Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Crystallization and preliminary X-ray crystallographic analysis of *Thermus thermophilus* transcription elongation complex bound to Gfh1

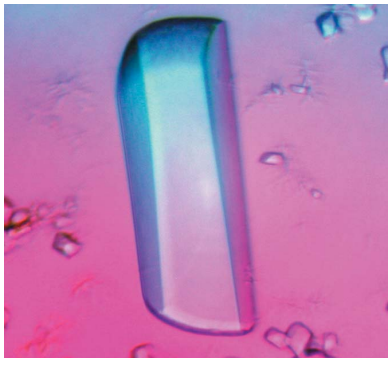
Shunsuke Tagami,[a,b] Shun-ichi Sekine,[a,b]* Thirumananseri Kumarevel,[c]‡ Masaki Yamamoto[d] and Shigeyuki Yokoyama[a,b,c]*

[a]Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan, [b]RIKEN Syetems and Structural Biology Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan, [c]Structural and Molecular Biology Laboratory, RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan, and [d]SR Life Science Instrumentation Unit, RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan

‡ Present address: Biometal Science Laboratory, RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan.

Correspondence e-mail: sekine@biochem.s.u-tokyo.ac.jp, yokoyama@biochem.s.u-tokyo.ac.jp

Received 16 October 2009
Accepted 18 November 2009



RNA polymerase (RNAP) elongates RNA by iterative nucleotide-addition cycles (NAC). A specific structural state (or states) of RNAP may be the target of transcription elongation factors. Gfh1, a *Thermus thermophilus* Gre-family protein, inhibits NAC. To elucidate which RNAP structural state Gfh1 associates with, the *T. thermophilus* RNAP elongation complex (EC) was cocrystallized with Gfh1. Of the 70 DNA/RNA scaffolds tested, two (for EC1 and EC2) were successfully crystallized. In the presence of Gfh1, EC1 and EC2 yielded crystals belonging to space group $P2_1$ with similar unit-cell parameters (crystals 1 and 2, respectively). X-ray diffraction data sets were obtained at 3.6 and 3.8 Å resolution, respectively.

1. Introduction

RNA polymerase (RNAP) elongates RNA by repetitive additions of a nucleotide to the 3′-end of the growing RNA (nucleotide-addition cycles or NAC). The transcribing RNAP (the elongation complex or EC) incorporates the substrate nucleoside triphosphate (NTP) into its binding site adjacent to the RNA 3′-end. The nucleoside monophosphate moiety of NTP is transferred to the RNA 3′-end, while the pyrophosphate moiety is released. This chemical step extends the RNA by one nucleotide residue, generating the 'pre-translocation state' with the RNA 3′-end in the NTP-binding site. The RNAP then translocates downstream along the DNA and RNA by one base step to form the 'post-translocation state' with an empty site for the next incoming NTP. Accumulating crystallographic studies on *Saccharomyces cerevisiae* RNAP II and *Thermus thermophilus* RNAP have revealed the structures of ECs in the pre-translocation state and the post-translocation state and in the latter incorporating an NTP (insertion and pre-insertion complexes) (Gnatt *et al.*, 2001; Kettenberger *et al.*, 2004; Westover *et al.*, 2004; Wang *et al.*, 2006; Vassylyev, Vassylyeva, Perederina *et al.*, 2007; Vassylyev, Vassylyeva, Zhang *et al.*, 2007).

During RNA elongation, various transcriptional regulators associate with EC and accurately control the RNAP activity, probably by influencing a step or steps in the NAC. Nevertheless, the EC conformational state or states in NAC that are the targets of these regulators and the mechanism by which they modulate the RNAP activity are obscure. Gfh1 is a protein that is found in the *Thermus* genus and is closely related to the transcript cleavage stimulation factor GreA. Gfh1 binds to the EC or the initiation complex and inhibits nucleotide addition and its reverse (pyrophosphorolysis) as well as transcript cleavage (Hogan *et al.*, 2002; Laptenko *et al.*, 2006). To determine crystallographically which EC structural state Gfh1 binds to, we surveyed cocrystallization conditions of RNAP, DNA, RNA and Gfh1 (EC–Gfh1). We designed a series of nucleic acid scaffolds for *in vitro* reconstitution of ECs and they were tested for crystallization. In this report, crystallization, data collection and preliminary X-ray crystallographic analyses are described.

