Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural studies of a yeast quaternary transcription-initiation complex

Sunny M. C. Wang, Grace J. Kim and Daniel T. Gewirth*

Duke University Medical Center, Department of Biochemistry, Durham, NC 27710, USA

Correspondence e-mail: gewirth@duke.edu

Received 14 September 2000
Accepted 8 January 2001

A 96.7 kDa quaternary transcription-factor complex consisting of the conserved core domains of yeast TBP, TFIIA, TFIIB and TATA-box DNA has been assembled from purified components. Crystals of the complex were obtained by the hanging-drop vapor-diffusion method. Native data sets were collected at synchrotron sources. Crystal form I belongs to space group $R32$, with unit-cell parameters $a = b = 173.2$, $c = 164.1$ Å, and diffracts to 2.5 Å resolution, but contains substoichiometric amounts of TFIIB. Crystal form II was assembled with a longer piece of DNA than that used for form I crystals and contains the complete quaternary complex. These crystals belong to space group $P42_12$, with unit-cell parameters $a = b = 141.7$, $c = 112.0$ Å, and diffract to 3.6 Å resolution.

1. Introduction

A key step in the initiation of transcription is the recruitment of RNA polymerase II to promoters by the general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH (Conaway & Conaway, 1993; Zawel & Reinberg, 1993). The first step in the recognition of the promoter is the binding of TATA-box binding protein (TBP) – the DNA-binding component of TFIID – to the eight base-pair TATA box (Buratowski *et al.*, 1989). TFIIA and TFIIB bind after TBP and together form a TBP–DNA–TFIIA–TFIIB 'quaternary' subassembly whose function is to recognize the TATA box and specify the polarity of the pre-initiation complex (PIC) assembly.

TBP is composed of a variable N-terminus that is not required for viability in yeast (Cormack *et al.*, 1991; Gill & Tjian, 1991; Reddy & Hahn, 1991) and a conserved C-terminal domain (called TBPc) that binds directly to the TATA box by unwinding the DNA and straddling the wide-open underwound minor groove (Kim, Geiger *et al.*, 1993; Kim, Nikolov *et al.*, 1993). The bending of the DNA brings the upstream and the downstream flanking sequences closer together and may influence the binding of other transcription factors.

TFIIA functions as a target for activators of transcription, stabilizes the TBP–DNA complex and blocks the action of proteins which disrupt the TBP–DNA interaction (Maldonado & Reinberg, 1995; Ranish & Hahn, 1996; Wang *et al.*, 1992). Yeast TFIIA is composed of a highly conserved small subunit (SSU) and a large subunit (LSU). The LSU contains two conserved subdomains at each end of the polypeptide that flank a non-conserved middle region. A deletion of 113 amino acids from the middle region (TFIIAΔ113) still results in a functional molecule (Kang *et al.*, 1995).

TFIIB recognizes the TBP–DNA complex and the flanking DNA on either side of the TATA box and mediates the recruitment and positioning of the polymerase on the start site (Lee & Hahn, 1995; Orphanides *et al.*, 1996). TFIIB is a multidomain polypeptide with an N-terminal domain that varies among species and a conserved C-terminal domain (TFIIBc). The N-terminal domain is not required for TFIIB binding to DNA or to TBP (Barberis *et al.*, 1993; Hisatake *et al.*, 1993; Yamashita *et al.*, 1993).

Crystal structures of the TBPc–DNA–TFIIAΔ113 (Geiger *et al.*, 1996; Tan *et al.*, 1996) and TBPc–DNA–TFIIBc (Littlefield *et al.*, 1999; Tsai & Sigler, 2000) ternary complexes have shown that TFIIA binds to the surface formed by the upstream portion of the TBP–DNA complex, 5′ to the TATA box, and that TFIIB makes interactions with both TBP and with the DNA on either side of the TATA box.

