Crystal structure of TFIID TATA-box binding protein

Dimitar B. Nikolov*, Shu-Hong Hu*†¶, Judith Lin*†¶, Alexander Gasch‡, Alexander Hoffmann§, Masami Horikoshi§, Nam-Hai Chua‡, Robert G. Roeder§ & Stephen K. Burley*†‖

* Laboratories of Molecular Biophysics, ‡ Laboratory of Plant Molecular Biology, § Laboratory of Biochemistry and Molecular Biology and † Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021-6399, USA

The structure of a central component of the eukaryotic transcriptional apparatus, a TATA-box binding protein (TBP or TFIIDτ) from *Arabidopsis thaliana*, has been determined by X-ray crystallography at 2.6 Å resolution. This highly symmetric α/β structure contains a new DNA-binding fold, resembling a molecular 'saddle' that sits astride the DNA. The DNA-binding surface is a curved, antiparallel β-sheet. When bound to DNA, the convex surface of the saddle would be presented for interaction with other transcription initiation factors and regulatory proteins.

EUKARYOTES have three RNA polymerases (pol I, II and III) that catalyse transcription of nuclear genes[1]. Despite their structural complexity, these multisubunit enzymes require auxiliary proteins known as general initiation factors to initiate transcription from the corresponding class I, II and III promoters[2-5]. Until recently, it was thought that the class I, II and III initiation factors constituted non-overlapping sets of proteins. The TATA-box binding polypeptide (TBP or TFIIDτ), first identified as a component of the class II initiation factor TFIID, is now known however to be required for transcription by all three nuclear RNA polymerases (reviewed in refs 6–9). It thus resembles the essential subunits common to the three RNA polymerases[5] in being a universal transcription factor.

The role of TBP in the initiation and regulation of transcription is best understood for genes transcribed by pol II together with the general initiation factors TFIIA, -B, -D, -E, -F, -G/J, -H and -I[3,4]. In this case, TBP is tightly associated with other polypeptides (the TAFs)[10-12] to form a multiprotein complex called TFIID. TFIID was first identified as a general initiation factor[13] that binds to the TATA element and, thereby, coordinates formation of a functional preinitiation complex (PIC) by class II initiation factors and RNA polymerase II (reviewed in ref. 12). TBP cannot substitute for native TFIID in activator-dependent transcription[14,15] (at least in higher eukaryotes), but can mediate PIC assembly[16] and basal (core promoter) transcription in concert with other general class II factors. It interacts directly with the general initiation factors TFIIA and TFIIB, the C terminus of the large subunit of pol II, some cofactors that inhibit PIC formation (NC1, NC2, DR1), and an initiator-binding factor (TFII-I) that may be important for TATA-less promoters. It can also interact with certain gene-specific activators, suggesting that it may be a direct target for some of these proteins (reviewed in refs 3, 4).

The roles of TBP in transcription by pol I and pol III are less well understood. A TBP–TAF complex (SL1) has, however, been implicated in pol I transcription, and these TAFs are believed to be distinct from those in TFIID[17]. No pol III-specific TBP–TAF complex has yet been reported, although one is thought to exist[8,9].

Soon after the initial description and purification of TBP from yeast[18-20], complementary DNAs encoding TBP were cloned from a variety of eukaryotes (Table 1). The protein has a phylogenetically conserved C-terminal portion of 180 amino acids, containing two 40% identical repeats of 66–67 amino acids flanking a highly basic segment known as the basic repeat[14]. This part of the protein binds to the TATA consensus sequence (TATAa/tAa/t)[21,22] with high affinity ($k_d = 2\text{–}4 \times 10^{-9}$ M)[23], interacting with the minor groove[24,25] and promoting DNA bending[26]. The N-terminal part of the protein varies in length, shows little or no conservation among different organisms, and is largely unnecessary for transcription in certain yeast strains (reviewed in refs. 6,7). Although some plants have two distinct TBPs, most eukaryotes have only one.

We now report the structure of a canonical TATA-box binding protein determined by X-ray crystallography. The structure of TBP-2 from *Arabidopsis thaliana*, obtained at 2.6 Å resolution, reveals a new, highly symmetric DNA-binding fold composed of two topologically identical domains derived from the two direct repeats in the phylogenetically conserved sequence. The domains are related by an approximate 2-fold axis, and consist of a five-stranded, antiparallel β-sheet and two α-helices. The intramolecular symmetry generates a saddle-shaped structure dominated by a curved antiparallel β-sheet, which forms the saddle's concave face and interacts with the DNA. The convex face of the saddle would interact with other proteins during transcription.

Crystallization and structure determination

TBP-2 from *Arabidopsis thaliana*[27] was expressed and purified, and diffraction-quality tetragonal crystals (space group $I4_122$; $a = 105.1$ Å, $c = 215.6$ Å) with two molecules per asymmetric unit were obtained as described in Table 2. The structure was solved by multiple isomorphous replacement (MIR) using five mercurial derivatives (Table 2). The 3 Å resolution MIR map revealed clear electron density for both α-helices and β-strands. After 'solvent flattening'[28] a partial, discontinuous poly-alanine model was fitted to the modified electron density, and the non-crystallographic symmetry operator and molecular boundaries were determined. The electron density map was further improved by phase combination of the refined partial structure model at 3.0 Å resolution[29] with the MIR phase data[30]. Two rounds of model building[31], refinement and phase combination gave an unambiguous trace for about 93% of the asymmetric unit. The resolution of the native data was extended to 2.58 Å using simulated annealing and further refinement. The current model consists of residues 7 to 198 and 13 to 199 in the first and second monomers. No water molecules or sulphate anions have been included in the model, despite appropriately

‖ To whom correspondence should be addressed.
¶ Present addresses: St Vincent's Institute of Medical Research, Fitzroy, Victoria 3065, Australia (S.-H.H.); Department of Medicine, University of Southern California Medical Center, Los Angeles, California 90033-1084, USA (J.L.).