2. Materials and methods

2.1. Preparation of RNAP

The RNAP core enzyme was purified from *T. thermophilus* cells as described previously (Vassylyeva *et al.*, 2002). Briefly, the RNAP components in the *T. thermophilus* cell lysate were fractionated by polyethyleneimine precipitation followed by ammonium sulfate precipitation. The endogenous RNAP core enzyme with no affinity tag was purified by conventional column chromatography on SP-Sepharose, Heparin-Sepharose and Superdex 200 pg columns (GE Healthcare Biosciences). The RNAP was dissolved in 20 m*M* Tris–HCl buffer pH 7.7 containing 500 m*M* NaCl, 5% glycerol, 10 m*M* 2-mercaptoethanol and 1 m*M* EDTA and the solution was concentrated to 45 mg ml^{-1}.

2.2. Cloning, expression and purification of Gfh1

The gene encoding the transcription regulatory protein Gfh1 (TTHA0793) from *T. thermophilus* was amplified from genomic DNA by PCR using the primers 5′-TATCATATGGCTCGCGAGGTGAAGCTCAC-3′ and 5′-TATGGATCCTTATTAACCGTGGATGGCCACCACCCGGAACT-3′. The PCR fragment was digested with *Nde*I and *Bam*HI and cloned into the pET-11a (+) expression vector (Novagen). The resultant plasmid was transformed into *Escherichia coli* BL21-CodonPlus (DE3)-RIL-X (Stratagene) strain and the Gfh1 protein was overexpressed in LB medium (1% polypeptone, 0.5% yeast extract and 1% NaCl) at the mid-log phase by the addition of IPTG to a final concentration of 1 m*M*. Harvested cells (59 g) were suspended in buffer *A* (20 m*M* Tris–HCl buffer pH 8.0 containing 50 m*M* NaCl) and disrupted by sonication for 10 min. An equal volume of preheated buffer *A* at 343 K was added to the lysate and this solution was heat-treated at 343 K for 10 min in order to denature most of the nonthermophilic contaminant proteins. After centrifugation at 40 000 rev min^{-1} for 60 min at 277 K, the supernatant was subjected to ammonium sulfate precipitation (20 m*M* Tris–HCl buffer pH 8.0 containing 1.2 *M* ammonium sulfate) and the pellet was resuspended and dialyzed against buffer *A*. The dialyzed sample was then loaded onto an anion-exchange column (Toyopearl SuperQ-650M, Tosoh) pre-equilibrated with buffer *A*. The bound protein was eluted with a sodium chloride gradient (0.15–1.5 *M*) and most of the Gfh1 protein eluted at 0.35 *M*. After desalting on a HiPrep desalting column (GE Healthcare Biosciences), the protein-containing fractions were pooled and applied onto a ResourceQ column (GE Healthcare Biosciences). The bound protein was eluted with a sodium chloride gradient (0.01–0.5 *M*) and most of the sample eluted at 175 m*M*. The Gfh1-containing fractions were dialyzed against 10 m*M* sodium phosphate buffer pH 7.0 containing 0.15 *M* NaCl and then applied onto a hydroxyapatite column (CHT10; GE Healthcare Biosciences) to remove additional contaminants. The bound protein was eluted with a sodium phosphate gradient (0.01–0.5 *M*) and most of the sample eluted at 20 m*M* sodium phosphate. The eluted fractions were pooled, concentrated and chromatographed on a gel-filtration column (Superdex 75, GE Healthcare Biosciences), which was pre-equilibrated with 20 m*M* Tris–HCl buffer pH 8.0 containing 0.15 *M* NaCl. The homogeneity of the final purified protein was over 99% as estimated by SDS–PAGE (Fig. 1). A total of 36.8 mg protein was purified from 59 g of cells and the solution was concentrated to 10.7 mg ml^{-1} in 20 m*M* Tris–HCl buffer pH 8.0 containing 0.15 *M* NaCl for crystallization studies.