Biochemical studies suggest that the conformation of the components in the quaternary complex differs from that seen in the two ternary complexes. In particular, contacts between TFIIB and the 5′ flanking region of the promoter become more extensive in the TFIIA–TFIIB–TBP–DNA complex than in complexes lacking TFIIA (Lagrange *et al.*, 1996, 1998). An examination of the intact quaternary complex will be required to fully

explain the specificity of the TBP, TFIIA and TFIIB interactions with the TATA-box promoter and to understand how the pseudo-symmetric TBP can direct the assembly of a PIC with defined polarity.

Here, we report the assembly, crystallization and preliminary X-ray diffraction studies of a yeast quaternary transcription-factor complex.

2. Materials and methods

2.1. Expression and purification

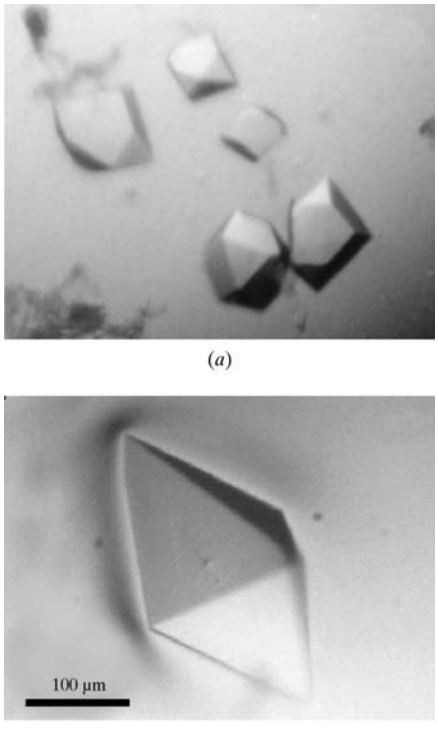
Yeast TBPc (residues 60–224) was purified in a manner analogous to that described for human TBPc (Juo *et al.*, 1996). Yeast TFIIAΔ113 was prepared essentially as described by Geiger *et al.* (1996). Yeast TFIIBc (residues 111–346) was overexpressed in *Escherichia coli* strain BL21(DE3) and purified by chromatography on S-Sepharose Fast Flow, Resource S and Superdex 75 columns (Pharmacia) in 20 m*M* HEPES buffer pH 7.9. The concentrations of yTBPc and yTFIIBc were determined by UV absorbance in 6 *M* guanidine–HCl pH 6.5. The concentration of yTFIIAΔ113 was determined by amino-acid analysis. Oligonucleotides were purchased from the Keck Oligonucleotide Synthesis Facility at Yale University and were detritylated and purified on a reversed-phase column (Dynamax-300Å PureDNA, Rainin). Complementary DNA strands were concentrated in a Centricon-3, exchanged into 50 m*M* Tris pH 7.6, 100 m*M* NaCl buffer by repeated concentration and dilution, mixed in equimolar amounts and annealed by heating to 368 K and slow cooling to room temperature.

The quaternary DNA–protein complex was prepared by mixing 11 nmol of each protein (22 nmol of yTFIIAΔ113 for form I crystals) and 10 nmol of DNA in the high-salt buffers in which they were purified in a Centricon-30 (Amicon) in the following order: yTFIIAΔ113, DNA, yTBPc, yTFIIBc. The complex was diluted with 5 m*M* Tris pH 7.5, 5% glycerol, 50 m*M* KCl, 5 m*M* DTT and exchanged into the same buffer by repeated concentration and dilution. The complex was concentrated to 10 mg ml^{-1}.

2.2. Crystal growth and analysis

Crystals were grown from hanging drops containing 2 µl of complex plus 2 µl of reservoir solution over 1 ml reservoirs at 291 K using the vapor-diffusion technique. Two crystal forms were obtained. For crystallization of form I, the reservoir contained 30% PEG 4000, 200 m*M* ammonium acetate and 100 m*M* sodium citrate pH 5.6. For crystallization of form II, the reservoir contained 20–30% PEG 4000, 100–150 m*M* ammonium citrate and 100 m*M* MES pH 5.6. Quaternary complexes prepared with a blunt-ended DNA duplex consisting of an eight base-pair TATA box flanked by nine base pairs upstream and one base pair downstream (5′-ATCGATCGA**TATAAA-AG**G-3′) yielded the best form I crystal. The best form II crystal was obtained using a blunt-ended oligoduplex with nine base-pair flankers both upstream and downstream of the eight base-pair TATA box (5′-CTGTA-TCG**TATAAAAG**TCTTGTTCG-3′). Crystals appeared in 2–14 d and reached full size in 2–6 weeks.