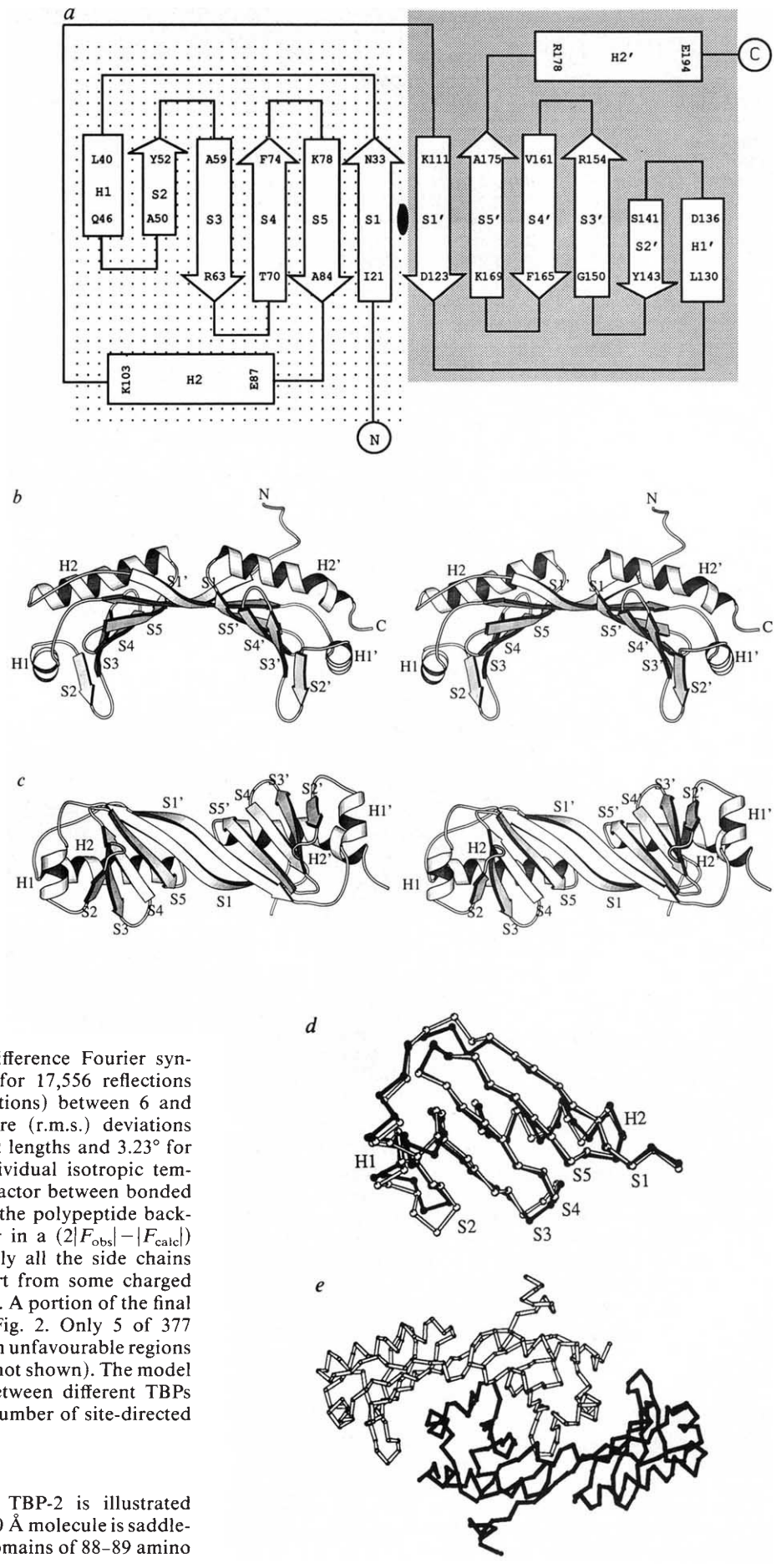


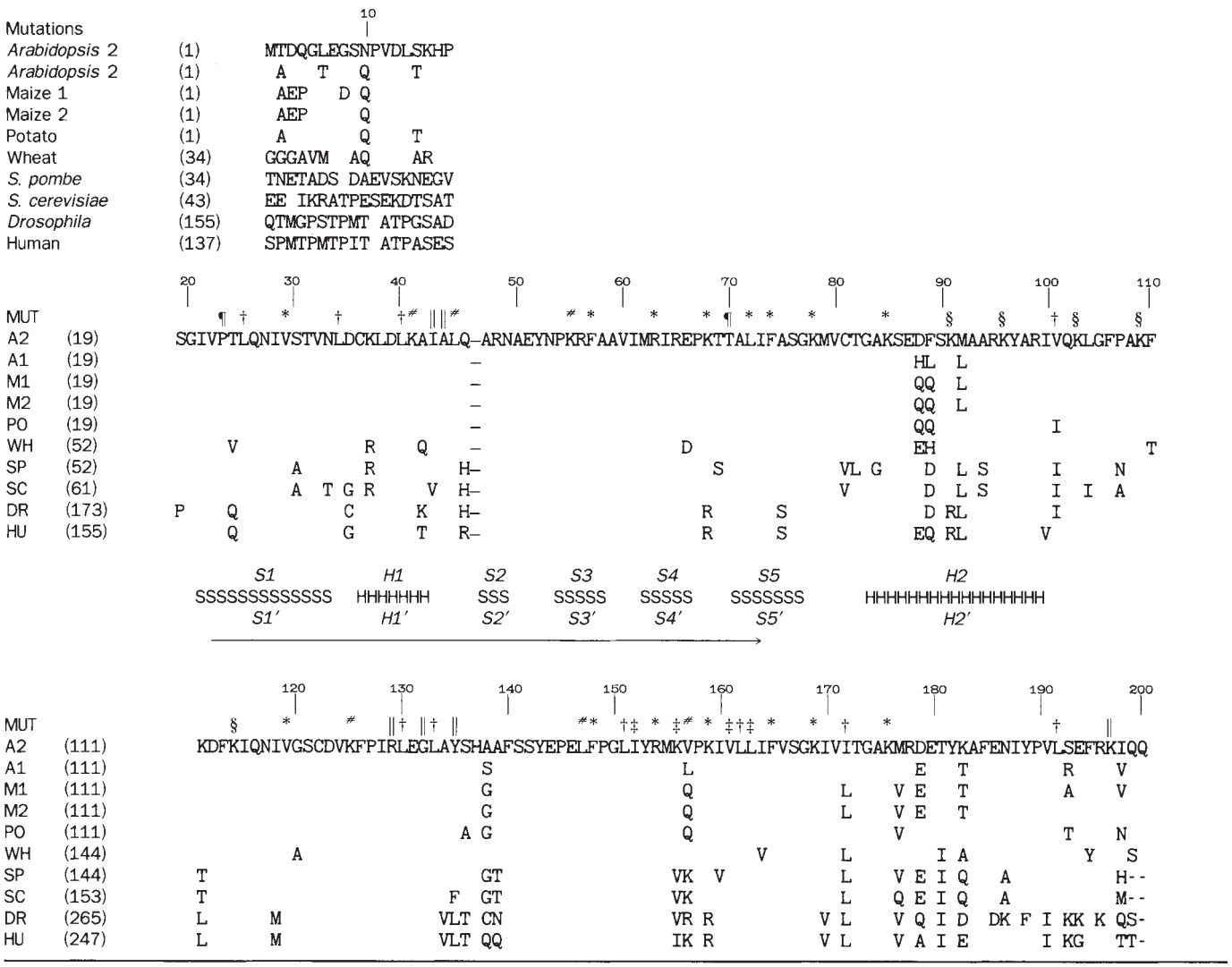
FIG. 1 Structure of TBP-2 from *Arabidopsis thaliana* determined by X-ray crystallography. Residues were assigned to secondary structural elements according to criteria defined by Kabsch and Sander[50]. The N and C termini of the protein are labelled, as are the residues at the end of each numbered α-helix (H) and β-strand (S). The symbol ′ refers to the second structural domain. *a*, Schematic drawing of the secondary structure. α-Helices are shown as rectangles and β-strands are shown as arrows. The location of the axis of approximate 2-fold symmetry is indicated by the symbol (●). The boundaries of the two structural domains are depicted with shading. *b*, Stereodrawing of a Richardson-type[51] representation of the three-dimensional structure viewed perpendicular to the intramolecular 2-fold symmetry axis. The symmetry breaks down slightly in helix H2′, where Pro190 induces a bend. *c*, Stereodrawing viewed along the intramolecular symmetry axis, showing the β-sheet structure on the underside of the molecular saddle. *d*, Superposition of the α-carbon backbone of the first and second domains comprising TBP-2. α-Carbon atomic positions of residues 24–46 and 49–98 were used to determine the least-squares best superposition of the two domains, depicted with open and solid lines (r.m.s. deviation of 1.0 Å for the α-carbon atomic pairs). Significant differences occur at the loop connecting helix H1 with strand S2, where there is a single amino-acid insertion in the second domain (His 137). The two domains are about 30% and 50% identical at the amino-acid and nucleotide levels, respectively. *e*, Schematic drawing of the crystallographic asymmetric unit showing the α-carbon backbone trace of the two tightly associated molecules.