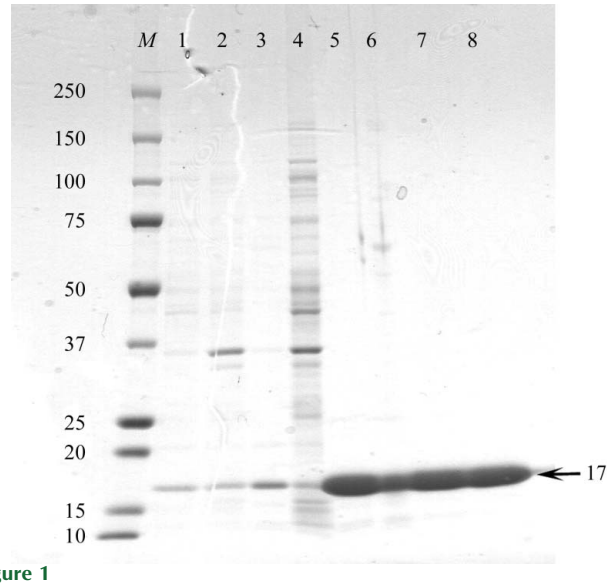


Figure 1
Purification of the Gfh1 protein. SDS–PAGE of the purified Gfh1 protein. Lane *M*, molecular-mass markers (all values are in kDa); lane 1, supernatant after sonication; lane 2, precipitate after sonication; lane 3, supernatant after heat treatment; lane 4, precipitate after heat treatment; lane 5, pooled fractions from Toyopearl SuperQ column; lane 6, pooled fractions from ResourceQ column; lane 7, pooled fractions from hydroxyapatite column; lane 8, pooled fractions from gel-filtration column.

2.3. Nucleic acid constructs and EC assembly

For the *in vitro* reconstitution of ECs, the nucleic acids were designed to contain a nine-base-pair DNA/RNA hybrid with downstream DNA and upstream RNA. We tested various nucleic acid scaffolds for the cocrystallization of EC–Gfh1. Of the 70 scaffolds examined, two scaffolds (RNA1/TDNA/NDNA and RNA2/TDNA/NDNA) yielded well diffracting crystals suitable for data collection (Fig. 2). The sequences of the oligomers are as follows. RNA oligomers: RNA1, CCCCGGAAGAUCAUCUUCCGGGGGAUGCGGCGG; RNA2, CCCGGAAGAUCAUCUUCCGGGGGAUGCGGCGG. DNA oligomers: template DNA (TDNA), GGTCTGTATCACGAGCCACCGCCGCAT; nontemplate DNA (NDNA), CGTGATACAGACC. PAGE-purified synthetic RNAs were purchased from Dharmacon. DNAs were purchased from Sigma–Genosys. Each nucleic acid was dissolved in water to a concentration of 500 µ*M*.

Generally, the nucleic acid scaffold was assembled by mixing 40 µ*M* RNA, template DNA and nontemplate DNA in 10 m*M* Tris–HCl buffer pH 7.7 and incubating the mixture at 343 K for 5 min followed by slow cooling to 293 K. For the preparation of EC, 25 µ*M* of each nucleic acid scaffold was mixed with 20 µ*M* RNAP in 10 m*M* Tris–HCl buffer pH 7.7 containing 150 m*M* NaCl and 1% glycerol and incubated at 293 K for 30 min. The formation of ECs was confirmed by native PAGE (not shown). The successful complexes, EC1 and EC2, were assembled using the scaffolds RNA1/TDNA/NDNA and RNA2/TDNA/NDNA, respectively (Fig. 2). The results of PAGE on a 5% native gel, showing the formation of EC1 and EC2, are shown in Fig. 3. The nucleic acids were stained with SYBR Gold (Molecular Probes) and the proteins were stained with SimplyBlue SafeStain (Invitrogen).