Crystals were stabilized by gradual equilibration with protein-free stabilizer solution (5% glycerol, 105% of the PEG 4000 and salt concentrations in the reservoir and 100 m*M* buffer). 0.5 µl of stabilizer solution was added to the hanging drop every 0.5 h for the first 3 h, after which 1 µl was added every 0.5 h. After 8 µl had been added, the droplets were flooded with 5 µl of stabilizer and the crystals were allowed to remain in this solution overnight. Cryostabilization was achieved in a similar manner. Every hour, 2 µl of stabilizer solution was removed and replaced with 2 µl of cryoprotectant solution (stabilizer plus 20–30% glycerol). A proportional amount of the reservoir solution was drawn off at the same time and replaced by cryoprotectant. After five additions (10 µl), 5 µl more was drawn off and 5 µl of cryoprotectant solution was added. Crystals were removed from the cryo-stabilizer solution within 1 h of stabilization with thin nylon loops (Hampton Research) and flash-cooled in a bath of liquid propane chilled to 100 K (form I crystals) or directly in liquid nitrogen (form II crystals).



(*a*)

(*b*)

Figure 1
Crystal images. (*a*) Quaternary complex crystal form I. (*b*) Quaternary complex crystal form II.

Table 1
Data-collection statistics.

Values in parentheses are for the last resolution shell.

	Form I	Form II
X-ray source	CHESS F1	NSLS X25
Detector	Princeton CCD 2K × 2K	Brandeis B4 CCD
Space group	*R*32	$P42_12$
Wavelength (Å)	0.908	1.1
Unit-cell parameters		
a (Å)	173.2	141.7
b (Å)	173.2	141.7
c (Å)	164.1	112.0
Mosaicity (°)	1.02	1.22
Resolution (Å)	50.0–2.5 (2.6–2.5)	99.0–3.6 (3.7–3.6)
Completeness (%)	90.4 (64.1)	98.4 (99.8)
$R_{\rm merge}$	0.059 (0.228)	0.081 (0.610)
$I/\sigma(I)$	15.3 (3.3)	22.9 (2.5)

A native data set for crystal form I was collected at the Cornell High Energy Synchrotron Source (CHESS) on beamline F1, which was equipped with a Princeton 2K × 2K CCD detector operating in the binned mode. For form II crystals, the best native data set was collected on a Brandeis B4 detector at beamline X25 at the National Synchrotron Light Source (NSLS). The temperature of the crystal was maintained at 100 K with a cold nitrogen-gas stream. Data reduction and scaling were performed using *HKL* version 1.7 or *HKL*2000 (Otwinowski & Minor, 1997).

3. Results and discussion

The length of the DNA duplex used in the assembly of the quaternary complex determined the crystal form obtained. 44 different oligoduplexes that varied in the length and ends of the flanker sequences on either side of the eight base-pair TATA box were screened for crystal growth. Form I crystals appeared only with two DNAs, both of which have a nine base-pair upstream flanker and either a one or zero base-pair downstream flanker. When the downstream

flanker was increased to two base pairs, form I crystals could no longer be obtained.