positioned features in ($|F_{obs}| - |F_{calc}|$) difference Fourier syntheses. The current R-factor is 23.0% for 17,556 reflections ($|F| > 1\sigma(|F|)$; 99% of possible observations) between 6 and 2.58 Å resolution, with root-mean-square (r.m.s.) deviations from ideal geometry of 0.015 Å for bond lengths and 3.23° for bond angles, and tightly restrained individual isotropic temperature factors (r.m.s. deviations in B-factor between bonded atoms = 2.3 Å^2). The electron density of the polypeptide backbone is everywhere continuous at 1.25σ in a ($2|F_{obs}| - |F_{calc}|$) difference Fourier synthesis, and virtually all the side chains have well defined electron density, apart from some charged residues located on the molecular surface. A portion of the final electron density map is illustrated in Fig. 2. Only 5 of 377 backbone torsion angle combinations lie in unfavourable regions of the Ramachandran (ϕ, ψ) plot[32] (data not shown). The model is also supported by the comparison between different TBPs described below and the locations of a number of site-directed mutations.

Structure description

Topological overview. The structure of TBP-2 is illustrated schematically in Fig. 1*a*–*c*. The 32 × 45 × 60 Å molecule is saddle-shaped with two very similar structural domains of 88–89 amino

TABLE 1 TBP sequence alignment



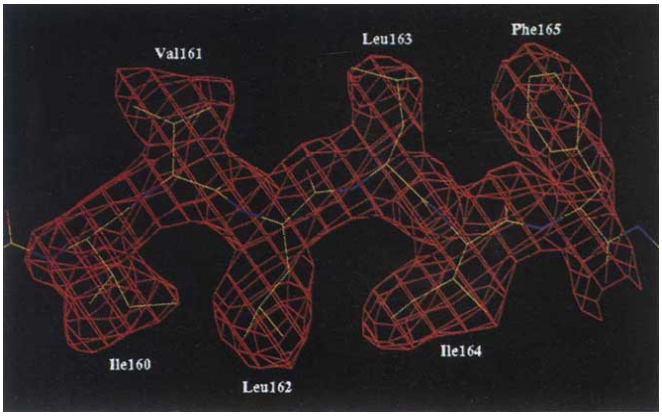
Ten TBP sequences aligned by sequence homology[14,15,27,38,52-62]. The first sequence block includes the entire nonconserved N-terminal region of *Arabidopsis*, maize, and potato TBPs, and the corresponding 18 residues from the TBPs of the remaining organisms. The second and third blocks contain each of the two structural repeats comprising the conserved C-terminal region. Amino acids occupying identical positions in each repeat are aligned vertically. Only differences from the *Arabidopsis* TBP-2 sequence are shown. A gap is present in the first domain to accommodate the insertion of a single amino acid in the second. Secondary structural elements are labelled and indicated with S for strand and H for helix, and the long arrow denotes the extent of the direct repeats. The locations of point mutants and their phenotype are indicated as: Mutations affecting: * DNA binding, † All function, ‡ DNA recognition, § TFIIA interaction, ‖ SPT3 interaction, ¶ Pol. specificity, # No effect.

acids related by approximate intramolecular 2-fold symmetry. Each α/β domain is comprised of two α-helices and a five-stranded, antiparallel β-sheet connected in the order S1-H1-S2-S3-S4-S5-H2. The spatial order of the five β-strands is S1-S5-S4-S3-S2, with the short H1 helix on the edge of the sheet adjacent to strand S2 and the long, amphipathic H2 helix abutting the convex side of the sheet and defining the domain's hydrophobic core. The two domains are connected by a short polypeptide derived from the basic repeat (subsequently referred to as the basic linker peptide).

The loops connecting strands S2 and S3 and S2′ and S3′, respectively, can be visualized as the 'stirrups' of the molecular saddle. The underside of the saddle forms a half-cylindrical recess 32 × 25 Å, lined by the central 8 strands of the 10-stranded β-sheet, which is generated by 2-fold symmetry between the 5-stranded β-sheets of each structural domain. Solvent-accessible residues on the underside include the hydrophobic and small uncharged polar side chains projecting from each of the eight β-strands and the positively-charged residues located on the six loops connecting the strands (Fig. 3*a*). The convex surface of the saddle is composed of the four α-helices, the basic linker peptide, parts of strands S1 and S1′, and the 18 nonconserved N-terminal residues. To the best of our knowledge, the structure of TBP-2 represents a new protein fold.

Intramolecular 2-fold symmetry. The two structural domains of TBP-2 are topologically identical (Fig. 1*d*). Although this finding was not entirely unexpected given the similarity of the direct repeats, the three-dimensional structural overlap goes well beyond the repeats, encompassing almost all of the conserved C-terminal domains. The structural repeats in TBP thus include much of the basic segment between the two direct repeats and the extreme C terminus of TBP (as hypothesized by Stucka and Feldmann[33]).

