J. Mol. Biol. (1996) **264**, 1072–1084

JMB

The Crystal Structure of a Hyperthermophilic Archaeal TATA-box Binding Protein

Brian S. DeDecker[1], Ronan O'Brien[2], Patrick J. Fleming[1] James H. Geiger[1], Stephen P. Jackson[3] and Paul B. Sigler[1]*

[1]*Department of Molecular Biophysics and Biochemistry and the Howard Hughes Medical Institute*
[2]*Department of Chemistry Yale University, JWG 421 P.O. Box 208114, New Haven, CT 06511-8114, USA*
[3]*Wellcome/CRC Institute Tennis Court Road, and Department of Zoology, Cambridge University, Cambridge, CB2 1QR, UK*

**Corresponding author*

This study analyzes the three-dimensional structure of the TATA-box binding protein (TBP) from the hyperthermophilic archaea *Pyrococcus woesei*. The crystal structure of *P. woesei* TBP (*Pw*TBP) was solved at 2.2 Å by X-ray diffraction and as expected from sequence homology (36% to 41% identical to eukaryotic TBPs) its overall structure is very similar to eukaryotic TBPs. The thermal unfolding transition temperature of this protein was measured by differential scanning calorimetry to be 101°C, which is more than 40°C higher than that of yeast TBP. Preliminary titration calorimetry data show that the affinity of *Pw*TBP for its DNA target, unlike its eukaryotic counterparts, is enhanced by increasing the temperature and salt concentration. The structure reveals possible explanations for this thermostability and these unusual DNA binding properties. The crystal structure of this hyperthermostable protein was compared to its mesophilic homologs and analyzed for differences in the native structure that may contribute to thermostability. Differences found were: (1) a disulfide bond not found in mesophilic counterparts; (2) an increased number of surface electrostatic interactions; (3) more compact protein packing. The presumed DNA binding surface of *Pw*TBP, like its eukaryotic counterparts, is hydrophobic but the electrostatic profile surrounding the protein is relatively neutral compared to the asymmetric positive potential that surrounds eukaryotic TBPs. The total reliance on a hydrophobic interface with DNA may explain the enhanced affinity of *Pw*TBP for its DNA promoter at higher temperatures and increased salt concentration.

Keywords: Archaea; thermophile; structure; TBP; transcription

Introduction

The organism *Pyrococcus woesei* is a hyperthermophilic archaeal species found near deep-sea thermal vents and has an optimal growth temperature of 105°C. Supporting the assertion that the archaeal cell and the eukaryotic nucleus share a common ancestor, homologs of the general eukaryotic transcription factors, TBP (Forterre, 1996; Marsh *et al*., 1994; Qureshi *et al*., 1995a; Rowlands *et al*., 1994), and transcription factor II B (TFIIB) (Ghosh *et al*., unpublished; Qureshi *et al*., 1995a,b), have recently been cloned from *P. woesei* and other archaeal species (TBP sequences aligned in Figure 1). In a manner similar to eukaryotic Pol II transcription initiation, archaeal TBP binds upstream of transcription start sites at promoters containing A + T-rich sequences (Qureshi *et al*., 1995a; Rowlands *et al*., 1994). These DNA elements, called box A motifs, have a consensus sequence of $^{C}_{T}$TTA$^{A}_{T}$ANN (Zillig *et al*., 1993; Palmer & Daniels, 1995) that is similar to the eukaryotic TATA-box which has a consensus sequence of TATA$^{A}_{T}$A$^{A}_{T}$N (Bucher, 1990). The archaeal TFIIB-like factor, TFB, binds the archaeal TBP/DNA complex (Ghosh *et al*., unpublished results; Rowlands *et al*., 1994), and this TBP/TFB/DNA complex has been shown to be required for efficient and specific RNA transcription (Qureshi *et al*., 1995b), again similar to the homologous eukaryotic transcription system. In a further demonstration of the conservation between the archaeal and eukaryotic transcription systems, we have recently shown that archaeal TFB

Abbreviations used: TBP, TATA-box binding protein; TFIIB, transcription factor II B; TFB, archaeal TFIIB-like factor; TFIIA, transcription factor II A; RMSD, root mean squared difference.

0022–2836/96/501072–13 $25.00/0

can cross kingdom lines by binding eukaryotic TBP/TATA box complexes (Ghosh *et al*., unpublished).

The crystal structures of the conserved C-terminal domain of two eukaryotic TBPs are known, one from the plant *Arabidopsis thaliana* (*At* TBP; Nikolov *et al*., 1992) and the other from the yeast *Saccharomyces cerevisiae* (*Sc*TBPc, TBPc denotes the conserved C-terminal 180 residues of TBP; Chasman *et al*., 1993; J. H. Geiger, unpublished crystal structure refined to 2.6 Å). The structure of the eukaryotic complex of TBP bound to DNA (J. Kim *et al*., 1993; Y. Kim *et al*., 1993) and the ternary complexes of TBP-DNA associated separately with TFIIB (Nikolov *et al*., 1995) and transcription factor II A (TFIIA) (Geiger *et al*., 1996; Tan *et al*., 1996) have also been determined by crystallography. Here we report the crystal structure of an archaeal TBP from *P. woesei* at 2.2 Å resolution. The structure of this archaeal homolog of eukaryotic TBP offers insight into the evolution of the basal transcription machinery. While remarkably similar in its conformation to that of its eukaryotic counterparts, *Pw*TBP shows a more compact folding and a different electrostatic charge potential distribution. These features help explain this molecule's remarkable thermostability and altered DNA binding properties.

Results

Structure of *Pw*TBP

Soluble *Pw*TBP (191 amino acids) was expressed in *Escherichia coli*, purified to homogeneity and crystallized by the hanging drop method using ammonium sulfate as a precipitant. An electron-density map of *Pw*TBP was synthesized at 3.1 Å resolution with phases derived from isomorphous and anomalous differences of a mercurial derivative, and isomorphous differences from a crystal of selenomethionine substituted protein (Table 1). The current model is refined to an *R*-factor of 0.205 (free *R*-factor = 0.285), for data with $|F| > 2\sigma$ and Bragg spacings from 6.0 to 2.2 Å.

The conserved 180-residue core of TBP comprises a direct repeat sequence (Figure 1). The primary sequence of *Pw*TBP extends in both the N and C-terminal directions beyond this core. The region N-terminal to the core, which is conserved neither in length nor composition in eukaryotic TBPs, consists of four residues in *Pw*TBP. This extension is shorter than any reported for a eukaryotic TBP, which range in length from 18 residues in *At*TBP to 158 residues in human TBP (*Hs*TBP). The C terminus of *Pw*TBP contains an ''acidic tail'' of seven residues which is conserved in the three

Table 1. Crystallography statistics

A. *Phasing statistics*

Resolution limit (Å)	10.1	7.7	6.2	5.1	4.4	3.9	3.4	3.1	Total
Phasing power[a,b]									
$HgCl_2$ (iso)	2.74	2.88	3.54	3.02	2.14	1.94	1.90	—	2.42
$HgCl_2$ (anom)	0.56	0.62	0.60	0.48	0.39	0.31	0.24	—	0.39
Se-Met (iso)	0.32	0.42	0.67	0.64	0.41	0.35	0.31	0.29	0.38
Mean figure of merit[c]	0.61	0.62	0.58	0.46	0.39	0.28	0.14	0.07	0.27

B. *Data collection statistics*

	Source[d]	λ (Å)	R_{sym} (%)[e] Outer shell	R_{sym} (%)[e] Total	Complete (%) Outer shell	Complete (%) Total	I/σ overall	Average redundancy
			(2.3–2.2 Å)		(2.3–2.2 Å)			
Parent 1	NSLS X25 (Fuji IP)	0.95	23.7	7.9	98.6	98.8	16.9	3.2
			(3.2–3.1 Å)		(3.2–3.1 Å)			
Parent 2	Yale CSB (MacScience IP)	1.54	29.4	9.5	98.8	99.8	17.5	5.8
			(3.6–3.5 Å)		(3.6–3.5 Å)			
$HgCl_2$	Yale CSB (Xentronics multiwire)	1.54	27.8	9.5	58.6	90.8	9.5	—
			(2.6–2.5 Å)		(2.6–2.5 Å)			
Se-Met	CHESS A1 (CCD)	0.91	18.9	6.8	87.8	92.6	14.4	3.3

C. *Refinement results*

	Resolution (Å)	*R*-factor	Free *R*-factor (Brünger, 1992) (10% of data)	No. of reflections
Data with $\|F\| > 2\sigma$	6.0–2.2	0.205	0.285	24,801
All data	6.0–2.2	0.218	0.301	27,565
r.m.s. deviations	Bond lengths 0.014 Å	Bond angles 1.85°	Improper dihedral angles 1.69°	

[a] Isomorphous (iso) phasing power = $\Sigma|F_H|/\Sigma\|F_{PH}\text{obs}| - |F_{PH}\text{calc}\|$.

[b] Anomalous (anom) phasing power = $\Sigma|F''_H|/\Sigma\|AD_{obs}| - |AD_{calc}\|$.