Figure 2
Schematic drawings of the nucleic acid scaffolds for EC1 and EC2.

2.4. Crystallization

For the cocrystallization of EC and Gfh1, 20 μ*M* of each EC (EC1, EC2 or the other ECs tested) was mixed with 60 μ*M* Gfh1 in 10 m*M* Tris–HCl buffer pH 7.7 containing 150 m*M* NaCl and 1% glycerol and the mixture was incubated at 293 K for 30 min. For the screening of initial crystallization conditions, the Crystal Screen, Crystal Screen II, Natrix (Hampton Research) and Wizard I and II (Emerald BioSystems) kits were used. To avoid the dissociation of EC, conditions with high salt concentrations and/or extreme pH were excluded. Although an acidic pH is preferable for Gfh1 to bind RNAP (Laptenko *et al.*, 2006), the screening included not only acidic conditions but also weakly alkaline conditions. Crystallization was performed by the sitting-drop vapour-diffusion technique at 293 K, with a starting drop size of 2 μl (1 μl sample solution plus 1 μl reservoir solution). The volumes of the reservoir solutions for equilibration were 80 and 500 μl for the first screening and optimization, respectively.

2.5. RNA-elongation and pyrophosphorolysis assays

For the RNA-extension and pyrophosphorolysis assays, 5 μ*M* of the nucleic acid scaffold was mixed with an excess amount (10 μ*M*) of RNAP so that all of the scaffold could complex with the RNAP. For RNA extension, 2.5 μ*M* EC was incubated with 10 μ*M* NTP at 328 K for 10 min in 40 m*M* Tris–HCl buffer pH 8.0 containing 50 m*M* NaCl, 10 m*M* $MgCl_2$ and 5 m*M* 2-mercaptoethanol. The reaction was stopped by quenching on ice and adding five volumes of buffer containing 10 m*M* EDTA. For the pyrophosphorolysis reaction, the NTP in the above reaction was replaced by 500 μ*M* pyrophosphate (PP_i). Nucleic acids were extracted with a 1:1 mixture of phenol and chloroform. They were separated by PAGE on a 20% gel containing 8 *M* urea and stained with SYBR Gold (Molecular Probes).

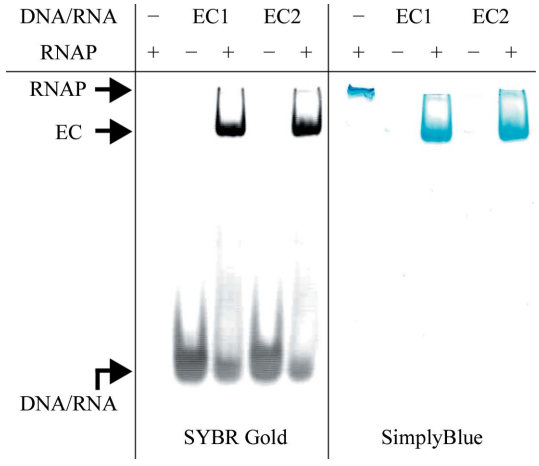


Figure 3
Reconstruction of EC1 and EC2. RNAP, nucleic acids and ECs were separated by native PAGE on a 5% gel. RNAP is stacked at the top of the gel. The same gel was stained with SYBR Gold and then with SimplyBlue SafeStain to visualize nucleic acids and proteins, respectively.