Non-crystallographic symmetry. The two molecules forming the crystallographic asymmetric unit are related by a non-crystallographic 2-fold axis of rotation (Fig. 1*e*). The r.m.s. deviation



◀ FIG. 2 Electron density map of TBP-2. View of the region of the structure shown to determine the specificity for TATA-box binding. The ($|F_{obs}| - |F_{calc}|$) Fourier difference synthesis was calculated at 2.58 Å resolution using phases calculated by omitting residues 160 to 165 from the refined structural model. The contour level is 2.5σ and the refined atomic model is shown as a stick figure.

between α-carbon atomic positions of the two is 0.48 Å for residues 13 to 198. The intermolecular packing appears to result from tight self-association of TBP. Hydrodynamic studies of various TBPs show that they dimerize readily in solution, unless complexed with DNA containing the TATA-box (see legend to Table 2).

Sequence comparison of TBPs

Seventy per cent of the 180 amino acids in the C-terminus of TBP are conserved in all species so far examined (Table 1). All 54 sites at which differences occur map either to the surface of the molecule, where mutations would be tolerated, or represent conservative changes of buried residues unlikely to destabilize the protein's hydrophobic core. All known TBPs must thus share the same three-dimensional structure in this region[34]. Subsequent references to the results of mutagenesis studies use *Arabidopsis* TBP-2 sequence numbers throughout (see Table 1 for conversion to other sequence numbering schemes).

Species specificity in TBP is largely determined by the conserved 180 amino acids[35,36]. In yeast–human hybrid molecules, residues 130–155 (helix H1′ and strands S2′ and S3′) and 173–188 (helix H2′) must come from yeast to support normal growth in *Saccharomyces cerevisiae*. The only differences between the yeast and human sequences lie in helix H1′, the loop connecting helix H1′ and strand S2′, and helix H2′. These regions constitute part of the convex protein-binding surface in TBP (see below).

The 18 N-terminal amino acids of TBP-2 have not yet been visualized in their entirety, as the two copies in the asymmetric unit adopt distinct conformations and show evidence of disorder. This portion of the TBP sequence is conserved only among plants (Table 1). Normal TBP function in *S. cerevisiae* requires

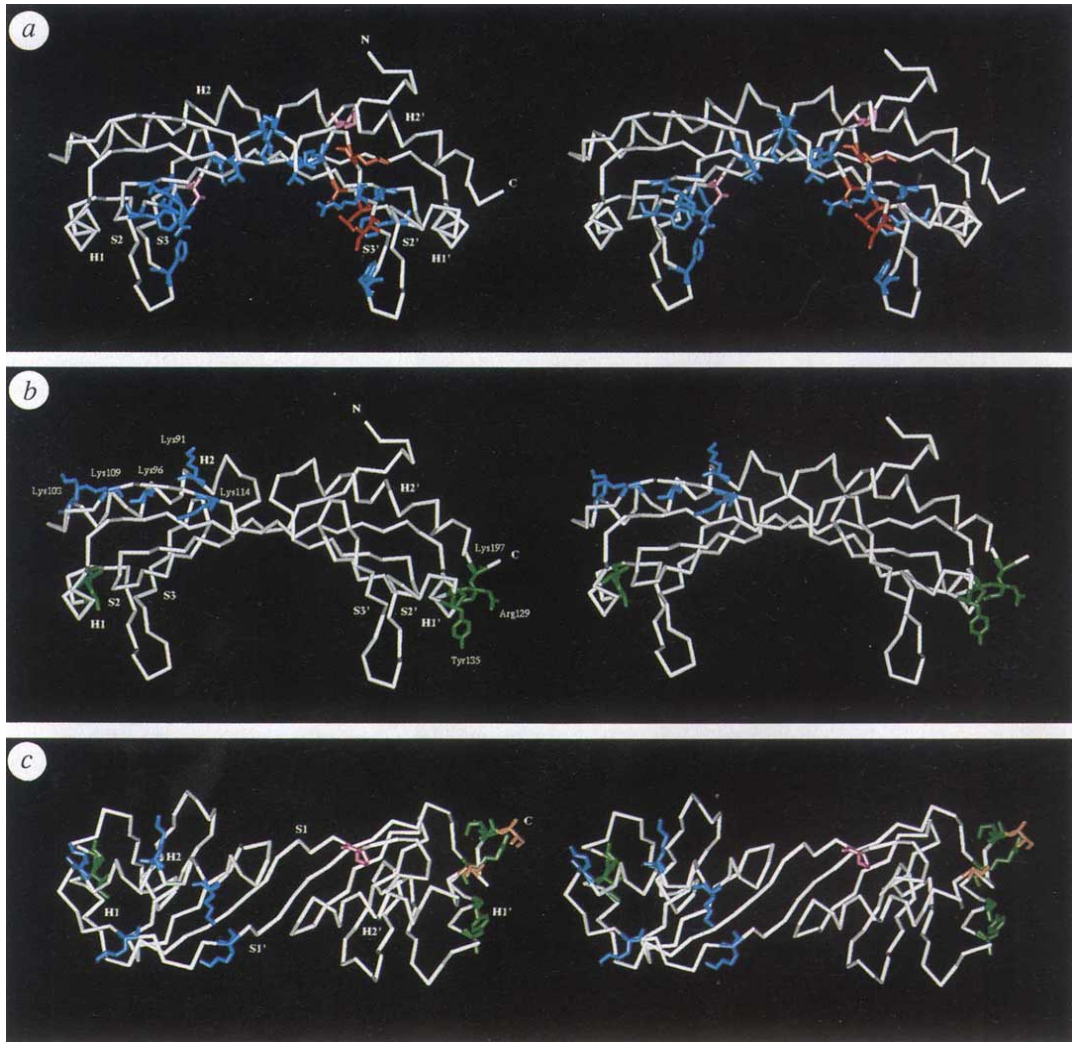


FIG. 3 Schematic stereodrawings of selected TBP point mutants. The α-carbon backbone is shown as a solid white line and highlighted residues are drawn as colour-coded atomic stick figures. *a*, Residues implicated in DNA binding and RNA polymerase specificity, displayed using the view in Fig. 1*b*. Red, alteration in TATA-box specificity; blue, reduced affinity for DNA; pink, alteration in RNA polymerase specificity. *b*, Viewed as in Fig. 1*b*. Residues implicated in TBP-transcription factor interactions: blue, TBP–TFIIA interaction; green, TBP–SPT3 interaction. *c*, View of the upper surface of the molecular saddle. Ser 193 and Ile 198, which may distinguish TBP isoform 1 from TBP isoform 2 in *Arabidopsis* are shown in brown.

several acidic amino acids in this region, but is insensitive to their precise locations[37,38]. Both *Schizosaccharomyces pombe* and plants also possess a relative abundance of acidic residues in this region.

DNA binding and promoter recognition

Biochemical and genetic data confirm that the TBP saddle sits astride the DNA (Figs 3*a* and 4). The model in Fig. 4 protects ~10 base pairs, in good agreement with the 15 base pair DNaseI footprint of *S. cerevisiae* TBP[20].