[c] Figure of merit = $\int P(\phi)\exp(i\phi)d\phi \int P(df)d(\phi)$, where P is the probability distribution of ϕ, the phase angle.

[d] CHESS A1, Cornell High Energy Synchrotron Source beamline A1; NSLS X25, National Synchrotron Light Source beamlines X25; Yale CSB, Yale Center for Structural Biology; IP, Image Plate; CCD, charged coupled device.

[e] $R_{sym} = \Sigma|I_h - \langle I_h\rangle|/\Sigma Ih$, where $\langle I_h\rangle$ is the average intensity over Friedel and symmetry equivalents.

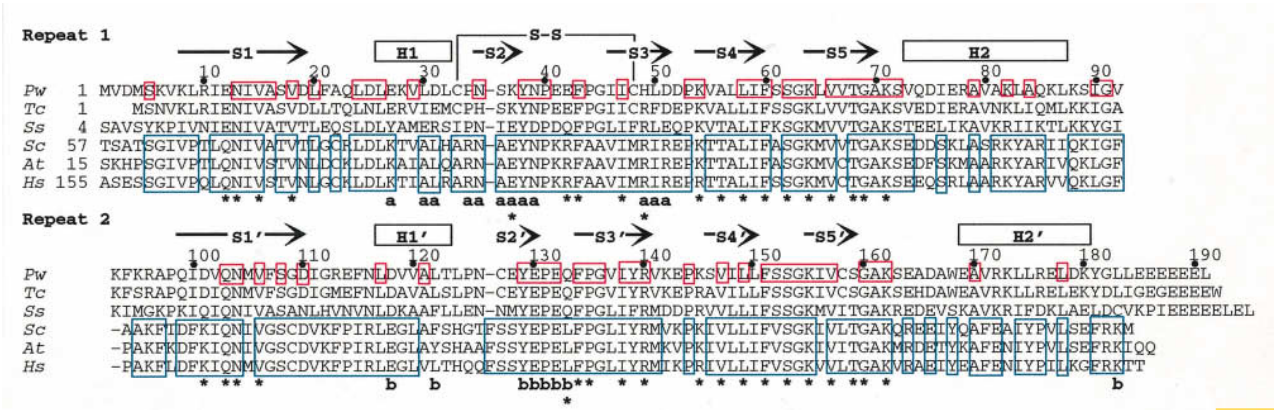

Figure 1. Sequences of the conserved C termini of archaeal and eukaryotic TBPs aligned to maximize sequence identity and similarity (Genetic Computer Group, 1994). The three known TBP sequences from archaeal species are: *Pw*, *Pyrococcus woesei*; *Tc*, *Thermococcus celer*; *Ss*, *Sulfolobus shibatea*. (GenBank: Pwu10285; Tcu04932; Ssu23419) Representing the eukaryotic sequences are: *Sc*, *Saccharomyces cerevisiae*; *At*, *Arabidopsis thaliana*; *Hs*, *Homo sapian*. (SWISSPROT: Tf2d_yeast; Tf2d_arath; Tf2d_human) Numbering and secondary structure (S = beta sheet, H = alpha helix) labeled according to the *Pw*TBP structure (Kabsch & Sander, 1983). Disulfide shown between *Pw*TBP residues 33 and 48. Boxed in blue are residues conserved among eukaryotes and in red are residues conserved between *Pw*TBP and *Sc*TBPc. Residues involved in binding TFIIA, TFIIB, and DNA are labeled a, b, and *, respectively.

known archaeal TBPs but is not found in any eukaryotic TBP sequences. Sequence identity between *Pw*TBP and the cores of all other eukaryotic TBPs ranges between 36% and 41% (Rowlands *et al*., 1994). Consistent with the high degree of sequence identity between archaeal and eukaryotic TBPs, the distribution of secondary structure and overall tertiary fold of *Pw*TBP is essentially the same as those of the eukaryotic TBPs (Figure 2).

The structure of TBPc is composed of two very similar substructures which reflect a sequence repeat that is more pronounced in *Pw*TBP (40% sequence identity between repeats) than eukaryotes (28% to 30% sequence identity between repeats). These substructures are approximately

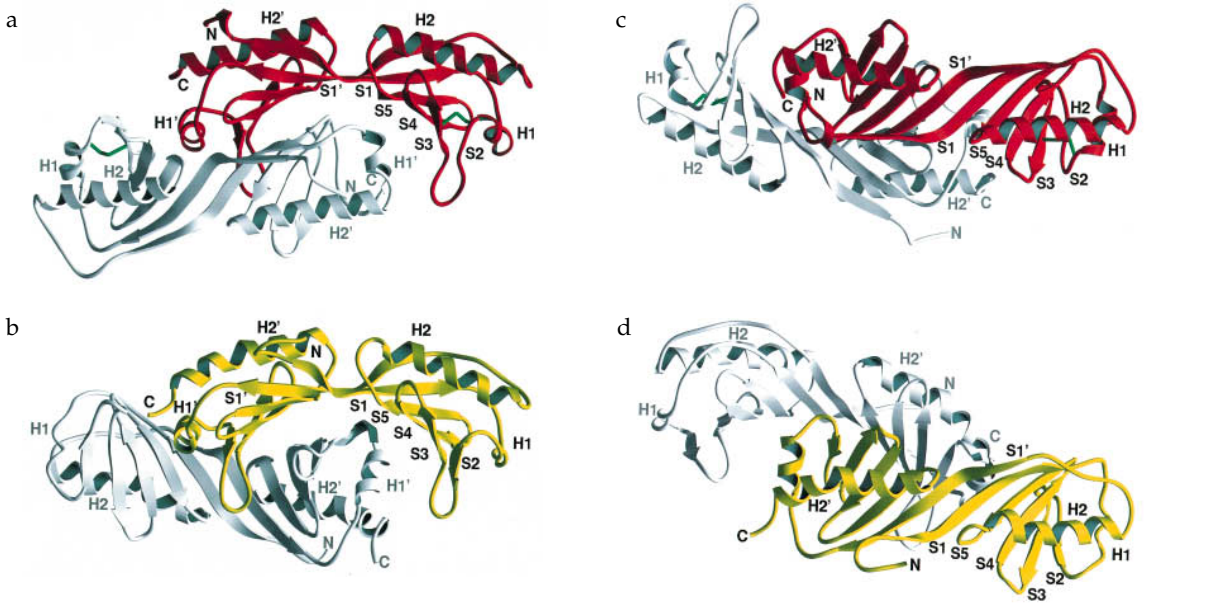

Figure 2. The crystal structures of archaeal and eukaryotic TBPs emphasizing secondary structure and rendered by the program Ribbons (Carson, 1991). a, Ribbons drawing of *Pw*TBP showing both molecules of the asymmetric unit. The disulfide bond between Cys33 of alpha helix H1 and Cys48 of beta strand S3 is green. b, Ribbons drawing of *Sc*TBPc (J. H. Geiger, crystal structure refined at 2.6 Å) showing both molecules of the asymmetric unit with the yellow representation oriented as the red molecule of a. The dimer arrangement of *Sc*TBPc is similar to that observed in the other eukaryotic TBP crystal structure from the species *A. thalinia* (Nikolov *et al*., 1992). c and d, Same as a and b, respectively, with view rotated by 90° along the horizontal axis in the plane of the paper showing the top of the molecule highlighted in a and b.

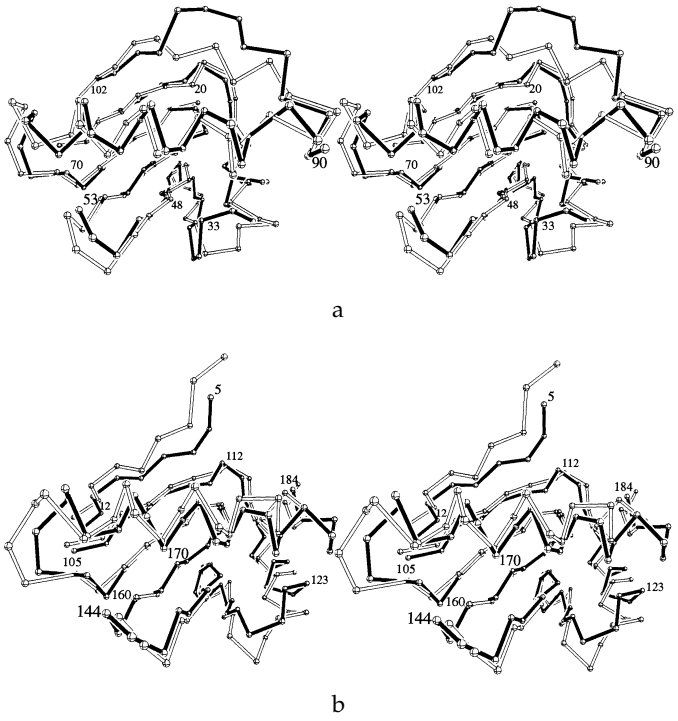

Figure 3. C^{α} carbon traces of *Pw*TBP com ared to that of *Sc*TBPc. a, Superposition of N-terminal substructure of *Pw*TBP (filled black bonds) and *Sc*TBPc (outlined bonds) numbered according to *Pw*TBP sequence and oriented roughly like the N-terminal substructure in the right-hand side of Figure 2a. Pro53 and the corresponding *Sc*TBPc residue are in the *cis* configuration. Residues 33 and 48 are connected by a disulfide bond in the *Pw*TBP structure. Extended loop in *Pw*TBP structure between residues 90 and 102 is the result of a sequence insertion in this region (Figure 1). b, C-terminal substructure and a short sequence of the N terminus of *Pw*TBP (filled black bonds) superimposed upon*Sc*TBPc (outlined bonds) oriented as in a. Numbered according to *Pw*TBP sequence. Pro144 and the corresponding *Sc*TBPc residue are in the *cis* configuration.

