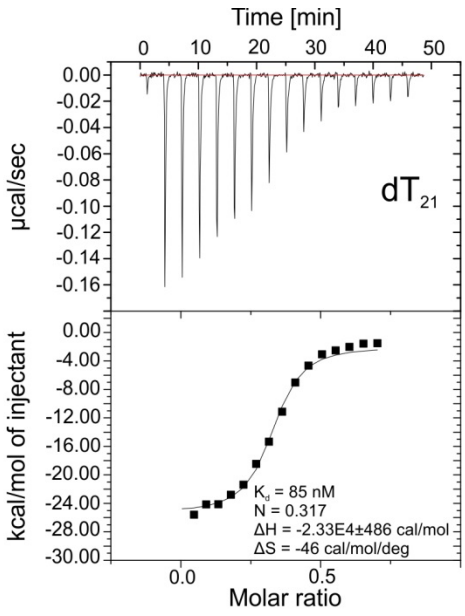


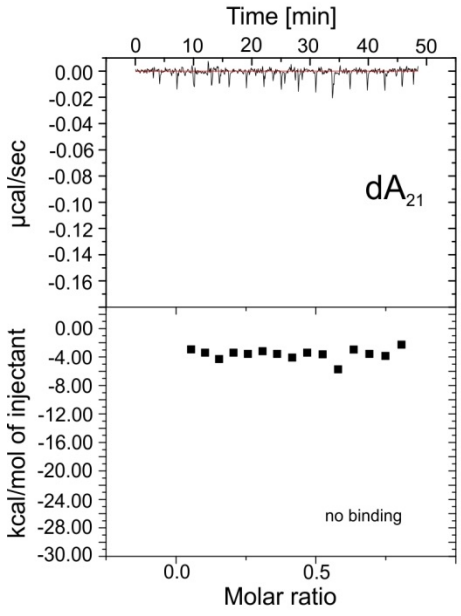
Supplementary Figure S2. Binding of ThermoDBP-RPs to box C/D sRNAs. Electrophoretic gel mobility shift assays testing the binding of *pfu*ThermoDBP-RP1 and of *ape*ThermoDBP-RP2 to sR2 and sR12 box C/D sRNAs from *Pyrococcus furiosus*.