The locations of point mutations reducing DNA-binding affinity, altering TATA-binding specificity, and changing polymerase specificity are shown in Fig. 3*a*. All mutations affecting DNA binding map to the underside or edges of the molecular saddle. The different mutations may eliminate positive charges from the loops connecting segments of the β-sheet, disrupt the structure of the stirrup-like loops between strands S2 and S3 or S2′ and S3′, or introduce positively charged residues in place of the nonpolar or small, uncharged polar side chains located on the β-strands of the saddle[21,22,39].

Discussion of the TBP-TATA-box interaction must await high-resolution structures of appropriate protein-DNA complexes. Mutations affecting TATA-box binding and polymerase specificity are nevertheless of considerable interest. Two amino-acid changes, Ile 152 → Phe and Leu 163 → Val (Fig. 3*a*, red side chains), are necessary and primarily responsible for changing DNA-binding specificity from TATA to TGTA in yeast[39]. These substitutions map to strands S3′ and S4′, respectively, where they would alter the precise shape of the DNA-binding surface. The mutations Lys 156 → Gln/Asn and Val 161 → Thr, contribute to the effect and map to the loop connecting strands S3′ and S4′ and to strand S4′, respectively. Two further mutations[7] affect the polymerase specificity of TBP (Fig. 3*a*, pink side chains). Thr 70 → Lys, which reduces DNA-binding affinity and abolishes pol II activity, maps to strand S4 and would change the structure and electrostatic appearance of TBP's DNA-binding surface. Pro 23 → Ser, which abolishes both pol II and pol III activities, maps to the beginning of strand S1 where a mutation could alter the convex protein-binding surface of TBP (see below and Fig. 3*b*, *c*).

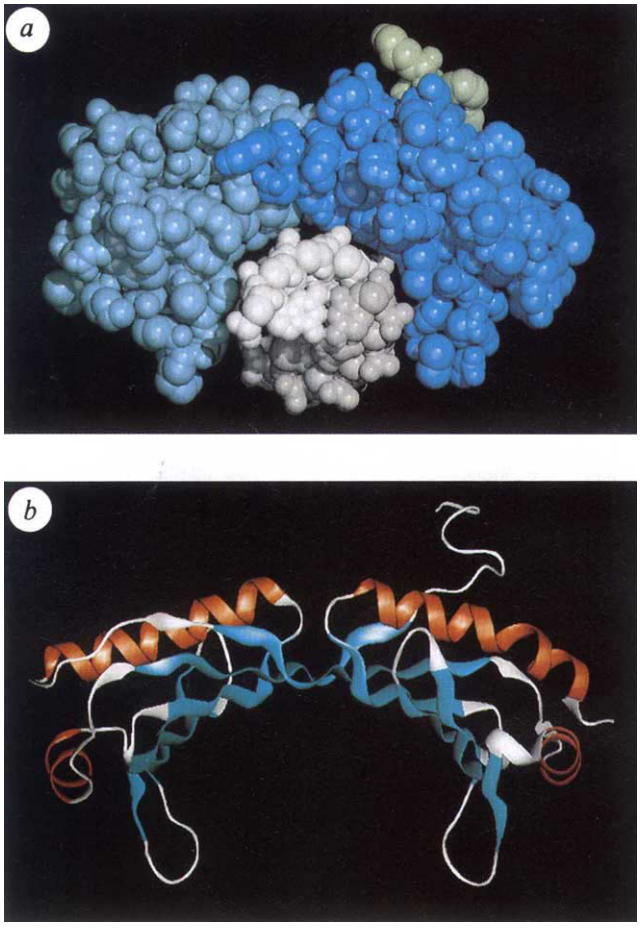


FIG. 4 *a*, Computer-generated model of TBP-2 interacting with B-form DNA without steric repulsion. The protein atoms are colour-coded by their location within the structure (light green, non-conserved segment of the polypeptide chain; light blue, first domain, dark blue, second domain), and the atoms of each DNA strand are colored white and grey, respectively. Each atom is depicted as a sphere of radius equal to the atomic van der Waals radius. It should be emphasized that this figure does not represent the TBP-DNA interaction with any precision, and is included only to illustrate the general picture of DNA-binding derived from this structural study. *b*, Ribbon drawing of TBP-2 viewed as in *a*. α-Helices and β-strands are shown in red and blue, respectively, with the remainder in white.

The half-cylindrical recess defined by the concave surface of TBP-2 will accommodate B-form DNA without steric overlap in computer-generated models (Fig. 4*a*). Unlike any other structurally characterized DNA-binding protein[40–42] TBP recognizes its binding site through specific contacts with the minor groove[24,25]. The DNA-binding face of the protein is a curved, eight-stranded β-sheet, which provides a large surface area for DNA contact. This area may also mediate nonspecific DNA-protein contacts required to position TBP on TATA-deficient promoters for functions analogous to those required during transcription from TATA-containing promoters[43]. Unfortunately, our data do not fix the relative orientations of TBP and the DNA to which it binds, or the relative location of the transcription start site. The DNA-binding surface may flex about a hinge between TBP's two structural domains.

TBP-transcription factor interactions

The crystal structure of TBP-2 also provides a basis for analysing contacts between it and the other proteins involved in transcription, such as general transcription initiation factors, promoter-specific factors and viral transactivators.

Five lysine residues in the basic repeat of TBP are involved in the interaction with TFIIA[44,45] a general transcription initiation factor that binds to and stabilizes the TBP-TATA complex during assembly of the PIC[16]. (Fig. 3*b*, *c*; Table 1). They map to a highly basic area of the convex surface of TBP, containing the polar face of helix H2 and the basic linker. Non-conservative point mutations at any of these sites reduce the strength of the TBP-TFIIA interaction, which is presumably electrostatic.

Four sites on the surface of TBP, where the helix H1′ and the C terminus of the polypeptide chain are in close proximity, are involved in the interaction between TBP and a *S. cerevisiae* protein SPT3, a specificity factor required for transcription initiation at certain promoters (Fig. 3*b*, *c*). The wild-type amino acids include Arg 129, Gly 132, Tyr 135 (Phe in yeast), and Lys 197 (Table 1). Specific polar and nonpolar contacts between the molecular surfaces of TBP and SPT3 are therefore probably responsible for stabilizing this complex. A double mutation at positions 43 and 44, located on the solvent-inaccessible face of helix H1 (Fig. 3*b*, *c*), implicates another part of TBP's surface in interactions with SPT-3.

