Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Purification, crystallization and initial crystallographic analysis of RNA polymerase holoenzyme from *Thermus thermophilus*

Marina N. Vassylyeva,[a,b] Jookyung Lee,[c] Shun-ichi Sekine,[b,d] Oleg Laptenko,[c] Seiki Kuramitsu,[a,b,e] Takehiko Shibata,[f] Yorinao Inoue,[b] Sergei Borukhov,[c]* Dmitry G. Vassylyev[b,d]* and Shigeyuki Yokoyama[a,b,d,g]*

[a]Genomic Sciences Center (RIKEN), 2-1 Hirosawa, Wako, Saitama 351-0198, Japan, [b]Structurome Project, RIKEN Harima Institute at SPring-8, 1-1-1 Kouto, Mikazuki-cho, Sayo, Hyogo 679-5148, Japan, [c]SUNY HSC at Brooklyn, Department of Microbiology and Immunology, 450 Clarkson Avenue, BSB 3-27, Brooklyn, NY 11203, USA, [d]Cellular Signaling Laboratory, RIKEN Harima Institute at SPring-8, 1-1-1 Kouto, Mikazuki-cho, Sayo, Hyogo 679-5148, Japan, [e]Department of Biology, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan, [f]Laboratory of Cellular and Molecular Biology, RIKEN, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan, and [g]Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Bukyo-ku, Tokyo 113-0033, Japan

Correspondence e-mail: serbor@aol.com, dmitry@yumiyoshi.harima.riken.go.jp, yokoyama@biochem.s.u-tokyo.ac.jp

Received 22 February 2002
Accepted 3 July 2002

RNA polymerase holoenzyme from *Thermus thermophilus*, consisting of six protein subunits (α_2, β, β', ω and σ^{70}) and having a total molecular mass of about 450 kDa, was purified and crystallized by the hanging-drop vapour-diffusion technique under mild near-physiological conditions. The crystals diffract beyond 3 Å resolution. Careful analysis of diffraction data revealed that the crystals belong to space group $P3_2$, with unit-cell parameters $a = b = 236.35$, $c = 249.04$ Å, and have perfect twinning along the threefold axis. A complete data set at 3 Å resolution was collected and an unambiguous molecular-replacement solution was found using the structure of *T. aquaticus* RNA polymerase core enzyme as a search model. The refinement of structure and model building of the σ^{70} subunit is now in progress.

1. Introduction

Bacterial DNA-dependent RNA polymerase (RNAP) is a multifunctional enzyme responsible for the synthesis of all RNAs in the cell (mRNA, rRNA and tRNA) and serves as a target for numerous regulatory factors. The catalytically competent core enzyme (MW $\simeq$ 400 kDa) consists of five evolutionarily conserved subunits: α-dimer (α_2), β, β' and ω (Archambault & Friesen, 1993; Zhang *et al.*, 1999; Cramer *et al.*, 2000; Minakhin, Bhagat *et al.*, 2001; Minakhin, Nechaev *et al.*, 2001). The core RNAP is capable of elongation and termination of transcription. However, for specific and efficient initiation it requires association with transcription-initiation factor σ, to form a holoenzyme which can bind to a specific promoter DNA sequence (Gross *et al.*, 1992). Promoters recognized by the major housekeeping σ-factor σ^{70} are typically defined by essential DNA sequences near nucleotide positions -35 and -10 relative to the transcription start site (+1) (Hawley & McClure, 1983; Dombroski *et al.*, 1992). As a part of the holoenzyme, σ is responsible not only for binding to the promoter, but also for melting the DNA duplex in the vicinity of the -10 region (Juang & Helmann, 1994) and interacting with the emerging 5′-terminus of the nascent RNA (Severinov *et al.*, 1994). It may also play a role in the transition of the initiation complex to a stable elongation complex (TEC; Hernandez *et al.*, 1996).