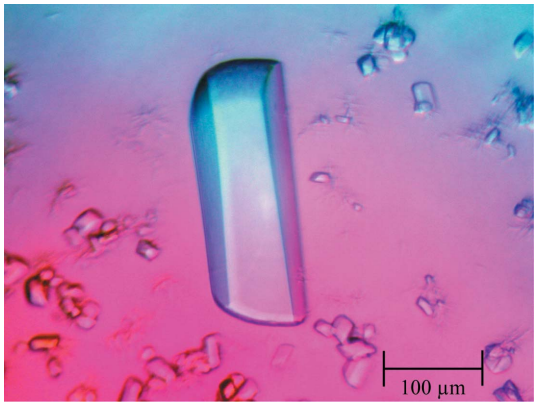


Figure 4
A crystal of EC–Gfh1. A typical crystal of EC1–Gfh1 obtained by the sitting-drop vapour-diffusion technique is shown.

3. Results and discussion

3.1. Design of nucleic acid scaffolds and crystallization test of EC and Gfh1

We first tried to cocrystallize EC and Gfh1 by using nucleic acid scaffolds similar to that used for the previously reported *T. thermophilus* EC (Vassylyev, Vassylyeva, Perederina *et al.*, 2007; Vassylyev, Vassylyeva, Zhang *et al.*, 2007). Although those scaffolds yielded EC crystals in the presence of Gfh1, preliminary crystallographic analyses revealed that they belonged to a tetragonal space group with a unit cell similar to that reported for EC and did not contain Gfh1. Therefore, in order to obtain different crystal forms, the lengths and sequences of the downstream DNA and/or the upstream RNA were varied and the resultant scaffolds were used for EC reconstitution and crystallization. As the downstream DNA and the upstream RNA should protrude from RNAP, it was expected that variations of these portions could influence the crystal-packing interactions. Most of the scaffolds, except for the two scaffolds for EC1 and EC2 (Fig. 2), yielded tetragonal or hexagonal RNAP crystals (Kashkina *et al.*, 2006) which did not contain Gfh1.

In contrast, the scaffolds for EC1 and EC2 (Fig. 2) yielded monoclinic crystals which had not been reported previously (Fig. 4). These new crystals contained Gfh1 (Fig. 5). The RNA molecules included in EC1 and EC2 can adopt a 5′-hairpin structure (Fig. 2). They differ by one nucleotide residue in the stem length in the RNA hairpin, as well as in the number of spacer residues connecting the hairpin and the DNA/RNA hybrid. The two nucleic acid scaffolds form stable com-

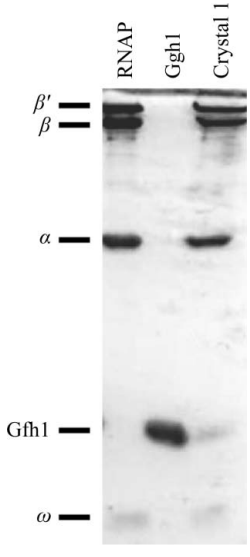


Figure 5
SDS–PAGE analysis of the crystals which contained both RNAP and Gfh1. The crystals of EC1–Gfh1 were washed three times with the reservoir solution and dissolved in water. The solution was analyzed by SDS–PAGE on a 14% gel.

plexes with RNAP *in vitro*, as revealed by gel-retardation experiments (Fig. 3).

3.2. Characterization of EC1 and EC2

To characterize the *in vitro*-assembled EC1 and EC2, we performed an RNA chain-extension experiment (Fig. 6). Each of the nucleic acid scaffolds in the ECs has a thymidine residue as the template nucleotide residue. When ATP was added, EC1 and EC2 elongated the RNA by one nucleotide residue. The RNA extension is sequence-specific, as the addition of CTP, GTP and UTP resulted in no extension of the RNA. These results indicated that EC1 and EC2 retain the characteristics of the normal EC.