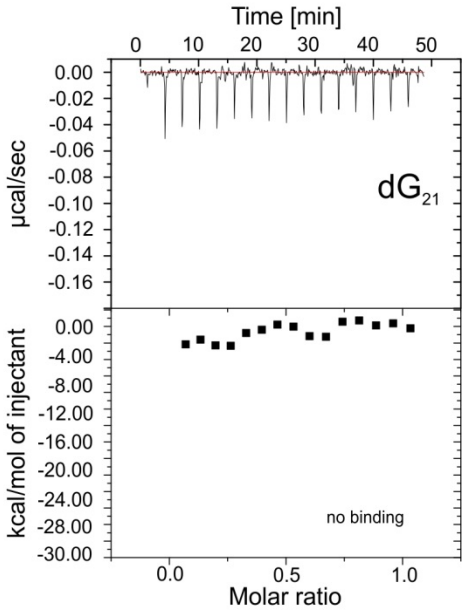


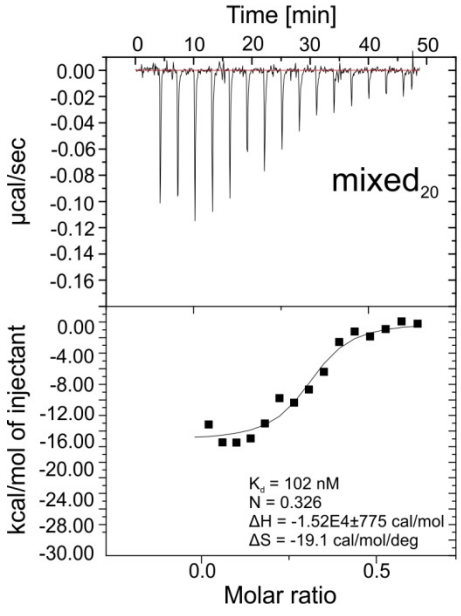
Supplementary Figure S3. Interactions of ThermoDBP-RPs with ribosomal subunits and ribosomes. Interactions were probed by co-sedimentation binding assays of ThermoDBP-RPs and 30S, 50S and 70S ribosomal particles of *Thermococcus kodakarensis*. Ribosomal particles were isolated from *T. kodakarensis* lysates as previously described (5) and incubated in the presence of 120 mM NaCl **(A)** or 200 mM NaCl **(B)** together with purified ThermoDBP-RPs from *A. pernix* and *P. furiosus* and subsequently separated by ultra-centrifugation. Upper panels: Proteins were separated by SDS-PAGE and stained with Coomassie Brilliant Blue. Lower panels: Proteins, after separation by SDS-PAGE, were transferred onto a nitrocellulose membrane for immuno-detection using an anti-His_6 antibody. Signals for proteins co-sedimented with ribosomal particles are detected in the pellet fractions (P). Whole-reaction-samples and samples after ribosomal particles separation are labeled (IN) and (SN), respectively.

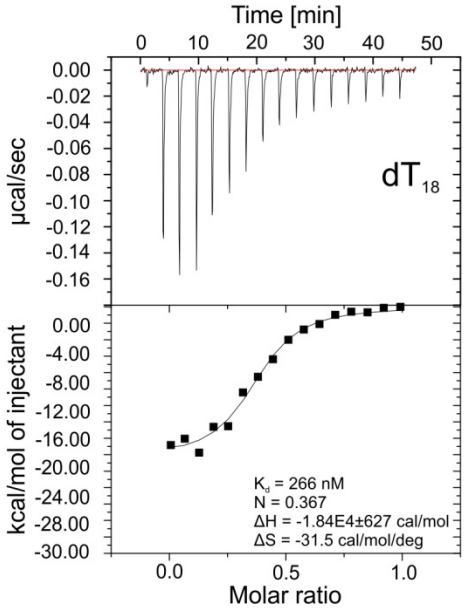

Time [min]
µcal/sec
dC21
kcal/mol of injectant
Kd = 43 nM
N = 0.371
ΔH = -1.3E4±473 cal/mol
ΔS = -9.94 cal/mol/deg
Molar ratio


Time [min]
µcal/sec
dT21
kcal/mol of injectant
Kd = 85 nM
N = 0.317
ΔH = -2.33E4±486 cal/mol
ΔS = -46 cal/mol/deg
Molar ratio