The *in vitro* association of the adenovirus transactivator E1A with TBP depends on amino acid(s) located between Lys 85 and Tyr 135[47,48]. This region of TBP encompasses helix H2, the basic linker connecting the domains of TBP, strand S1′ and the majority of helix H1′, all of which contribute to TBP's protein-binding surface. Consequently, the E1A binding site may include residues that also mediate the TBP-TFIIA interaction.

Many other TBP-protein interactions occur during transcription initiation, and the protein-binding surface of TBP is by no means fully characterized. The binding sites for TFIIB, the C terminus of the large subunit of pol II, the polymerase-specific TAFs, negative cofactors, and transcriptional activator proteins

TABLE 2 Summary of crystallographic analysis

	Native	THM	MCL	EMP	DMM	MMC
Resolution (Å)	2.58	2.70	2.60	2.70	3.00	3.00
Reflections (observed/unique)	138,216/19,227	49,979/16,157	105,493/18,460	40,469/13,727	36,605/10,834	44,696/10,215
Data coverage (%)	99	94	97	80	86	82
R_{sym} (%)	6.8	5.9	7.8	8.7	8.1	7.2
Mean fractional isomorphous difference		0.29	0.33	0.16	0.24	0.23
MIR analysis (10–3 Å)						
Number of sites		5	9	5	6	5
Phasing power		1.73	1.82	1.16	1.45	1.44
R_c		0.59	0.60	0.64	0.58	0.61
Mean overall figure of merit	0.64					
Refinement						
Resolution (Å)	6–2.58					
R-factor (%)	23.0					
Reflections ($\|F\| > \tau\|F\|$)	17,556					
Total number of atoms	2,988					
r.m.s. bond length (Å)	0.015					
r.m.s. bond angle (degrees)	3.23					

TBP-2 derived from an AT-2 cDNA from *Arabidopsis thaliana*[27] fused with an additional 20 N-terminal amino acids (MGSSHHHHHHSSGLVPRGSH; single-letter code) was overexpressed in *Escherichia coli* (BL21(DE3)pLysS) using the T7 RNA polymerase system[63,64]. Cells grown at 30 °C to an absorbance of 0.7 at 595 nm and induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside for 3 h were collected by low-speed centrifugation. Lysis was done by three cycles of freeze–thaw and DNA digestion was done with DNaseI. The soluble fraction was applied directly to a Ni^{2+}-ion affinity column and washed with a buffer containing increasing amounts of imidazole. TBP with an estimated homogeneity of 95% was then eluted from the resin with a buffer containing 100 mM EDTA. After overnight dialysis, removal of the histidine-containing N-terminal sequence was done by either thrombin or trypsin proteolytic cleavage of the peptide bond between R and G in the above sequence to yield TBP-2 plus three N-terminal amino acids, namely GS and H. A final cation-exchange chromatography step to remove the protease and any remaining contaminants yielded material of homogeneity estimated to be 98% or greater. Mass spectrometry[65] documented that the TBP used for crystallization was neither modified nor further proteolysed during expression and purification. The measured molecular mass was 22,650 ± 5, compared with 22,653 calculated from the predicted amino-acid sequence. Further control experiments were done to ensure that the protein used for the crystallographic study was biochemically active and indistinguishable from recombinant yeast TBP. Gel retardation analyses, done as described in ref. 20, confirmed that the purified protein had full DNA-binding activity (data not shown). Photon correlation spectroscopy (PCS) documented that TBP is a dimer in solution in the absence of DNA and undergoes a dimer-to-monomer transition on binding to oligonucleotides containing the TATA-consensus sequence. The PCS experiments were done with a Biotage dp801 molecular size detector under a wide range of experimental conditions. Purified, recombinant *Arabidopsus thaliana* and *S. cerevisiae* TBPs were monodisperse and dimeric in the absence of DNA. Addition of duplex oligonucleotide containing the TATA-consensus sequence yielded a TBP–DNA complex consisting of 1 molecule of TBP and 1 molecule of duplex oligonucleotide. Presumably, the crystallographic asymmetric unit corresponds to the TBP dimer detected by PCS. Crystals were obtained from 2.0 M ammonium sulphate in a buffer consisting of 20 mM HEPES pH8.5 and 5 mM $MgCl_2$ using vapour diffusion combined with microscopic seeding. X-ray data collection at conventional sample temperatures (4–20 °C) was complicated by a high degree of variation in the unit cell dimensions between crystals grown under identical conditions, and during the course of measurements from a single crystal maintained at constant temperature. Diffraction experiments were carried out at ~150 °C using a Rigaku RAXIS-II imaging plate area detector and a refrigerated nitrogen gas delivery system. Crystals were transiently 'cryoprotected' in a 30% glycerol, 15% PEG4K mixture before placement in a 1 mm diameter wire loop and freezing. Significant unit cell edge length fluctuations were eliminated by optimization and standardization of the freezing protocol, and the unit cell dimensions did not vary by more than ±0.3 Å. Heavy-atom derivatives were prepared by soaking the crystals with mercurial reagents at concentrations of 0.1–1.0 mM for 15–60 h (DMM, dimercury malonate; EMP, ethyl mercury phosphate; MMC, methyl mercury chloride; MCL, mercuric chloride, THM, thimerosal). All derivative data were measured from single crystals, whereas the native data were measured from two crystals. There was no evidence of non-isomorphism between the two native crystals, and no significant non-isomorphism between native and derivative crystals. Oscillation photographs were integrated and reduced to structure factor amplitudes using software provided by the manufacturer. Additional crystallographic software included the PROTSYS package (G. A. Petsko, personal communication), HEAVY[66], and MOLSCRIPT[67]. To our knowledge this is the first MIR structure determination undertaken entirely using crystals frozen at near liquid nitrogen temperature. Atomic coordinates and structure factor amplitudes have been submitted to the Brookhaven Protein Data Bank[68]. $R_{sym} = \sum |I - \langle I \rangle| / \sum I$, where I = observed intensity, $\langle I \rangle$ = average intensity obtained from multiple observations of symmetry related reflections. Mean fractional isomorphous difference = $\sum ||F_{PH}| - |F_P|| / \sum |F_P|$, $|F_P|$ = protein structure factor amplitude, $|F_{PH}|$ = heavy-atom derivative structure factor amplitude. $R_c = \sum ||F_{H(obs)}| - |F_{H(calc)}|| / \sum |F_{H(obs)}|$, $|F_{H(obs)}|$ = observed heavy atom structure factor amplitude and $|F_{H(calc)}|$ = calculated heavy-atom structure factor amplitude. Phasing power = root mean square ($|F_H|/E$), $|F_H|$ = heavy-atom structure factor amplitude and E = residual lack of closure. r.m.s. bond lengths and r.m.s. bond angles are the respective r.m.s. deviations from ideal values.

remain undetermined.