To understand the molecular mechanism of transcription and its regulation, it is essential to discern the structure and function of RNAP holoenzyme at the atomic level. The three-dimensional crystal structure of the core RNAP from a thermophilic organism, *Thermus aquaticus*, has been recently obtained at modest resolution (Zhang *et al.*, 1999). The X-ray structure of a fragment of *Escherichia coli* σ^{70} factor, containing several conserved regions, was reported previously (Malhotra *et al.*, 1996). However, a high-resolution structure of holoenzyme is not yet available and most information about σ–core and σ–DNA interactions has been obtained through biochemical, biophysical and genetic studies (Sharp *et al.*, 1999; Callaci *et al.*, 1999; Burgess & Anthony, 2001). We now report the purification, crystallization and preliminary crystallographic analysis of RNAP holoenzyme from the thermophilic organism *T. thermophilus*.

2. Experimental procedures and results

2.1. Isolation and purification

Purification of RNAP from several thermophilic bacteria has been described previously (Wnendt *et al.*, 1990; Zhang *et al.*, 1999). Unlike the core, the purification of the holoenzyme poses difficulties owing to the unstable σ-core association and its susceptibility to proteolysis. Several attempts to isolate native *T. thermophilus* RNAP holoenzyme have been undertaken in the past (Date *et al.*, 1975; Tsuji *et al.*, 1981; Xue *et al.*, 2000); however, all preparations described appeared to be heterogeneous. Here, we present an improved procedure for obtaining electrophoretically homogeneous RNAP holoenzyme and core from *T. thermophilus*.

The major steps for holoenzyme purification are illustrated in Fig. 1. *T. thermophilus* cells were grown as described in Wnendt *et al.*

(1990). 180 g of fresh cells were suspended in 1 l of lysis buffer [40 m*M* Tris–HCl pH 7.7, 0.1 *M* NaCl, 10 m*M* EDTA, 10 m*M* 2-mercaptoethanol (2-ME), 0.1 m*M* PMSF], disrupted by sonication and centrifuged at 15 000*g*. The supernatant (initial crude cell lysate; Fig. 1, left panel, lane 1) was precipitated by Polymine P (0.5% final concentration), the pellet was washed twice with 1 l of lysis buffer containing 0.2 *M* NaCl and extracted with 0.6 l of the same buffer containing 0.8 *M* NaCl. The extract (lane 2) was precipitated with ammonium sulfate (at 45% saturation), the pellet collected by centrifugation, redissolved in 1 l of buffer *A* (20 m*M* Tris–HCl pH 7.7, 1 m*M* EDTA, 5 m*M* 2-ME, 5% glycerol) and centrifuged as above. The resulting material was applied to a 300 ml SP-Sepharose CL6B column (5 × 15 cm) at 5 ml min^{-1}, the column was washed with 1.5 l of buffer *A* and eluted with a linear NaCl gradient (0–0.8 *M*) in 3 l of buffer *A*. This chromatographic step is crucial, since it allows separation of holoenzyme from core and facilitates the removal of most cellular proteases. The material eluting at ∼0.1 *M* NaCl (∼300 ml) contained ∼80% pure core. It was further fractionated, as described for *E. coli* RNAP (Orlova *et al.*, 1995), by a combination of affinity (heparin-Sepharose), size-exclusion and anion-exchange FPLC, yielding ∼35 mg of electrophoretically homogeneous (>95% pure) core protein (Fig. 1, right panel, lane 1). The material eluting from SP-Sepharose at ∼0.24 *M* NaCl (∼200 ml) contained ∼50% pure holoenzyme (Fig. 1, left panel, lane 3). It was successively fractionated by anion-exchange FPLC on Mono Q 10HR (lane 4), size-exclusion FPLC on Superdex 200HR 16/60 (lane 5) and cation-exchange FPLC on Mono S 10HR as described in Zhang *et al.* (1999). This procedure yielded ∼13 mg of electrophoretically homogeneous (>95% pure) holoenzyme (Fig. 1, right panel, lane 2).