dyad symmetric and deviate from one another with a root mean squared difference (RMSD) of 1.8 Å for corresponding C^{α} atoms. If the N-terminal substructure of *Pw*TBP is superimposed on its counterparts from *Sc*TBPc and *At*TBP, the RMSDs for corresponding C^{α} atoms are 1.9 Å and 2.0 Å, respectively (Figure 3a). The same alignment of the C-terminal substructures results in a RMSD of 2.5 Å between *Pw*TBP and *Sc*TBPc and 2.6 Å between *Pw*TBP and *At*TBP (Figure 3b). The sequences of *Sc*TBPc and *At*TBP are 81% identical, which is reflected in a higher structural similarity between them than between either and *Pw*TBP. A comparison of the two substructures of the eukaryotic TBPs results in 1.0 Å and 1.5 Å RMSD between corresponding C^{α} atoms for the N-terminal and C-terminal repeats, respectively.

Proline residues 53 and 144 of *Pw*TBP are in the *cis* configuration as are their eukaryotic counterparts. Cysteine 33 and cysteine 48 are unique to the two sequenced hyperthermophilic archaeal TBPs, *P. woesei* and *T. celer* (Figure 1), and form a disulfide bond (Figures 2 and 4). Electron density in simulated annealing omit maps (Brünger, 1993), in which both cysteine residues and an area encompassing a radius of 8 Å surrounding them was

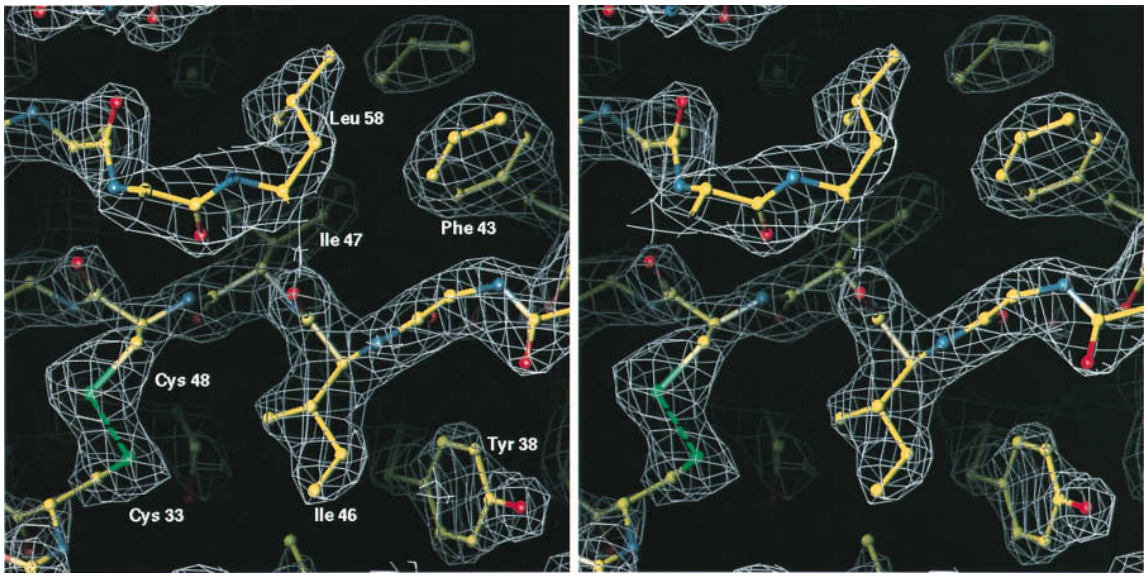

Figure 4. Stereo pair showing $2F_o - F_c$ electron-density map contoured at 2.0σ. The disulfide bond between Cys33 and Cys48 is shown.

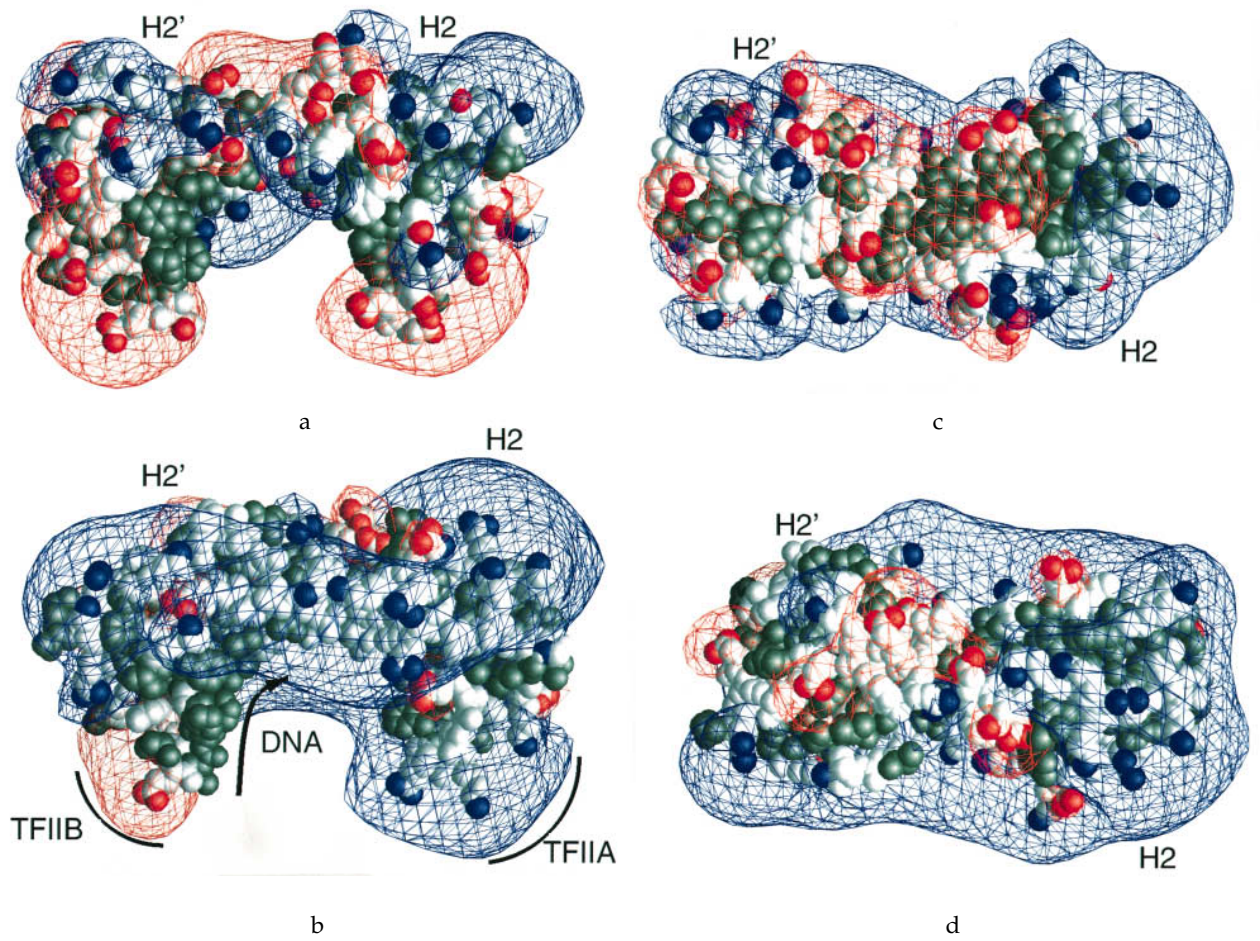

Figure 5. Electrostatic potential surrounding *Pw*TBP (a) and *Sc*TBPc (b) calculated in 100 mM salt and rendered with the program GRASP (Nicholls *et al*., 1993). Orientation of the protein is the same as the red molecule of Figure 2a. Positive potentials are blue and negative potentials are red with both contoured at a magnitude of 1 kT. Charged nitrogen atoms are shown in blue, charged oxygen atoms red, and hydrophobic residues green. Electrostatic potential of *Pw*TBP (c) and *Sc*TBPc (d) labeled the same as above and oriented as the red molecule in Figure 2c.

omitted, confirmed a disulfide bond between them. Based on this density they were refined as cystine. Proline 232 (yeast numbering), which is highly conserved in eukaryotic TBPs and is responsible for a kink in the H2' helix, is not present in the *Pw*TBP sequence. This allows the H2' helix of *Pw*TBP to remain linear, as is the H2 helix in all TBP structures, archaeal and eukaryotic.