TBP diversity in plants

Unlike mammals, *Drosophila* and yeast, *Arabidopsis thaliana* and maize each have two different TBPs (reviewed in ref. 49). Comparison of the plant sequences reveals only two positions at which both pairs of proteins differ (Table 1). In TBP-1, residue 193 is arginine in *Arabidopsis* and alanine in maize, whereas in both TBP-2s it is serine. Residue 198 is valine in both TBP-1s and isoleucine in both TBP-2s. It is unknown whether TBP-1 and TBP-2 have identical roles in *Arabidopsis* and in maize, but because residues 193 and 198 occur on the surface of helix H2′ (Fig. 3*c*), functional differences between the two TBP isoforms may result from isoform-specific interactions with other proteins.

Conclusion

TBP is a saddle-shaped protein with a new DNA-binding fold. The concave DNA-binding surface of the saddle is a curved antiparallel β-sheet, which may mediate both specific and non-specific contacts with DNA. On binding to DNA, the convex surface of the saddle would be presented for interaction with TAFs, transcription initiation factors and other regulatory proteins. Although the structure of the protein displays approximate intramolecular 2-fold symmetry, neither the DNA- nor the protein-interaction surfaces of TBP are symmetric, and the two domains present chemically distinct patterns of amino-acid side chains. TBP can thus bind DNA and other proteins in a directional manner.

The structure presented here provides a starting point for further crystallographic, biochemical and genetic studies of TBP. In particular, it should now be possible systematically to determine first, which regions of TBP's protein-binding surface mediate very tight (and possibly irreversible) interactions with polymerase-specific TAFs, and second, how the TBP–TAF complex (or in yeast TBP itself) interacts reversibly with other proteins involved in transcription. Given the somewhat limited protein-binding surface of TBP, many of the interaction sites may be overlapping and competition for them

may influence transcription.

Finally, the approximate intramolecular symmetry and amino-acid sequence similarity of TBP's two structural domains suggests that the primordial ancestor of TBP functioned as a dimeric DNA-binding protein before gene duplication and fusion produced the structure described here. □

Received 10 September; accepted 19 October 1992.

1. Sentenac, A. *CRC Curr. Rev. Biochem.* **18,** 31–90 (1985).
2. Reeder, R. H. *Trends Genet.* **6,** 390–395 (1990).
3. Roeder, R. G. *Trends biochem. Sci.* **16,** 402–408 (1991).
4. Zawel, L. & Reinberg, D. *Curr. Opinion Cell Biol.* **4,** 488–495 (1992).
5. Gabrielsen, O. S. & Sentenac, A. *Trends biochem. Sci.* **16,** 412–416 (1991).
6. Cormack, B. P. & Struhl, K. *Cell* **69,** 685–696 (1992).
7. Schultz, M. C., Reeder, R. H. & Hahn, S. *Cell* **69,** 697–702 (1992).
8. Sharp, P. A. *Cell* **68,** 819–821 (1992).
9. White, R. J. & Jackson, S. P. *Trends Genet.* **8,** 284–288 (1992).
10. Dynlacht, B. D., Hoey, T. & Tjian, R. *Cell* **66,** 563–576 (1991).
11. Pugh, B. F. & Tjian, R. *Genes Dev.* **5,** 1935–1945 (1991).
12. Takada, R. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
13. Matsui, T., Segall, J., Weil, P. A. & Roeder, R. G. *J. biol. Chem.* **255,** 11992–11996 (1980).
14. Hoffmann, A. *et al. Nature* **346,** 387–390 (1990).
15. Peterson, M. G., Tanese, N., Pugh, B. F. & Tjian, R. *Science* **248,** 1625–1630 (1990).
16. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. *Cell* **56,** 549–561 (1989).
17. Comai, L., Tanese, N. & Tjian, R. *Cell* **68,** 965–976 (1992).
18. Cavallini, B. *et al. Nature* **334,** 77–80 (1988).
19. Buratowski, S., Hahn, S., Sharp, P. A. & Guarente, L. *Nature* **334,** 37–42 (1988).
20. Horikoshi, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **86,** 4843–4847 (1989).
21. Yamamoto, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **89,** 2844–2848 (1992).
22. Reddy, P. & Hahn, S. *Cell* **65,** 349–357 (1991).
23. Hahn, S., Buratowski, S., Sharp, P. A. & Guarente, L. *Proc. natn. Acad. Sci. U.S.A.* **86,** 5718–5722 (1989).
24. Lee, D. K., Horikoshi, M. & Roeder, R. G. *Cell* **67,** 1241–1250 (1991).
25. Starr, D. B. & Hawley, D. K. *Cell* **67,** 1231–1240 (1991).
26. Horikoshi, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **89,** 1060–1064 (1992).
27. Gasch, A., Hoffmann, A., Horikoshi, M., Roeder, R. G. & Chua, N.-H. *Nature* **346,** 390–394 (1990).
28. Wang, B.-C. *Meth. Enzym.* **115,** 90–112 (1985).
29. Brunger, A. T., Kuriyan, J. & Karplus, M. *Science* **235,** 458–460 (1987).
30. Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. *Nature* **347,** 37–44 (1990).
31. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta Crystallogr.* A**47,** 110–119 (1991).
32. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. *J. molec. Biol.* **7,** 95–99 (1963).
33. Stucka, R. & Feldmann, H. *FEBS Lett.* **261,** 223–225 (1990).
34. Sander, C. & Schneider, R. *Proteins Struct. Funct. Genet.* **9,** 56–68 (1991).
35. Cormack, B. P., Strubin, M., Ponticelli, A. S. & Struhl, K. *Cell* **65,** 341–348 (1991).
36. Gill, G. & Tjian, R. *Cell* **65,** 333–340 (1991).
37. Zhou, Q., Schmidt, M. C. & Berk, A. J. *EMBO J.* **10,** 1843–1852 (1991).
38. Fikes, J. D., Becker, D. M., Winston, F. & Guarente, L. *Nature* **346,** 291–294 (1990).
39. Strubin, M. & Struhl, K. *Cell* **68,** 721–730 (1992).
40. Harrison, S. C. *Nature* **353,** 715–719 (1991).
41. Pabo, C. O. & Sauer, R. T. *A. Rev. Biochem.* **61,** 1053–1095 (1992).
42. Phillips, S. E. V. *Curr. Opinion struct. Biol.* **1,** 89–98 (1991).
43. Wiley, S. R., Kraus, R. J. & Mertz, J. E. *Proc. natn. Acad. Sci. U.S.A.* **89,** 58144–5818 (1992).
44. Lee, D. K., DeJong, J., Hashimoto, S., Horikoshi, M. & Roeder, R. G. *Molec. cell. Biol.* **12,** 5189–5196 (1992).
45. Ranish, J. A., Lane, W. S. & Hahn, S. *Science* **255,** 1127–1129 (1992).
46. Eisenmann, D. M., Arndt, K. M., Ricupero, S. L., Rooney, J. W. & Winston, F. *Genes Dev.* **6,** 1319–1331 (1992).
47. Horikoshi, N. *et al. Proc. natn. Acad. Sci. U.S.A.* **88,** 5124–5128 (1991).
48. Lee, W. S., Kao, C. C., Bryant, G. O., Liu, X. & Berk, A. J. *Cell* **67,** 365–376 (1991).
49. Katagiri, F. & Chua, N.-H. *Trends Genet.* **8,** 22–27 (1992).
50. Kabsch, W. & Sander, C. *Biopolymers* **22,** 2577–2637 (1983).
51. Richardson, J. S. *Meth. Enzym.* **115,** 359–390 (1985).
52. Haass, M. M. & Feix, G. *FEBS Lett.* **301,** 294–298 (1992).
53. Holdsworth, M. J., Grierson, C., Schuch, W. & Bevan, M. *Pl. molec. Biol.* **19,** 455–464 (1992).
54. Kawata, T., Minami, M., Tamura, T.-A., Sumita, K. & Iwabuchi, M. *Pl. molec. Biol.* **19,** 867–872 (1992).
55. Hoffmann, A. *et al. Genes Dev.* **4,** 1141–1148 (1990).
56. Horikoshi, M. *et al. Nature* **341,** 299–303 (1989).
57. Cavallini, B. *et al. Proc. natn. Acad. Sci. U.S.A.* **86,** 9803–9807 (1989).
58. Hahn, S., Buratowski, S., Sharp, P. A. & Guarente, L. *Cell* **58,** 1173–1181 (1989).
59. Schmidt, M. C., Kao, C. C., Pei, R. & Berk, A. J. *Proc. natn. Acad. Sci. U.S.A.* **86,** 7785–7789 (1989).
60. Hoey, T., Dynlacht, B. D., Peterson, M. G., Pugh, B. F. & Tjian, R. *Cell* **61,** 1179–1186 (1990).
61. Muhich, M. L., Iida, C. T., Horikoshi, M., Roeder, R. G. & Parker, C. S. *Proc. natn. Acad. Sci. U.S.A.* **87,** 9148–9152 (1990).
62. Kao, C. C. *et al. Science* **248,** 1646–1650 (1990).
63. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. *Meth. Enzym.* **185,** 60–89 (1990).
64. Hoffmann, A. & Roeder, R. G. *Nucleic Acids Res.* **19,** 6337 (1991).
65. Chait, B. T. & Kent, S. B. H. *Science* **257,** 1885–1894 (1992).
66. Terwilliger, T. & Eisenberg, D. *Acta Crystallogr.* A**39,** 813–817 (1983).
67. Kraulis, J. P. *J. appl. Crystallogr.* **24,** 946–950 (1991).
68. Bernstein, F. C. *et al. J. molec. Biol.* **112,** 535–542 (1977).