Form I crystals (Fig. 1) belong to space group $R32$ and diffracted to 2.5 Å (Table 1). The structure of the complex was solved using molecular replacement in combination with mercury, uranyl and bromine phases (D. T. Gewirth *et al.*, in preparation). Density for TFIIB was not seen in the electron-density map, suggesting that this protein was either disordered or present in substoichiometric amounts. In support of this, denaturing polyacrylamide-gel analysis of washed and dissolved form I crystals showed variable amounts of TFIIB (Fig. 2). Since the complete binding site for TFIIB is thought to include a minimum of seven base pairs of DNA on either side of the TATA box (Lee & Hahn, 1995), our result may reflect a lack of sufficient downstream flanker DNA in the oligoduplexes used to assemble the complex.

Form II crystals (Fig. 1) grew from quaternary complexes prepared with a variety of different TATA-box-containing DNAs and contained all the proteins (Fig. 2). The upstream flanker could vary in length from six to ten base pairs while, independently, the downstream flanker could vary from two to ten base pairs. DNAs longer than ten base pairs upstream or downstream of the TATA box did not yield any crystals. The fact that the same crystal form was obtained within a broad range of DNA duplex lengths (18–28 base pairs) suggests that, unlike the majority of protein–DNA complex structures, the ends of the DNA duplex are not involved in lattice contacts. The apparent upper limit to the DNA length that yields form II crystals suggests that at this point the ends of the DNA duplex were interacting with a symmetry-related complex.

Form II crystals, with average dimensions of 0.2 × 0.2 × 0.1 mm, showed a diffraction limit of approximately 8 Å both at home X-ray sources (Rigaku RU-200) and at synchrotrons (NSLS X25, CHESS F1). In the course of these studies, an unusually large (0.5 × 0.5 × 0.25 mm) form II crystal was grown and was mounted in a capillary directly from the mother liquor. Analysis of this crystal at NSLS beamline X25 showed diffraction to 5 Å and thus inspired a systematic effort to grow larger crystals. Large crystals could be reproducibly grown by routinely screening around the standard crystallization condition in 1% PEG 4000 increments and 50 mM ammonium citrate increments. Importantly, nucleation was limited by leaving the crystallization tray undisturbed for two weeks after setup.

X-ray analysis of these larger crystals, ranging from 0.35 × 0.35 × 0.15 to 0.6 × 0.45 × 0.2 mm showed reproducible diffraction to 4–5 Å using synchrotron radiation. The best crystal yielded a complete 3.6 Å data set at NSLS beamline X25 (Table 1). The space group is $P42_12$. Assuming one 96.7 kDa quaternary complex per asymmetric unit yields a V_M of 2.9 Å^3 Da^{-1} and a solvent content of 68%.

Large form II crystals often cracked or showed high mosaicity (2–4°) following stabilization and freezing, presumably owing to the stresses associated with physical manipulation of large high-solvent content crystals and non-uniform heat-transfer rates during cooling. This severely degraded the quality of the resulting diffraction data. In our hands, careful attention to slow stabilization protocols was the key to obtaining good diffraction data. Even with these precautions, however, only about one in five large crystals had the required combination of low mosaicity and large volume necessary to yield measurable diffraction better than 4 Å.

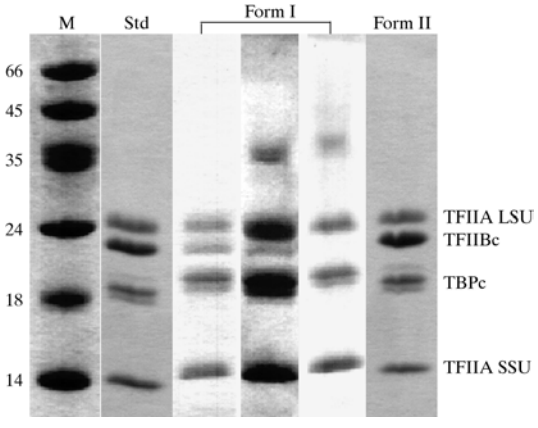


Figure 2
SDS–polyacrylamide gel analysis of quaternary complex crystals. Crystals were harvested, washed three times in reservoir solution and dissolved in buffer. Lane 1, molecular-weight standards. Lane 2, solution of complex used for crystallization. Lanes 3, 4 and 5, crystal form I showing variable amounts of TFIIB. Lane 6, crystal form II.