The electrostatic potential surrounding *Pw*TBP is dramatically different than that of its eukaryotic counterparts (Figure 5). This is due to more acidic residues on the surface of the *Pw*TBP molecule that result in more ion-pairs (Table 2) and a corresponding neutralization of the positive potential seen in the eukaryotic structures. The possible implications of this on DNA affinity is discussed below.

Homodimer formation

As seen for the eukaryotic TBPs, *Hs*TBP, *At*TBP, and *Sc*TBPc (Coleman *et al*., 1995; Nikolov & Burley, 1994), the archaeal TBP is a dimer in solution as determined by gel filtration (data not shown). Consistent with the dimer formation seen in solution, the crystal structures of all TBPs reveal a dimer in the asymmetric unit, with the monomers related by a non-crystallographic dyad (Figure 2). In all crystal structures (eukaryotic and archaeal) the C-terminal repeats make up most of the dimer interface and mask the hydrophobic DNA-binding surface of each molecule. As shown in Figure 2, the two eukaryotic TBPs dimerize in a similar manner with a solvent-excluded interface for each monomer of 1600 Å^2. The interface between the two monomers in the asymmetric unit of *Pw*TBP also buries 1600 Å^2 of solvent accessible surface area (Table 2), but the nature of the interface is distinctly different (Figure 2), and thus a specific dimer arrangement is not evolutionarily conserved.

Hyperthermostability

To test the thermostability of *Pw*TBP, the melting temperature was determined by differential scanning calorimetry (Figure 6). The midpoint of the thermal denaturation was 101°C in 50 mM

Table 2. Structure statistics[a]

	*Pw*TBP	*Sc*TBPc	*At*TBP
Total number of atoms	1403	1415	1421
Hydrogen bonds (no.)[b]	278/264	265/272	273/276
Ion pairs (no.)[c]			
Monomers	7/8	2/4	5/4
Dimer interface	2		0
Solvent-accessible surface area (Å^2)[d]			
Monomers	9800/9600	10,100/10,400	9800/10,000
Buried in dimer interface	1600	1500	1700
Solvent-accessible surface area composition (%)			
Non-polar	56/55	60/61	60/59
Polar	18/19	17/15	17/16
Charged	26/26	23/24	23/25
Of charged % positive	(55/55)	(72/73)	(69/70)
Atoms buried (%)[e]	46/45	45/43	46/46
Protein density[f]	0.818/0.815	0.796/0.798	0.793/0.796
Void volume (Å^3)[g]	4100/4200	4800/4700	4900/4800

[a] Calculations include both molecules in the asymmetric unit (A/B) and residues 5 to 184 of the *Pw*TBP structure, 19 to 198 of the *At*TBP structure (Nikolov *et al*., 1992), 61 to 240 of the *Sc*TBPc structure (J. H. Geiger, crystal structure refined at 2.6 Å).

[b] Hydrogen bonds were defined by the method of Kabsch & Sander (1983).

[c] Ion-pairs were defined as two ionizable groups $\leqslant$4 Å apart (Barlow & Thornton, 1983).

[d] The algorithm of Lee & Richards (1971) using a probe radius of 1.4 Å.

[e] Atoms with no solvent accessible surface.

[f] Molecular volumes using the program PQMS (Connolly, 1985), and the atomic radii reported by Rashin *et al*. (1986). Data for all structures compared were collected on frozen crystals. van der Waals volume, volume calculated with zero Å radius probe. Solvent-excluded volume, volume calculated with 1.4 Å radius probe. Protein density, van der Waals volume/solvent-excluded volume.

[g] Void volume = solvent-excluded volume – van der Waals volume.

potassium phosphate (pH 7.0). Under reducing conditions the transition decreased to 97°C and in high salt (800 mM potassium phosphate) the transition increased to 109°C. Thus, pure bacterially expressed *Pw*TBP is, indeed, extremely thermostable indicating that neither endogenous cellular factors nor *in vivo* modifications are needed for this property.

*Pw*TBP-DNA interaction

A thorough analysis of the salt and temperature dependence of *Pw*TBP-DNA interactions is presently under way. Preliminary results indicate a salt and temperature dependence for the binding of *Pw*TBP to its DNA target that is distinctly different from that observed for its eukaryotic counterparts. At room temperature gel mobility shift assays show weak affinity of *Pw*TBP for the elongation factor 1α promoter, AAGC**TTTAAAAA**GTAA (box A sequence underlined; Rowlands *et al*., 1994). At approximately the intracellular salt, concentration found in *P. woesei*, 800 mM potassium phosphate (Scholz *et al*., 1992), titration calorimetry experiments show that *Pw*TBP binds specifically to the same box A sequence with a dissociation constant (K_d) of 13 μM at 25°C (data not shown). Increasing the salt concentration to 1.3 M potassium phosphate enhances affinity 24 times while decreasing it to 50 mM potassium phosphate abolishes any detectable enthalpy of binding. Earlier gel mobility shift assays by Rowlands *et al*. (1994), indicated enhanced affinity at 55 to 60°C, and titration calorimetry experiments corroborate this by showing that affinity is enhanced 30 times when the temperature is raised to 45°C. If both the salt concentration and temperature are elevated to 1.3 M potassium phosphate and 45°C, then the affinity is enhanced 370 times to give a K_d of 35 nM, as shown by titration calorimetry. By contrast, the affinity of *Sc*TBP for the TATA-box (K_d = 2.4 nM; Hahn *et al*., 1989), is diminished 300 times when the salt concentration is increased from 50 mM to 300 mM potassium chloride and also decreases at temperatures exceeding 30°C (Petri *et al*., 1995). Although the temperatures for these *in vitro* studies are well below the optimal growth temperature of *P. woesei* (105°C), it is clear that elevated temperature and increased salt concentration enhance the stability of the archaeal complex at levels which destabilize the eukaryotic interaction.

Interface with general and specific transcription factors

The crystal structure of *At*TBP complexed with *Homo sapiens* TFIIB and DNA demonstrated that TFIIB interacts with the C-terminal ''stirrup'' of *At*TBP (Nikolov *et al*., 1995). Of the eight residues contacted by TFIIB, all but two are conserved or are similar in *Pw*TBP, and there are no charge reversals (Figure 1). Not surprisingly the TFIIB homolog in

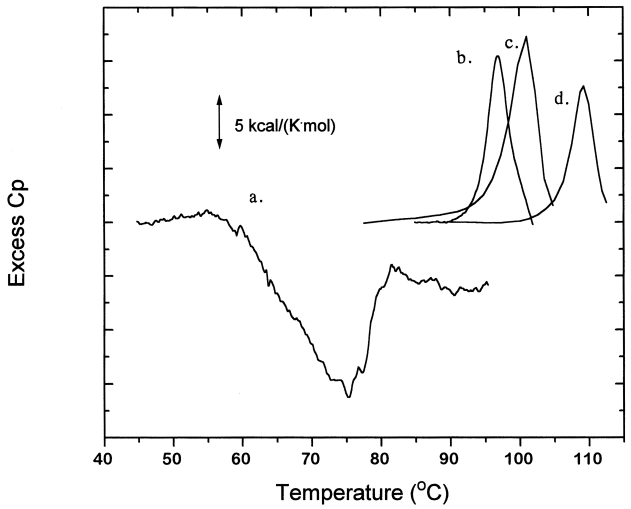

Figure 6. Thermal transitions of *Pw*TBP and *Sc*TBPc measured by differential scanning calorimetry. a, Exothermic denaturation/aggregation of *Sc*TBPc (50 mM potassium phosphate pH 7.0, 10% glycerol) with transition and presumed unfolding occurring about 60°C. b to d, Unfolding transition of *Pw*TBP in (b) reducing conditions (2 mM dithiothreitol (DTT), 50 mM potassium phosphate, pH 7.0), (c) oxidizing conditions (50 mM potassium phosphate, pH 7.0) which is identical with 10% glycerol added; and (d) high salt and oxidizing conditions (800 mM potassium phosphate, pH 7.0).

archaea, TFB, forms strong interactions with both archaeal and eukaryotic TBP/DNA complexes (Rowlands *et al*., 1994; Ghosh *et al*., unpublished). This is consistent with the fact that well ordered crystals of the *Pw*TFB/*Pw*TBP/A-box have been reported (Kosa *et al*., 1996).