To determine the RNA 3′-end position relative to the active site, we examined the susceptibility of ECs to pyrophosphorolysis, which is the reverse reaction to nucleotide addition. ECs in the pre-translocation state are sensitive to pyrophosphorolysis, while ECs in the post-translocation state are resistant (Kashkina *et al.*, 2006). After a 10 min incubation in the presence of 500 µ*M* PP_i, most of the RNAs in EC1 and EC2 remained intact (Fig. 6). This indicates that the DNA/RNA hybrid in these ECs resides predominantly at the position of post-translocation.

3.3. Crystallization

EC1 and EC2 generated high-quality crystals under the same conditions (Fig. 4). The crystals used for data collection were obtained by mixing 1 µl sample solution (20 µ*M* EC1 or EC2 + 60 µ*M* Gfh1) with 1 µl reservoir solution containing 50 m*M* HEPES–NaOH buffer pH 6.4, 5.5% polyethylene glycol (PEG) 8000, 300 m*M* LiCl, 10 m*M* $MgCl_2$, 1% dimethyl sulfoxide and 5 m*M* taurine and equilibrating this mixture against 500 µl reservoir solution at 293 K. Typically, crystals with dimensions of 0.3 × 0.1 × 0.1 mm appeared in a week. The presence of both RNAP and Gfh1 in the crystals was confirmed by SDS–PAGE (Fig. 5).

3.4. Data collection and preliminary crystallographic analyses

EC1–Gfh1 and EC2–Gfh1 yielded crystals that belonged to space group $P2_1$ with similar unit-cell parameters (crystal 1 and crystal 2, respectively). Initial X-ray diffraction experiments were performed at station BL41XU of SPring-8 (Hyogo, Japan). Preliminary data sets were obtained from both types of RNAP crystals and subsequent

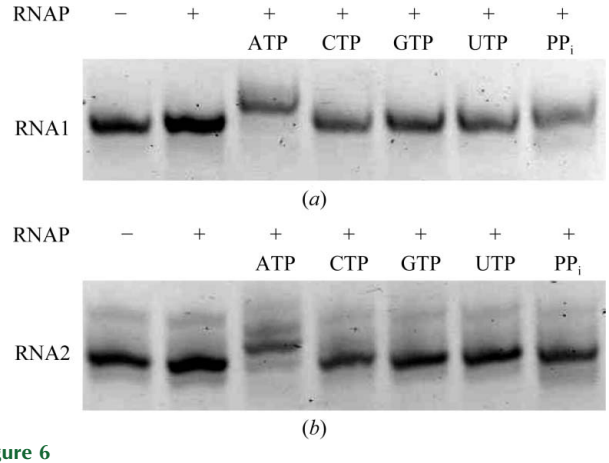


Figure 6
RNA-extension and pyrophosphorolysis analyses for the ECs. (*a*) EC1 and (*b*) EC2 were incubated in the presence of either 10 µ*M* ATP, 10 µ*M* CTP, 10 µ*M* GTP, 10 µ*M* UTP or 500 µ*M* pyrophosphate and the resultant RNAs were analyzed by PAGE on a 20% gel containing 8 *M* urea.

Table 1
Statistics of the diffraction data for the EC–Gfh1 crystals.

Values in parentheses are for the highest resolution shell.

Complex (crystal)	EC1–Gfh1 (crystal 1)	EC2–Gfh1 (crystal 2)
Space group	$P2_1$	$P2_1$
Unit-cell parameters (Å, °)	a = 192.9, b = 260.8, c = 198.6, β = 117.6	a = 189.6, b = 264.9, c = 193.3, β = 116.7
Resolution (Å)	50–3.6 (3.7–3.6)	50–3.8 (3.9–3.8)
Reflections	196251 (14802)	163417 (12057)
Redundancy	3.4 (3.2)	3.3 (3.4)
Mean $I/\sigma(I)$	6.5 (1.9)	7.0 (1.8)
R_{merge}† (%)	23.5 (66.6)	21.3 (71.9)
Completeness (%)	97.6 (93.9)	97.5 (96.5)