ACKNOWLEDGEMENTS. D.B.N. thanks A. Ferre-D'Amare for help and advice throughout this work and we thank S. Jacques for help with protein production, B. Chait and S. Cohen for doing the mass spectrometry at the National Resource for Mass Spectrometric Analysis of Biological Macromolecules and R. Brown, K. L. Clark, X. P. Kong, J. Kuriyan, P. Model and G. A. Petsko for their many useful suggestions. A.G. was supported by postdoctoral fellowships from the Deutsche Forschungsgemeinschaft and the Norman and Rosetta Winston Foundation. A.H. is a Boeringer-Ingelheim Graduate Fellow. M.H. was an Alexandrine and Alexander L. Sinsheimer Scholar. This work was supported by the Howard Hughes Medical Institute (S.K.B.) and by the NIH (M.H. and R.G.R.), Monsanto (N.-H.C.) and the Pew Trust.

LETTERS TO NATURE

Possible cataclysmic variable in the core of the globular cluster 47 Tucanae

Francesco Paresce*†, Guido De Marchi*† & Francesco R. Ferraro‡

* Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA
‡ Osservatorio Astronomico di Bologna, Via Zamboni 33, 40126, Bologna, Italy

SEVERAL globular clusters have low-luminosity X-ray sources at their cores[1–3], of which X0021.8−7221 in 47 Tucanae (NGC104) is one of the brightest. The nature of these sources is a mystery, and attempts to detect optical counterparts have been frustrated by the overcrowding of stars in cluster cores. The resolution of ground-based observations is insufficient to pick out the faint blue objects that are likely to be the optical counterparts to the X-ray sources, but the Faint Object Camera (FOC) on the Hubble Space Telescope has proved effective in identifying blue objects in dense clusters such as 47 Tuc[4]. Thus encouraged to use the FOC to search for the optical counterpart to X0021.8−7221, we have discovered a faint, variable and very blue object located in the error circle[5] for X0021.8−7221 from the Einstein satellite High-Resolution Imager. The object has the spectral characteristics of a cataclysmic variable, and the high inferred X-ray to optical brightness ratio suggests that it is a magnetic cataclysmic variable, one of the candidates that has been suggested as the X-ray source in globular clusters.

Several working hypotheses have been put forward for the nature and origin of the low-luminosity X-ray sources in globular clusters. They were at first thought to be normal cataclysmic variables (CVs) created in abundance in the dense cores by tidal capture between a red star and a white dwarf[7,8]. Reconsideration of this process, as there is no evidence that they are present in such large numbers (the dwarf nova V101 in M5 is probably the only confirmed CV in a globular cluster[9]), has shown[10] that its usual result might be unstable mass transfer, sharply reducing the population of CVs. Other possibilities, then, include magnetic CVs[6], low-mass X-ray binaries viewed at high inclination[11], soft X-ray transients in quiescence[12] and close binaries in which there is a rapidly rotating secondary with hotspots[10,13].

To distinguish between these possibilities and to shed light on the nature of close binaries that may effect the dynamical evolution of globular cluster cores during and after collapse[14], ultraviolet or optical identification of a counterpart is essential. We have carefully searched the field contained in the High-Resolution Imager (HRI) error circle overlapping with the available FOC images to look for objects with the markedly blue spectrum expected from a CV accretion disk. The investigation revealed within the small HRI error circle an interesting object, listed as DPF-2213 in the table provided by De Marchi *et al.*[15], which we shall call V1 for simplicity (Fig. 1). The F342W and F195W filters used here are described in ref. 16. The combination of these two filters has a peak response at 3,400 Å and a full width at half maximum of 706 Å, approximating the standard Johnson U bandpass. The instrumental arrangement in this

† F.P. is on assignment from the Astrophysics Division, Space Science Department, ESA, and on leave from the Osservatorio Astronomico di Torino, Italy. G.D.M. is on leave from the Dipartimento di Astronomia, Scuola di Dottorato, Università di Firenze, Italy.