Yeast TFIIA (*Sc*TFIIA) is a negatively charged molecule that interacts with four basic residues along the N-terminal ''stirrup'' of *Sc*TBPc (Geiger *et al*., 1996; Tan *et al*., 1996). None of these basic residues in *Sc*TBPc are conserved in *Pw*TBP, and two are of opposite charge. In fact, most of the TBP residues that contact TFIIA in the *Sc*TBPc/*Sc*TFIIA/TATA complex are not conserved (Figure 1). A negative potential surrounds the N-terminal stirrup of *Pw*TBP (Figure 5) and this negative potential would repel the negative charge of a yeast like TFIIA molecule. Additionally, the disulfide bond in *Pw*TBP alters the conformation of the S2 strand of *Pw*TBP relative to that of the eukaryotic TBPs. This S2 strand forms the predominant interacting surface of *Sc*TBPc with *Sc*TFIIA. An archaeal TFIIA homolog has not yet been identified, and if one were to exist, its binding surface to TBP would have to be significantly different.

Basic residues project from the surface of the H2 helix of *Pw*TBP in a manner similar to that found in eukaryotic TBPs (Figure 5c and d). This region, in eukaryotic TBPs, has been implicated in binding the activation domains of eukaryotic transcription factors such as p53 and E1A (Lee *et al*., 1991; Liu *et al*., 1993). Although endogenous activators have not been found in Archaea, these eukaryotic activators have been shown to bind *Pw*TBP (Rowlands *et al*., 1994). Since *Pw*TBP contains a basic region on the H2 helix similar to that of eukaryotic TBPs, it is possible that *Pw*TBP's interactions with some eukaryotic activators are also similar to that of eukaryotic TBPs.

Discussion

Hyperthermostability

Hyperthermophilic organisms are those which grow optimally at temperatures exceeding 90°C (Adams, 1993). Based on phylogenetic studies, most hyperthermophilic genera have been classified as Archaea, while two are Bacteria (Woese *et al*., 1990). *P. woesei* is a hyperthermophilic archaeal species, and as expected, the TBP molecule from this organism is stable at the high temperatures. *Pw*TBP has an unfolding transition of 101°C which is 40°C higher than *Sc*TBPc under comparable conditions (Figure 6). While the free energy level of both the unfolded and native forms contribute to the thermostability of a protein (Dill & Shortle, 1991), this study can only address contributions to stability from the native state since the denaturation is irreversible.

In the crystallographic study of the hyperthermophilic enzyme aldehyde ferredoxin oxidoreductase, the authors conclude that an unusually high fraction of buried atoms (55%) and a corresponding reduced solvent accessible surface area correlates with the thermostability of that molecule (Chan *et al*., 1995). The amount of solvent accessible surface area of *Pw*TBP, although less than its eukaryotic homologs is still 7% greater than the average protein of its size (Miller *et al*., 1987). In addition, *Pw*TBP does not bury an unusual large fraction of its atoms and is identical in this respect to the mesophilic TBPs (Table 2).

One unusual feature of the *Pw*TBP structure that could contribute to its thermostability, is its single disulfide bond. Typical of intracellular proteins, disulfide bonds have not been found in the mesophilic TBPs. However, when the unfolding transition temperature of *Pw*TBP under oxidizing conditions was compared with the transition in a reducing environment (Figure 6), only a modest decrease of 4°C in the transition temperature was observed under reducing conditions.

Four recent studies comparing the crystal structures of hyperthermophilic proteins to those of their mesophilic homologs revealed an increase in the number of ion-pairs in the hyperthermophilic versions (Day *et al*., 1992; Hennig *et al*., 1995; Korndörfer *et al*., 1995; Yip *et al*., 1995). The number of ion-pairs, defined as two ionizable groups $\leqslant 4$ Å apart (Barlow & Thornton, 1983), were calculated for the known TBP structures. *Pw*TBP contains eight ion-pairs on its surface and in comparison, the eukaryotic TBPs have four and five surface ion-pairs (Table 2). Although *Pw*TBP contains more surface ion pairs than the mesophilic TBPs, the

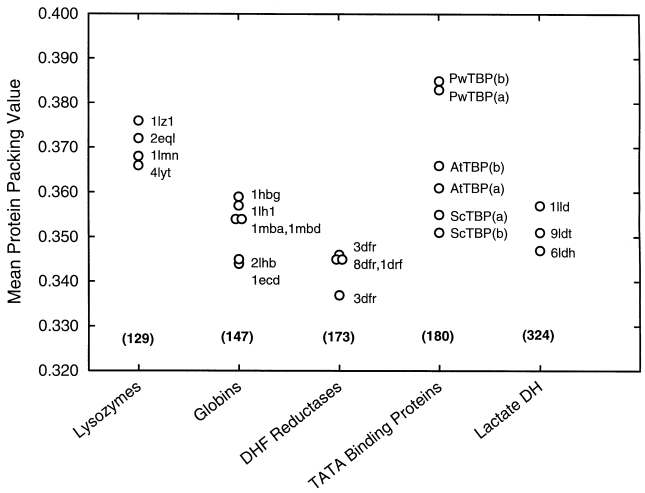

Figure 7. Atomic packing in homologous protein structures. The mean packing value for each protein without prosthetic groups was calculated from analysis of occluded surface as described in the text. Data points are labeled with the respective PDB identifier codes. The homolog families shown for comparison are all groups of high resolution crystal structures (< 2.2 Å) available in the PDB that have high sequence divergence among the members. The numbers in parentheses are the average numbers of amino acid residues per protein in the respective groups. Monomers of the dimeric TATA binding proteins found in the asymmetric unit are labeled a and b.

observation of increased thermostability at higher salt concentrations for *Pw*TBP (Figure 6) argues against electrostatic interactions as the predominant stabilizing force, since the stabilizing effect of a surface ion-pair is not likely to be enhanced by raising the ionic strength of the buffer.

Richards observed that the interiors of proteins are tightly packed with densities similar to those of crystals of small organic molecules (Richards, 1974). Studies of the density of protein interiors have implicated ''packing efficiency'' as a factor in protein stability (for review, see Richards & Lim, 1994). A mean protein packing value was calculated for the TBP structures and the packing value of *Pw*TBP is 6% greater than that of *At*TBP, the closest mesophilic TBP (Figure 7). A survey of packing values is shown for several families of homologous proteins in Figure 7. *Pw*TBP is clearly the most efficiently packed; but more important, its packing value deviates from that of its homologs by a substantial margin, whereas the packing values of other families are more tightly clustered. A complete analysis of the packing efficiency comparing hyperthermophiles and their mesophilic counterparts is currently under way. The increased packing density of *Pw*TBP is due mostly to the contraction of several segments toward the protein core as demonstrated in the difference distance matrix (Figure 8). The contraction of *Pw*TBP has eliminated about 12% of the void volume found within the solvent-excluded volume of the mesophilic proteins (Table 2). The elimination of this void space and consequent increase in packing efficiencies could contribute to the thermostability of *Pw*TBP by increasing the associated favorable van der Waals interactions (Richards & Lim, 1994).

Rather than a single specific structural attribute, a combination of chemical characteristics known to stabilize mesophilic proteins may work together to increase the thermostability of *Pw*TBP. Crystal structures of proteins from hyperthermophilic organisms, including *Pw*TBP described here, highlight several of these chemical characteristics for increasing protein thermostability: (1) disulfide bonds; (2) more ion-pair interactions; (3) increased buried surface area; (4) more compact packing.

*Pw*TBP-DNA interaction

The structure of *Pw*TBP gives a plausible explanation for the different effects of salt and temperature on the affinity of eukaryotic and archaeal TBPs for their respective DNA targets. The majority of contacts between eukaryotic TBPs and DNA are hydrophobic in nature (J. L. Kim *et al*., 1993; Y. Kim *et al*., 1993). Unlike ionic and polar contacts, hydrophobic interactions are not diminished by high salt and elevated temperature; in fact, they are often strengthened under these conditions. In *Pw*TBP, 21 of the 25 residues which are involved in hydrophobic contacts between eukaryotic TBPs and the TATA-box are preserved, and modeling suggests that those which differ would not be disruptive to the *Pw*TBP/DNA complex. Moreover, these ionic/polar contacts between eukaryotic TBP and the DNA backbone are replaced by residues that can only form van der Waals contacts with the sugars. Since the hydrophobic effect is predominantly entropic, an enhanced affinity of TBP for DNA might be expected at higher temperatures which is indeed the case for *Pw*TBP. That eukaryotic TBP's affinity for DNA decreases at elevated salt concentrations may be explained by the fact that in the case of the eukaryotic complex the hydrophobic interactions are supplemented by polar contacts (J. L. Kim *et al*., 1993; Y. Kim *et al*., 1993) and the protein's positive electrostatic potential. In addition to providing an attractive binding force, the positive potential flanks the phosphates of the bound TATA-box helping to splay open the minor groove of the DNA and, thereby bend the DNA target (J. L. Kim *et al*., 1993; Y. Kim *et al*., 1993). Thus, high salt concentrations would be expected to diminish these electrostatic contributions to the stability of eukaryotic TBP/DNA complexes. In *Pw*TBP, however, the large number of surface acidic side-chains that form the frequent ion pairs noted above in connection with thermal stability, create a nearly electrostatically neutral surface (Figure 5). Thus, the *Pw*TBP/DNA complex would not be stabilized by the electrostatic interactions to the extent seen in the eukaryotic TBP/DNA complexes, and increasing the salt concentration would not be expected to reduce *Pw*TBP's affinity for DNA as it does to *Sc*TBP's.