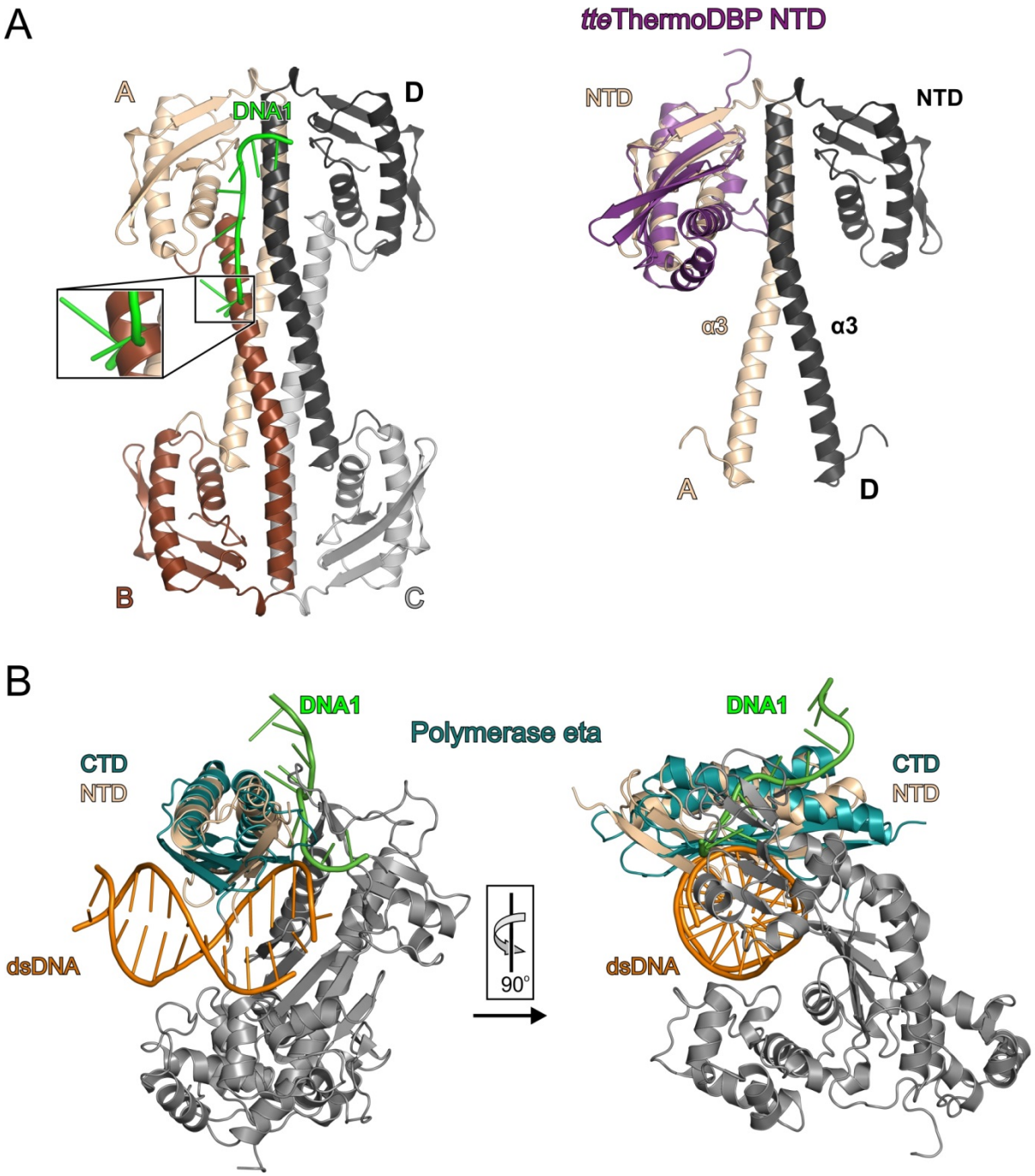

Time [min]
µcal/sec
dA21
kcal/mol of injectant
no binding
Molar ratio


Time [min]
µcal/sec
dG21
kcal/mol of injectant
no binding
Molar ratio

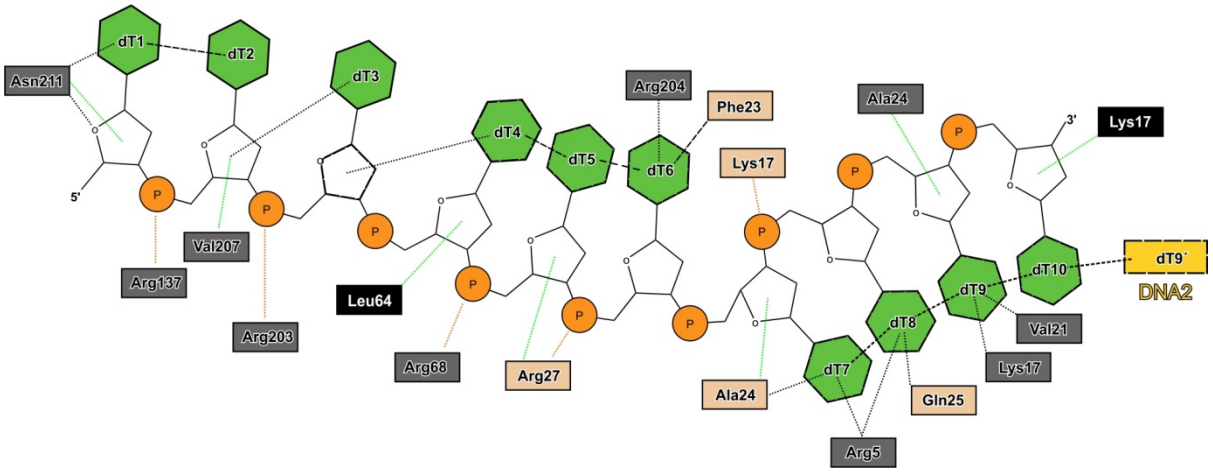

Time [min]
µcal/sec
mixed20
kcal/mol of injectant
Kd = 102 nM
N = 0.326
ΔH = -1.52E4±775 cal/mol
ΔS = -19.1 cal/mol/deg
Molar ratio


Time [min]
µcal/sec
dT18
kcal/mol of injectant
Kd = 266 nM
N = 0.367
ΔH = -1.84E4±627 cal/mol
ΔS = -31.5 cal/mol/deg
Molar ratio

Supplementary Figure S4. Quantification of *ape*ThermoDBP-RP2-ssDNA interactions by ITC. Both the raw data and the integrated data are shown. Oligonucleotide sequences are listed in Supplementary Table S3. Data were fitted based on the “One Set of Sites” model. The stoichiometry (N) gives the number of DNA molecules bound per *ape*ThermoDBP-RP2 monomer.



Supplementary Figure S5. Structural comparisons. **(A)** Cartoon view of *pfu*ThermoDBP-RP1 superposed with *tte*ThermoDBP NTD (left panel; *tte*ThermoDBP NTD – magenta; PDB ID 3TEK) and one ssDNA molecule from the *ape*ThermoDBP-RP2 co-crystal structure (right panel; DNA - green). Additional helices of *tte*ThermoDBP NTD do not clash with the helical extensions of *pfu*ThermoDBP-RP1 (left). However, a ssDNA bound on subunit A of a *pfu*ThermoDBP-RP1 tetramer in the same conformation as observed in the *ape*ThermoDBP-RP2 structure would clash with helix α3 of subunit B. **(B)** Superposition of yeast polymerase eta (grey; C-terminal domain – cyan; PDB ID 3OHB) with