Densitometric analysis of the stained polyacrylamide gel revealed that the purified holoenzyme contains stoichiometric amounts of α, β, β', σ and ω subunits in the expected molar ratio of 2:1:1:1:1. The σ (MW = 48.5 kDa; Nishiyama *et al.*, 1999) and α (MW = 35 kDa; Wada *et al.*, 1999) subunits displayed aberrant electrophoretic mobilities in both Tris–tricine and Tris–glycine buffer systems, with apparent MW ∼67/55 kDa and ∼45/40 kDa, respectively. However, the electrophoretic mobilities of the native and recombinant (SigA; Nishiyama *et al.*, 1999) σ factors were indistinguishable (Fig. 1, right panel, lane 3). The mobility of three other protein bands corresponded to their expected MW of 171 kDa for β', 125 kDa for β and 10 kDa for ω. The identities of all RNAP subunits were confirmed by Edman amino-acid sequence analysis (ten N-terminal residues of each protein were determined).

Purified core and holoenzyme were tested for their ability to initiate transcription and to form a stable TEC on a 200 bp phage T7 A1 promoter DNA fragment as described in Borukhov *et al.* (1991). We found that at 333 K the *T. thermophilus* holoenzyme, but not core, can efficiently recognize T7A1 promoter and form TEC carrying a radiolabelled 20-meric transcript (20A-TEC) (data not shown). The activity displayed by *T. thermophilus* holoenzyme was similar to that of the *E. coli* RNAP at 310 K.

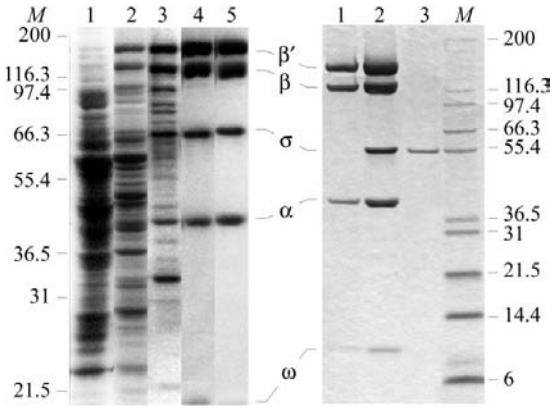

Figure 1
SDS–PAGE analysis of RNAP fractions. 8% gel in Tris–glycine system (left panel) and 10% gel in Tris–tricine system (right panel) were used, followed by Coomassie Blue staining. Marker proteins (Novex-Invitrogen) are shown in lanes *M* with their molecular weights indicated in kDa. From left to right: aliquots corresponding to 1/1 000 000, 1/30 000, 1/10 000, 1/4 000, 1/2 000, 1/10 000 and 1/3 000 of the total initial material were loaded on the gel in lanes 1, 2, 3, 4 and 5 of the left panel and in lanes 1 and 2 of the right panel, respectively. Lane 3 in the right panel shows purified recombinant SigA protein of *T. thermophilus*.

2.2. Crystallization and data collection

Hampton Research Crystal Screen (Jancarik & Kim, 1991) was used to determine the initial crystallization conditions for RNAP holoenzyme. Crystallization was carried out by the hanging-drop vapour-diffusion technique at 293 K and the Crystal Screen precipitant solutions were diluted sixfold with water in the initial screening. 2 µl of these diluted precipitant solutions were added to 2 µl drops containing protein solution at a protein concentration of 10 mg ml^{-1}. Very thin hexagonal crystals were grown from 35 m*M* of magnesium formate within 4 d. These crystals were subjected to microseeding using hanging drops prepared under the same conditions. Microseeding resulted in a large number of small but well shaped three-dimensional hexagonal crystals, which were further enlarged to about 0.3 × 0.3 × 0.2 mm in size by macroseeding. At this stage, the crystals diffracted beyond 4 Å using synchrotron radiation and a full 4 Å data set was collected at SPring-8 BL44 beamline using a MAR CCD detector. The crystals were further improved by varying the magnesium formate concentrations, pH values and precipitants. The final condition, 13 m*M* magnesium formate, 5% PEG 400, 2 m*M* spermine, 2 m*M* DTT and 20 m*M* MES buffer pH 5.8, yielded high-quality hexagonal crystals with unit-cell parameters $a = b = 236.35$, $c = 249.04$ Å (Fig. 2). They diffracted beyond 3 Å at the synchrotron X-ray beamlines. Diffraction data were collected at 100 K from a native crystal at SPring-8 BL45 beamline using a Rigaku R-AXIS V imaging-plate detector. The mother-liquor solution with 30% 2-methyl-2,4-pentanediol was used as a cryoprotectant. Although the crystals typically had high mosaicity (∼1°), the large active area of the R-AXIS V detector (40 × 40 cm) enabled us to resolve the spots beyond 2.5 Å resolution.