Protein evolution

In many ways, the hyperthermophilic genera of the Bacteria and Archaea appear to be the most ancient of their respective kingdoms (Woese *et al*., 1990). It has been proposed that the common ancestor of all existent life was a hyperthermophile, and that relatively modern mesophilic biology is adapted from an earlier hot environment (Woese, 1987). This assertion has been debated by those who point to the adaptive stress imposed by high temperature on DNA repair (for a review, see Forterre, 1996).

Regardless of the sign of the presumed thermal gradients in evolution, the contrast of *Pw*TBP and its eukaryotic counterparts provides a window into the adaptation of a DNA binding protein to a range of temperatures. Stability of the *Pw*TBP/ TATA-complex at high temperature and high salt concentrations exploits the thermal and salt dependence of the hydrophobic effect; i.e. stability of the hydrophobic interaction is

a

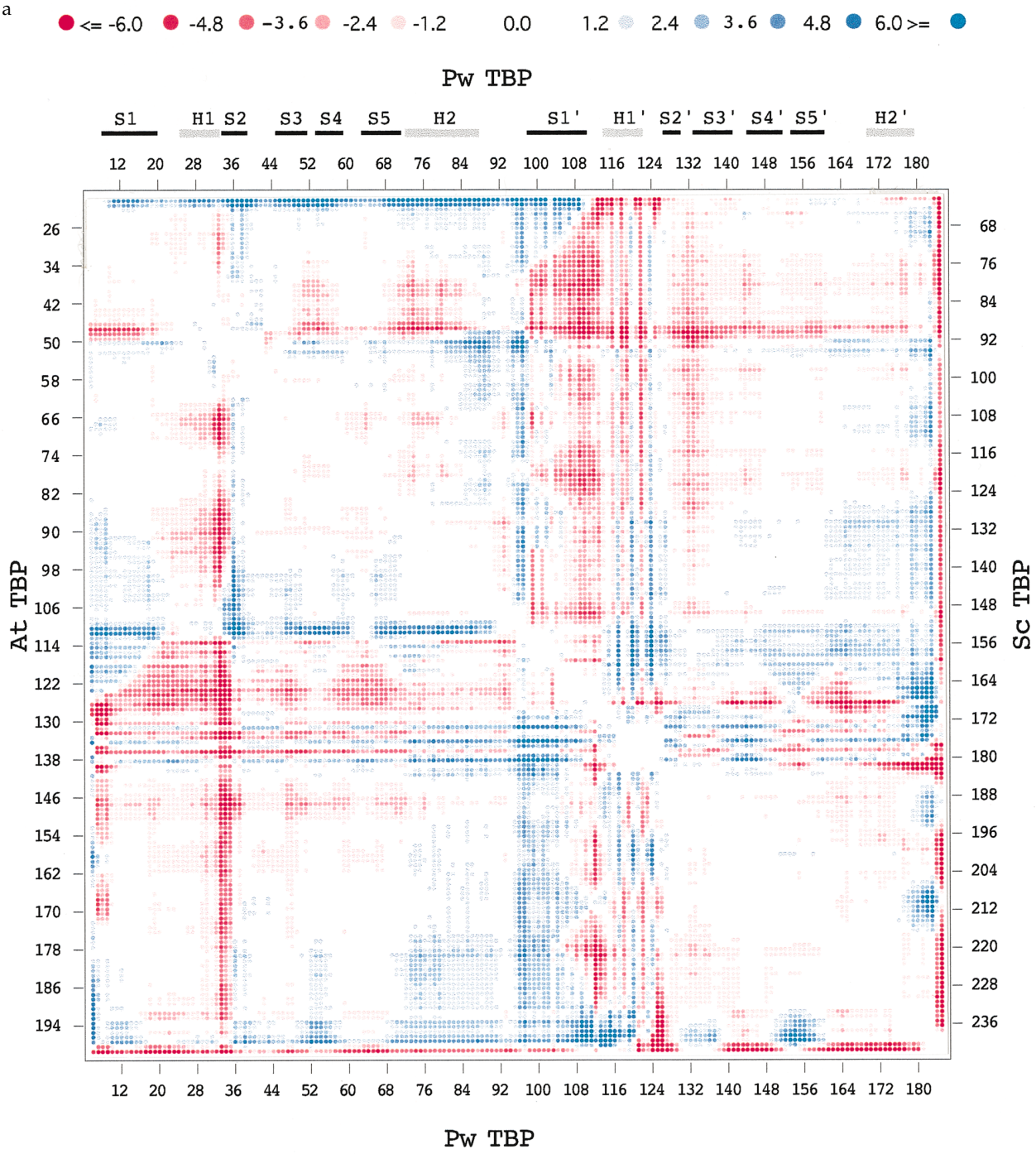

Figure 8a *legend on page* 1081

b

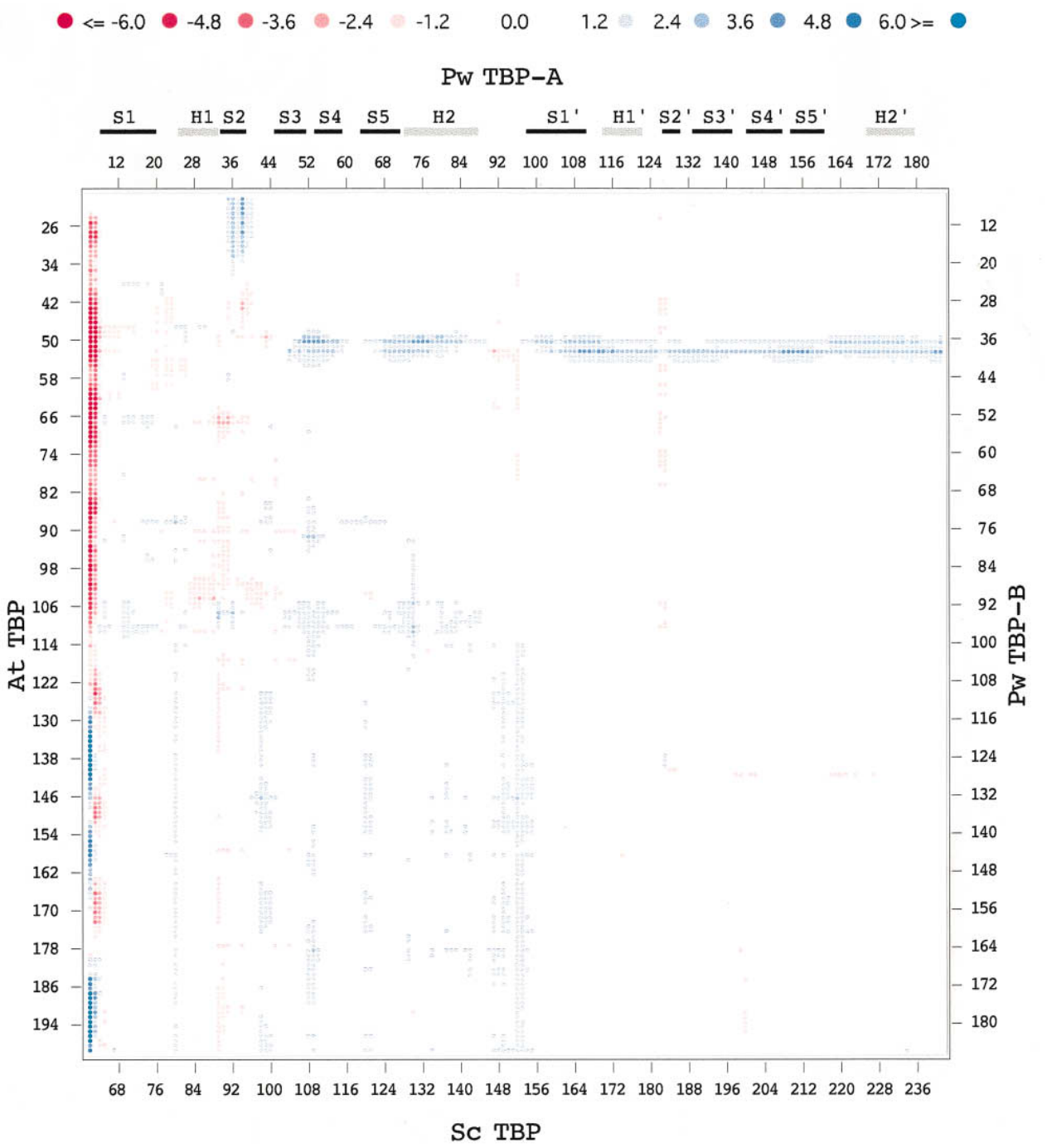

Figure 8. Difference distance matrix of C^{α} atoms. a, Red indicates that the respective intramolecular C^{α}–C^{α} distances of *Pw*TBP are contracted compared to either *Sc*TBPc (above and right of diagonal) or *At*TBP (below and left of diagonal). Blue indicates a corresponding expansion. The relative contraction or expansion is indicated by the color scale at the top of the Figure. Relative displacements of several segments are evident in the Figure. For example the vertical red streak between H1 and S2 shows that the disulfide bond in *Pw*TBP brings Cys33 and the H1 helix closer to most other C^{α} positions when compared to the eukaryotic TBPs. In contrast, the residue insertion between H2 and S1'(see Figure 4a) has resulted in an outward displacement of this loop relative to the core of the first domain as seen by the vertical blue streak in this region. b, ''Control''comparison of the two *Pw*TBP molecules in the asymmetric unit (above and right of diagonal) and between the two eukaryotic structures (below and left of diagonal). Major differences between the two *Pw*TBP molecules in the asymmetric unit representations can be seen at Ser36 and Tyr38 which are in the N-terminal stirrup loop. The only significant differences in C^{α} positions between *Sc*TBPc and *At*TBP are in the N-terminal two residues which probably precede the functional structures.

enhanced by high temperature and high salt. However, the *Pw*TBP described here is only the first DNA-binding regulatory protein in hyperthermophiles to be analyzed structurally. The structure/function relationships of many more cases must be analyzed before physical chemical correlates can be confidently extracted.