4. Conclusions

We have produced a moderately well diffracting crystal of the complete quaternary complex. A native data set has been collected and a derivative data set containing iodouracil-substituted DNA has also been collected and is currently being analyzed. Given the availability of good models for the protein components of the quaternary complex, it is likely that a 3.6 Å resolution electron-density map of the complex will be interpretable. The structure of the quaternary complex will provide insights into the nature of the macromolecular interactions that serve to orient and stabilize the first components of the transcription-preinitiation assembly.

In this paper we also demonstrate the feasibility of assembling and crystallizing a complex from four different polypeptides and a DNA duplex. To date, only one other assembly of similar complexity – the nucleosome core particle (Luger *et al.*, 1997) – has been reconstituted from purified components and crystallized. Such crystallization challenges are likely to be more common in the future.

We would like to thank J. H. Geiger and P. B. Sigler for advice and encouragement early on in this work. We also thank Drs D. Thiel and M. Szebenyi (CHESS F1) and Drs L. Berman, M. Capel and B. Sweet (NSLS) for access to and assistance in synchrotron data collection.

References

Barberis, A., Muller, C. W., Harrison, S. C. & Ptashne, M. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 5628–5632.
Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. (1989). *Cell*, **56**, 549–561.
Conaway, R. C. & Conaway, J. W. (1993). *Annu. Rev. Biochem.* **62**, 161–190.
Cormack, B. P., Strubin, M., Ponticelli, A. S. & Struhl, K. (1991). *Cell*, **65**, 341–348.
Geiger, J. H., Hahn, S., Lee, S. & Sigler, P. B. (1996). *Science*, **272**, 830–836.
Gill, G. & Tjian, R. (1991). *Cell*, **65**, 333–340.
Hisatake, K., Hasegawa, S., Takada, R., Nakatani, Y., Horikoshi, M. & Roeder, R. G. (1993). *Nature (London)*, **362**, 179–181.
Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J. & Dickerson, R. E. (1996). *J. Mol. Biol.* **261**, 239–254.
Kang, J. J., Auble, D. T., Ranish, J. A. & Hahn, S. (1995). *Mol. Cell. Biol.* **15**, 1234–1243.
Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993). *Nature (London)*, **365**, 520–527.
Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). *Nature (London)*, **365**, 512–520.
Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D. & Ebright, R. H. (1998). *Genes Dev.* **12**, 34–44.
Lagrange, T., Kim, T. K., Orphanides, G., Ebright, Y. W., Ebright, R. H. & Reinberg, D. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 10620–10625.
Lee, S. & Hahn, S. (1995). *Nature (London)*, **376**, 609–612.

Littlefield, O., Korkhin, Y. & Sigler, P. B. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 13668–13673.
Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. (1997). *Nature (London)*, **389**, 251–260.
Maldonado, E. & Reinberg, D. (1995). *Curr. Opin. Cell Biol.* **7**, 352–361.
Orphanides, G., Lagrange, T. & Reinberg, D. (1996). *Genes Dev.* **10**, 2657–2683.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Ranish, J. A. & Hahn, S. (1996). *Curr. Opin. Genet. Dev.* **6**, 151–158.
Reddy, P. & Hahn, S. (1991). *Cell*, **65**, 349–357.
Tan, S., Hunziker, Y., Sargent, D. F. & Richmond, T. J. (1996). *Nature (London)*, **381**, 127–151.
Tsai, F. T. & Sigler, P. B. (2000). *EMBO J.* **19**, 25–36.
Wang, W., Gralla, J. D. & Carey, M. (1992). *Genes Dev.* **6**, 1716–1727.
Yamashita, S., Hisatake, K., Kokubo, T., Doi, K., Roeder, R. G., Horikoshi, M. & Nakatani, Y. (1993). *Science*, **261**, 463–466.
Zawel, L. & Reinberg, D. (1993). *Prog. Nucleic Acid Res. Mol. Biol.* **44**, 67–108.