2.3. Data processing and initial analysis

All data processing was performed using the *HKL*2000 program package (Otwinowski & Minor, 1997). The data were

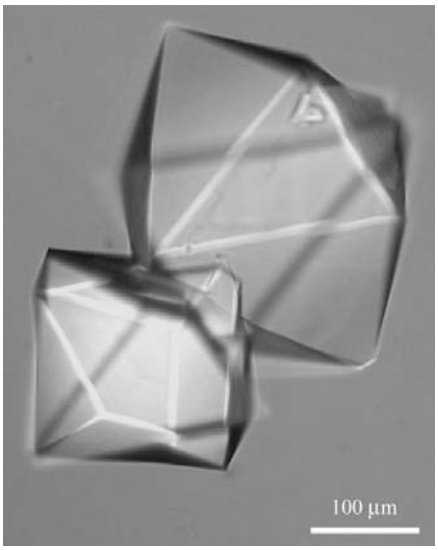

Figure 2
Crystals of RNAP holoenzyme.

Table 1
Molecular-replacement solution in space group $P6_5$ for RNAP holoenzyme.

Rotation search†		Translation search‡			
	Rotation function (σ units)		Correlation coefficients	Translation function (σ units)	R factor (%)
1st peak	6.1	1st peak	23.9	20.2	47.9
		2nd peak	15.8	8.2	49.9
2nd peak	3.8	1st peak	7.7	4.1	52.0

† Resolution range 20.0–4.0 Å, 50.0 Å search radius. ‡ Resolution range 8.0–4.0 Å.

Table 2
Data collection and rigid-body refinement statistics in space group $P3_2$ for RNAP holoenzyme.

Data collection. Values in parentheses indicate data in the highest resolution shell.

Space group	$P3_2$
Wavelength used (Å)	1.02
Resolution range (Å)	40.0–3.0 (3.23–3.0)
Total observations	727451
Unique reflections	290411
Completeness (%)	93.3 (87.7)
Mean $I/\sigma(I)$	7.2 (2.7)
$R_{\rm merge}$† (%)	10.0 (32.5)

Rigid-body refinement.

	Trial 1	Trial 2
Resolution range (Å)	40.0–3.0	40.0–3.0
Twinning fraction (%)	0	50
Initial $R_{\rm free}$‡ (%)	50.0	48.6
$R_{\rm free}$, molecules (%)	49.0	43.2
$R_{\rm free}$, subunits (%)	47.5	41.4
$R_{\rm free}$, domains (%)	45.3	38.7

† $R_{\rm merge} = \sum_{hkl}\sum_j |I_j(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_j \langle I(hkl)\rangle$, where $I_j(hkl)$ and $\langle I(hkl)\rangle$ are the intensity of measurement j and the mean intensity for the reflection with indices hkl, respectively. ‡ $R_{\rm free} = \sum_{hkl} ||F_{\rm calc}(hkl)| - |F_{\rm obs}(hkl)|| / \sum_{hkl} |F_{\rm obs}|$, where the crystallographic R factor is calculated over reflections excluded from refinement. The free reflections constituted 5% of the total number of reflections.

processed at 3 Å resolution. Based on the statistical analysis of data provided within *HKL*2000 ($R_{\rm merge}$ = 11.3%), the apparent space group for RNAP holoenzyme crystals was one of the hexagonal space groups. The model of the closely related *T. aquaticus* core enzyme (PDB entry 1i6v) was used to find a molecular-replacement solution using the *AMoRe* program (Navaza, 2001). Both rotation and translation functions showed clear single prominent peaks (Table 1). The translation-function solution unambiguously defined the space group as $P6_5$.