Materials and Methods

Protein expression and purification

Full length *Pw*TBP was expressed using a T7 polymerase-based expression system (Novagen) in the *E. coli* strain BL21(DE3). Protein was purified by heating the cell extract to 80°C for 15 minutes, which precipitated the majority of the *E. coli* proteins, leaving the *Pw*TBP in the soluble fraction. The supernatant was then fractionated on hydroxyapatite (BioRad) followed by Mono-Q (Pharmacia). *Pw*TBP eluted from the hydroxyapatite column at 200 mM potassium phosphate (pH 7.0) using a 100 mM to 300 mM gradient, and eluted as a homogeneous protein (as determined by sodium dodecyl sulfate/polyacrylamide gel electrophoresis and Coomassie blue stain) from the Mono-Q column at 250 mM potassium chloride and 20 mM Tris (pH 7.5) using a 100 mM to 400 mM gradient. The protein molecular mass was determined to be 21,320 ($\pm$10) daltons by electrospray mass spectroscopy (Yale mass spectrometry facility), consistent with the calculated 21,311 daltons deduced from the DNA sequence. Protein concentration was determined spectroscopically using the calculated value for ϵ_{280} of 11,050 M^{-1} cm^{-1} based on the extinction coefficients of the chromophores present in the protein sequence in 6 M guanidine (Pace *et al*., 1995). Final recovery of pure protein was 10 mg per liter of *E. coli* cell growth.

Crystallization and structure determination

Hexagonal crystals (0.2 mm $\times$ 0.2 mm $\times$ 0.4 mm, $P6_1$, $a = b = 92.4$ Å, $c = 145.0$ Å) grew in two days at 18°C from 4 μl hanging drops containing 20 mg ml^{-1} protein, 100 mM Tris-HCl (pH 7.5), 100 mM KCl, and 25% saturated ammonium sulfate when equilibrated over a reservoir containing the same solution with 50% saturated ammonium sulfate and no protein. Before data collection, crystals were ''washed'' for one minute in a cryosolvent that contained 25% (w/w) sucrose and all the reservoir components except that the ammonium sulfate concentration was increased to 55% saturated. Washed crystals were mounted in nylon loops, and flash frozen in liquid propane cooled to near liquid nitrogen temperature. Frozen crystals embedded in solid propane were stored in liquid nitrogen until exposure to X-rays in a nitrogen stream at 100 K. These crystals contained two molecules in the asymmetric unit and diffracted to 3.0 Å on laboratory sources and to at least 2.2 Å using synchrotron radiation (Table 1). The mercury derivative was prepared by soaking the crystals in stabilizer (reservoir solution with 55% saturated ammonium sulfate) and 1 mM $HgCl_2$ for 12 hours. All data were processed using DENZO and SCALEPACK (Otwinowski, 1993). Two mercury sites per monomer were found by applying the direct methods algorithm of the program SHELXS-86 (Sheldrick, 1990) to the isomorphous differences between parent data set 2 and the mercury data set. One selenium site per monomer was located from isomorphous difference Fouriers using the phases derived from the mercury derivative. Heavy-atom parameters were refined and phases computed with MLPHARE (Otwinowski, 1991). The initial map was used to define the orientation and position of the non-crystallographic 2-fold symmetry axis and then improved by molecular averaging. To improve the quality of the map, solvent flattening and histogram matching was performed using the program DPHASE (G. Van Duyne). The resulting map allowed easy fitting of the major secondary structural elements. The model was completed after iterative cycles of partial model refinement and phase combination. Selenium sites corresponded to methionine sulfur positions providing an internal check of the sequence assignment. The model was built using the program O (Jones *et al*., 1991) and the structure was refined with X-PLOR (Brünger, 1993) using simulated annealing, conjugate gradients minimization and constrained *B*-factor refinement with data from parent data set 1 at a resolution range of 6.0 to 2.2 Å. The current model includes 183 out of the 191 amino acid residues per molecule and 268 water molecules in the asymmetric unit. Average *B*-factor for all non-hydrogen protein atoms is 42 $Å^2$, consistent with the Wilson plot analysis of the diffraction data. No residues have ϕ, ψ angles in disallowed regions. The current model is refined to an *R*-factor of 0.205 (free *R*-factor = 0.285), for data with $|F| > 2\sigma$ and Bragg spacings of 6.0 to 2.2 Å. The N-terminal methionine and the C-terminal acidic tail consisting of six glutamic acids (residues 185 to 190) and leucine (191) were not represented in ($2|F_o| - |F_c|$) maps for both molecules in the asymmetric unit. Both molecules were refined independently and have a RMSD between C^α atoms of 0.7 Å.

Gel filtration

Gel filtration was performed to determine if the *Pw*TBP forms a dimer in solution, similar to the behavior of the eukaryotic homologs. *Pw*TBP (200 μM) was dialyzed against 40 mM potassium phosphate (pH 7.0). 100 μl of protein sample ($0.25A_{280}$) was injected onto a Pharmacia Superdex 75 gel filtration column (16 mm diameter, 60 cm length) previously equilibrated with the same buffer, and then run with a flow rate of 1 ml/min. For determination of molecular volume the elution profile was compared to molecular weight standards (BioRad) and *Sc*TBPc.

Differential scanning calorimetry

Differential scanning calorimetry was used to determine the unfolding transition temperatures of *Sc*TBPc and *Pw*TBP. *Sc*TBPc was prepared as described (J. L. Kim *et al*., 1993). Protein samples (60 μM) were dialyzed overnight into the various buffer conditions (Figure 6). Differential scanning calorimetry measurements were carried out with the DASM-4 instrument (Biopribor, Puschino, Moscow Region, Russia: Privalov, 1980) at a scan rate of 1 K/min. The reference cell was filled with solvent and the instrumental baselines were determined with solvent filling both cells. All samples were visibly precipitated after denaturation and rescanning of the samples indicated that the thermal denaturation process was irreversible. Thus, the equilibrium thermodynamic parameters could not be derived from these data.

Packing calculation

Analysis of atomic packing for protein structures was performed using an extension of the occluded surface analysis described previously (Pattabiraman *et al*., 1995). In this method, a molecular dot surface of each residue is calculated with a 1.4 Å probe using the MS program of Connolly (1985). A normal is extended radially from each dot until it either intersects the van der Waals surface of a neighboring atom or reaches a length of 2.8 Å (the diameter of a water molecule). Occluded surface, *OS*, is that molecular surface area on the originating atom associated with normals that intersect with another atom as opposed to reaching the 2.8 Å limit, all other molecular surface area is considered non-occluded. The packing value, PV_{res}, for each residue is defined as:

$$PV_{res} = \frac{\sum_{atom}^{res} [OS_{atom}\langle 1 - PP\rangle_{atom}]}{TS_{res}}$$

where TS_{res} is total molecular surface of the residue (sum of occluded and non-occluded areas) and *PP* is the length (maximum = 2.8) of the normal divided by 2.8. For example the PV_{res} of a completely non-occluded residue (100% ''exposed'') equals zero, and the mean PV_{res} for the three residues in tightly packed crystalline Gly-Phe-Gly is 0.571. The mean protein packing value reported in Figure 7 is the mean PV_{res} for each protein.

Acknowledgements

This work was supported in part by a National Institutes of Health grant to P.B.S. (GM15225). B.S.D. was an NIH predoctoral trainee (GM07223). Work in the S.P.J. laboratory was funded by grants from the Medical Research Council (UK) and the Leverhulme trust. We thank Professor J. M. Sturtevant (Yale University) for support and access to the DASM-4 differential scanning calorimeter and Microcal titration calorimeter. Work in the J.M.S. laboratory was funded by National Institutes of Health grant GM04725.

References

Adams, M. W. W. (1993). Enzymes and proteins from organisms that grow near and above 100°C. *Annu. Rev. Microbiol.* **47**, 627–658.