the NTD of *ape*ThermoDBP-RP2 (beige). A ssDNA molecule as observed in the *ape*ThermoDBP-RP2 structure is colored in green. dsDNA (orange) binds to a different site on the polymerase eta CTD, which is structurally related to the NTDs of ThermoDBP and ThermoDBP-RP proteins.



Supplementary Figure S6. Schematic representation of the DNA contacts. Schematic representation of the interactions of *ape*ThermoDBP-RP2 with one dT_{10} oligonucleotide. Color coding as in Figure 3B and C. Orange dashed lines indicate protein-phosphate contacts, green dashed lines are protein-sugar contacts and red (polar) and black (hydrophobic) dashed lines represent protein-nucleobase interactions. Black slashed lines represent stacking interactions.

Supplementary References

1. Vonrhein, C., Blanc, E., Roversi, P. and Bricogne, G. (2007) Automated structure solution with autoSHARP. *Methods Mol Biol*, **364**, 215-230.
2. Pape, T. and Schneider, T.R. (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SXELX programs. *J Appl Cryst*, **37**, 843-844.
3. Sheldrick, G.M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr*, **66**, 479-485.
4. Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **66**, 12-21.
5. Marquez, V., Frohlich, T., Armache, J.P., Sohmen, D., Donhofer, A., Mikolajka, A., Berninghausen, O., Thomm, M., Beckmann, R., Arnold, G.J. *et al.* (2011) Proteomic characterization of archaeal ribosomes reveals the presence of novel archaeal-specific ribosomal proteins. *J Mol Biol*, **405**, 1215-1232.