The additional statistical analysis of data to check for the presence of merohedral twinning (Yeates, 1997) as implemented in the program *CNS* (Brünger *et al.*, 1998) showed $\langle I^2\rangle/\langle I\rangle^2 = 1.52$ and $\langle F^2\rangle/\langle F\rangle^2 = 0.87$ in the resolution range 40–3 Å, which are characteristic of perfectly twinned crystals. At the same time, the only possible merohedral twinning for the $P6_5$ space group (Yeates, 1997) with twinning operator (k, h, $-l$) was shown to be negligible by the *CNS* program (Brünger *et al.*, 1998). Given the unambiguous molecular-replacement solution derived in space group $P6_5$, we can account for the observed twinning only if we assume the following: (i) the crystals actually belong to the $P3_2$ space group, (ii) they have perfect twinning with twinning operator ($-h$, $-k$, l) (180° rotation around the threefold axis) and (iii) the crystals contain two molecules in the asymmetric unit with non-crystallographic symmetry, which closely resembles the $P6_5$ space group. To check this unusual possibility, the data were reprocessed in space group $P3_2$ (Table 2). Rigid-body refinement was performed with the *CNS* program (Brünger *et al.*, 1998) in space group $P3_2$ for data in the resolution range 40–3 Å using two molecules of *T. aquaticus* RNAP core enzyme in the asymmetric unit, with their initial orientations corresponding to the crystallographic $P6_5$ symmetry (y, $-x + y$, $z + 1/6$). Three steps of rigid-body refinement were carried out for the whole molecules, then subunits and finally domains. In the case when no twinning operator was applied, the refinement converged from 50.0% to only 45.3% and the two molecules in the asymmetric unit strictly maintained $P6_5$ symmetry even after domain refinement. In contrast, when perfect twinning (twinning fraction = 0.5) was applied to the data, the refinement converged from 48.6 to 38.7%, which is 6.3% better than that of the untwinned refinement result (Table 2). In the case of perfect twinning, the refinement resulted in a substantial deviation from the initial $P6_5$ crystallographic symmetry. The rotational components of this deviation range from 1 to 4.5° for different subunits/domains of the protein. This is consistent with the assumption of only non-crystallographic symmetry in the case of perfectly twinned data in space group $P3_2$. Additional evidence for the identity of $P3_2$ as the actual space group for the holoenzyme crystal and for the presence of non-crystallographic symmetry comes from the analysis of systematic absences. Reflections with indexes (0, 0, l) were checked. Indeed, all reflections with $l = 6n$ exhibit strong intensities (space group $P6_5$). Much weaker, but detectable, intensities with $I/\sigma(I)$ of about 3–4 are observed for a number of reflections with $l = 3n$ (space group $P3_2$). No valid intensities (above the 2σ level) could be found for the other (0, 0, l) reflections. This result is also consistent with the crystals belonging to the $P3_2$ space group with two molecules in the asymmetric unit and having non-crystallographic symmetry which is close to the crystallographic $P6_5$ symmetry. To check whether the crystals are perfectly twinned, the rigid-body refinement was repeated with different twinning fractions (0.35, 0.40 and 0.45). In all these cases the refinement results were substantially worse than those obtained for the twinning fraction 0.5 (perfect twinning). Therefore, we conclude that the crystals belong to space group $P3_2$ with perfect twinning along the threefold axis. The amino-acid sequence for RNAP holoenzyme is now available from the *T. thermophilus* genomic project (Kuramitsu, Yokoyama, Shibata, Inoue, unpublished results) and refinement of the structure at 3 Å resolution is under way.