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean intensity of multiple $I_i(hkl)$ observations of the symmetry-related reflections.

crystallographic analyses revealed that both crystals exhibited electron density arising from Gfh1 bound to RNAP. Each of the crystals in the crystallization drops differed considerably in its potential to diffract X-rays. For the final data collection, more than 100 crystals that diffracted X-rays beyond 4 Å were selected by taking a snapshot of each crystal and the chosen crystals were stored in liquid nitrogen. The best data sets for the two crystals (crystals 1 and 2) were collected on beamline X06SA at the Swiss Light Source using a PILATUS 6M pixel-array detector with a fine φ-slicing strategy. The data were indexed, integrated and scaled with the *XDS* program (Kabsch, 1993). The data-collection statistics are shown in Table 1. For both crystals, the Matthews coefficient (Matthews, 1968) and the solvent content were calculated to be 3.5 $Å^3\ Da^{-1}$ and 65%, respectively, assuming the presence of three complexes in the asymmetric unit. The high R_{merge} values seem to arise from the relatively weak diffraction intensities, which are probably a consequence of the huge asymmetric unit of ∼1.2 MDa and the presence of flexible domains in RNAP. Structure determination and refinement are in progress.

This work is based on experiments performed at SPring-8 with the approval of the Japan Synchrotron Radiation Research Institute and at the Swiss Light Source, Paul Scherrer Institute, Villigen, Switzerland. We thank Dr Nobukatsu Shimizu for supporting our data collection on SPring-8 beamline BL41XU. We thank Drs Takashi Tomizaki and Clemens Schulze-Briese for supporting our data collection on SLS beamline X06SA. We are grateful to Dr Yoshifumi Fujii for assisting with our data collection and for helpful comments. We thank Tomoyuki Tanaka and Keiko Sakamoto for assistance in protein preparation. This work was supported in part by a Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Young Scientists (to SS), a JSPS Grant-in-Aid for Scientific Research (to SS and SY) and the Targeted Proteins Research Program from the Ministry of Education, Culture, Sports, Science and Technology of Japan. ST was supported by the JSPS Global Centers of Excellence Program (Integrative Life Science Based on the Study of Biosignaling Mechanisms).

References

Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. (2001). *Science*, **292**, 1876–1882.
Hogan, B. P., Hartsch, T. & Erie, D. A. (2002). *J. Biol. Chem.* **277**, 967–975.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Kashkina, E., Anikin, M., Tahirov, T. H., Kochetkov, S. N., Vassylyev, D. G. & Temiakov, D. (2006). *Nucleic Acids Res.* **34**, 4036–4045.
Kettenberger, H., Armache, K. J. & Cramer, P. (2004). *Mol. Cell*, **16**, 955–965.
Laptenko, O., Kim, S. S., Lee, J., Starodubtseva, M., Cava, F., Berenguer, J., Kong, X. P. & Borukhov, S. (2006). *EMBO J.* **25**, 2131–2141.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H. & Artsimovitch, I. (2007). *Nature (London)*, **448**, 157–162.
Vassylyev, D. G., Vassylyeva, M. N., Zhang, J., Palangat, M., Artsimovitch, I. & Landick, R. (2007). *Nature (London)*, **448**, 163–168.
Vassylyeva, M. N., Lee, J., Sekine, S., Laptenko, O., Kuramitsu, S., Shibata, T., Inoue, Y., Borukhov, S., Vassylyev, D. G. & Yokoyama, S. (2002). *Acta Cryst.* D**58**, 1497–1500.
Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D. & Kornberg, R. D. (2006). *Cell*, **127**, 941–954.
Westover, K. D., Bushnell, D. A. & Kornberg, R. D. (2004). *Cell*, **119**, 481–489.