Barlow, D. J. & Thornton, J. M. (1983). Ion-pairs in proteins. *J. Mol. Biol.* **168**, 867–885.

Brünger, A. T. (1992). The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature,* **355**, 472–474.

Brünger, A. T. (1993). *XPLOR Version 3.1: A System for X-ray Crystallography and NMR*. Yale University Press, New Haven.

Bucher, P. (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212**, 563–578.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M. & Clayton, R. A. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*, **273(5278)**, 1058–1073.

Carson, M. (1991). Ribbons 2.0. *J. Appl. Crystallog.* **24**, 958–961.

Chan, M. K., Mukund, S., Kletzin, A., Adams, M. W. W. & Rees, D. C. (1995). Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science,* **267**, 1463–1469.

Chasman, D. I., Flaherty, K. M., Sharp, P. A. & Kornberg, R. D. (1993). Crystal structure of yeast TATA-binding protein and model for interaction with DNA. *Proc. Natl Acad. Sci. USA,* **90**, 8174–8178.

Coleman, R. A., Taggart, A. K. P., Benjamin, L. R. & Pugh, B. F. (1995). Dimerization of the TATA binding protein. *J. Biol. Chem.* **270**, 13842–13849.

Connolly, M. L. (1985). Computation of molecular volume. *J. Am. Chem. Soc.* **107**(5), 1118–1124.

Day, M. W., Hsu, B. T., Joshua-Tor, L., Park, J.-B., Zhou, Z. H., Adams, M. W. W. & Rees, D. C. (1992). X-ray crystal structures of the oxidized and reduced forms of the rubredoxin from the marine hyperthermophilic archaebacterium *Pyrococcus furiosus*. *Protein Sci.* **1**, 1494–1507.

Dill, K. A. & Shortle, D. (1991). Denatured states of proteins. *Annu. Rev. Biochem.* **60**, 795–825.

Forterre, P. (1996). A hot topic: the origin of hyperthermophiles. *Cell,* **85**, 789–792.

Geiger, J., Hahn, S., Lee, S. & Sigler, P. B. (1996). Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science,* **272**, 830–836.

Genetics Computer Group (1994). Wisconsin Package, 8th edit., Madison, Wisconsin.

Hahn, S., Buratowski, S., Sharp, P. A. & Guarente, L. (1989). Yeast TATA-binding protein binds to TATA elements with both consensus and nonconsensus DNA sequences. *Proc. Natl Acad. Sci. USA,* **86**, 5718–5722.

Hennig, M., Darimont, B., Sterner, R., Kirschner, K. & Jansonius, J. N. (1995). 2.0 Å structure of indole-3-gycerol phosphate synthase from the hyperthermophile *Sulfolobus solfataricus*: possible determinats of protein stability. *Structure,* **3**, 1295–1306.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A,* **47**, 110–119.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers,* **22**, 2577–2637.

Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature,* **365**, 520–527.

Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature,* **365(7)**, 512–520.

Korndörfer, I., Steipe, B., Huber, R., Tomschy, A. & Jaenicke, R. (1995). The crystal structure of hologlyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* at 2.5 Å resolution. *J. Mol. Biol.* **246**, 511–521.

Kosa, P., Ghosh, G., DeDecker, B. S. & Sigler, P. B. (1996). Structural analysis of the TBP/TFIIB/DNA complex from the hyperthermophilic Archaea *Pyrococcus woesei*. Abstract no. 328. In *Keystone Symposia* (*Transcriptional Mechanisms*), Taos, New Mexico.

Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.

Lee, W. S., Kao, C. C., Bryant, G. O., Liu, X. & Berk, A. J. (1991). Adenovirus E1A activation domain binds the

basic repeat in the TATA box transcription factor. *Cell,* **67**, 365–376.
Liu, X., Miller, C. W., Koeffler, P. H. & Berk, A. J. (1993). The p53 activation domain binds the TATA box-binding polypeptide in holo-TFIID, and a neighboring p53 domain inhibits transcription. *Mol. Cell Biol.* **13(6)**, 3291–3300.
Marsh, T. L., Reich, C. I., Whitelock, R. B. & Olsen, G. J. (1994). Transcription factor IID in the Archae: sequences in the *Thermococcus celer* genome would encode a product closely related to the TATA-binding protein of eukaryotes. *Proc. Natl Acad. Sci. USA,* **91**, 4180–4184.
Miller, S., Janin, J., Lesk, A. M. & Chothia, C. (1987). Interior and surface of monomeric proteins. *J. Mol. Biol.* **196**, 641–656.
Nicholls, A., Bharadwaj, R. & Honig, B. (1993). GRASP: graphical representation and analysis of surface properties. *Biophys. J. 64, part 2*, A166.
Nikolov, D. B. & Burley, S. K. (1994). 2.1 Å resolution refined structure of a TATA box-binding protein (TBP). *Nature Struct. Biol.* **1(9)**, 621–637.
Nikolov, D. B., Hu, S.-H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N.-H., Roeder, R. G. & Burley, S. K. (1992). Crystal structure of TFIID TATA-box binding protein. *Nature,* **360**, 40–46.
Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G. & Burley, S. K. (1995). Crystal structure of a TFIIB-TBP TATA-element ternary complex. *Nature,* **377**, 119–128.
Otwinowski, Z. (1991). ML-PHARE CCP4 Proc 80-88, Daresbury Laboratory, Warrington, UK.
Otwinowski, Z. (1993). In *Data Collection and Processing*, Warrington, England.
Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. (1995). How to measure and predict the molar adsorption coefficient of a protein. *Protein Sci.* **4**, 2411–2423.
Palmer, J. R. & Daniels, C. J. (1995). In vivo definition of an archaeal promoter. *J. Bacteriol.* **177(7)**, 1844–1849.
Pattabiraman, N., Ward, K. B. & Fleming, P. J. (1995). Occluded molecular surface: analysis of protein packing. *J. Mol. Recognit.* **8**, 334–344.
Petri, V., Hsieh, M. & Brenowitz, M. (1995). Thermodynamic and kinetic characterization of the binding of the TATA binding protein to the adenovirus E4 promoter. *Biochemistry,* **34(31)**, 9977–9984.
Privalov, P. L. (1980). Scanning microcalorimeters for studying macromolecules. *Pure Appl. Chem.* **52**, 479–497.
Qureshi, S. A., Baumann, P., Rowlands, T., Khoo, B. & Jackson, S. P. (1995a). Cloning and functional analysis of the TATA binding protein from *Sulfolobus shibatae*. *Nucl. Acids Res.* **23(10)**, 1775–1781.
Qureshi, S. A., Khoo, B., Baumann, P. & Jackson, S. P. (1995b). Molecular cloning of the transcription factor TFIIB homolog from *Sulfolobus shibatae*. *Proc. Natl Acad. Sci USA,* **92**, 6077–6081.
Rashin, A. A., Iofin, M. & Honig, B. (1986). Internal cavities and buried waters in globular proteins. *Biochemistry,* **25(12)**, 3619–3625.
Richards, F. M. (1974). The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.* **82**, 1–14.
Richards, F. M. & Lim, W. A. (1994). An analysis of packing in the protein folding problem. *Quart. Rev. Biophys.* **26(4)**, 423–498.
Rowlands, T., Baumann, P. & Jackson, S. P. (1994). The TATA-binding protein: a general transcription factor in eukaryotes and Archaebacteria. *Science,* **264**, 1326–1329.
Scholz, S., Sonnenbichler, J., Schäfer, W. & Hensel, R. (1992). Di-myo-inositol-1,1'-phosphate: a new inositol phosphate isolated from *Pyrococcus woesei*. *FEBS Letters,* **306(2,3)**, 239–242.
Sheldrick, G. M. (1990). Phase annealing in Shelx-90: direct methods for larger structures. *Acta Crystallog.* **46**, 467–473.
Tan, S., Hunziker, Y., Sargent, D. F. & Richmond, T. J. (1996). Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature,* **381**, 127–134.
Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* 221–271.
Woese, C. R., Kandler, O. & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archae, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA,* **87**, 4576–4579.
Yip, K. S. P., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V., Scandurra, R. & Rice, D. W. (1995). The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure,* **3**, 1147–1158.
Zillig, W., Palm, P., Klenk, H.-P., Langer, D., Hüdepohl, U., Hain, J., Lanzendörfer, M. & Holz, I. (1993). Transcription in archaea. In *The Biochemistry of Archaea* (*Archaebacteria*) (Kates, M., Kushner, D. J. & Matheson, A. T., eds), vol. 26, pp. 367–386, Elsevier, Amsterdam.

Edited by I. A. Wilson

(Received 17 July 1996; received in revised form 14 October 1996; accepted 14 October 1996)

Note added in proof: The conclusion, that if an archaeal TFIIA were to exist, then it would have to have a radically different TBP binding surface, was confirmed by the sequence of the complete genome of *Methanococcus jannaschii* (Bult *et al*., 1996). The genome of this archaeal species contains no obvious homologs of TFIIA.