The authors thank Drs Adachi and Yamamoto for assistance during data collection at the SPring-8 synchrotron beamlines BL44 and BL45, respectively. We are grateful to Dr Yeates for discussion and helpful advice concerning the merohedral twinning problem. This work was supported in part by grant GM54098-06A1 from NIH, USA to SB and by a grant from the Organized Research Combination System of Science and Technology Agency (Japan) to SY.

References

Archambault, J. & Friesen, J. D. (1993). *Microbiol. Rev.* **57**, 703–724.
Borukhov, S., Lee, J. & Goldfarb, A. (1991). *J. Biol. Chem.* **266**, 23932–23935.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Burgess, R. R. & Anthony, L. (2001). *Curr. Opin. Microbiol.* **4**, 126–131.
Callaci, S., Heyduk, E. & Heyduk, T. (1999). *Mol. Cell*, **3**, 229–238.
Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., Thompson, N. E., Burgess, R. R., Edwards, A. M., David, P. R. & Kornberg, R. D. (2000). *Science*, **288**, 640–649.
Date, T., Suzuki, K. & Imahori, K. (1975). *J. Biochem.* **78**, 845–858.
Dombroski, A. J., Walter, W. A., Record, M. T. Jr, Siegele, D. A. & Gross, C. A. (1992). *Cell*, **70**, 501–512.
Gross, C. A., Lonetto, M. A. & Losick, R. (1992). *Transcriptional Regulation*, edited by K. Yamamoto & S. McKnight, pp. 129–176. New York: Cold Spring Harbor Laboratory Press.
Hawley, D. K. & McClure, W. R. (1983). *Nucleic Acids Res.* **11**, 2237–2255.

Hernandez, V. J., Hsu, L. M. & Cashel, M. (1996). *J. Biol. Chem.* **271**, 18775–18779.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Juang, Y.-L. & Helmann, J. (1994). *J. Mol. Biol.* **235**, 1470–1488.
Malhotra, A., Severinova, E. & Darst, S. (1996). *Cell*, **87**, 127–136.
Minakhin, L., Bhagat, S., Brunning, A., Campbell, E. A., Darst, S. A., Ebright, R. H. & Severinov, K. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 892–897.
Minakhin, L., Nechaev, S., Campbell, E. A. & Severinov, K. (2001). *J. Bacteriol.* **183**, 71–76.
Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.
Nishiyama, M., Kobashi, N., Tanaka, K., Takahashi, H. & Tanokura, M. (1999). *FEMS Microbiol. Lett.* **172**, 179–186.
Orlova, M., Newlands, J., Das, A., Goldfarb, A. & Borukho, S. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 4596–600.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Severinov, K., Fenyo, D., Severinova, E., Mustaev, A., Chait, B. T., Goldfarb, A. & Darst, S. A. (1994). *J. Biol. Chem.* **269**, 20826–20828.
Sharp, M. M., Chan, C. L., Lu, C. Z., Marr, M. T., Nechaev, S., Merritt, E. W., Severinov, K., Roberts, J. W. & Gross, C. A. (1999). *Genes Dev.* **13**, 3015–3026.
Tsuji, S., Imahori, K. & Nonomura, Y. (1981). *J. Biochem. (Tokyo)*, **89**, 1903–1912.
Wada, T., Yamazaki, T., Kuramitsu, S. & Kyogoku, Y. (1999). *J. Biochem. (Tokyo)*, **125**, 143–150.
Wnendt, S., Hartman, R. K., Ulbrich, N. & Erdman, V. A. (1990). *Eur. J. Biochem.* **191**, 467–472.
Xue, Y., Hogan, B. P. & Erie, D. A. (2000). *Biochemistry*, **39**, 14356–14362.
Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.
Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. & Darst, S. A. (1999). *Cell*, **98**, 811–824.