# Structure of *Escherichia coli* exonuclease I in complex with thymidine 5′-monophosphate

**Robert D. Busam‡**

Institute of Molecular Biology, Howard Hughes Medical Institute and Department of Physics, 1229 University of Oregon, Eugene, OR 97403-1229, USA

‡ Current address: Structural Genomics Consortium, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-17177 Stockholm, Sweden.

Correspondence e-mail: robert.busam@ki.se

In *Escherichia coli*, exonuclease I (ExoI) is a monomeric processive 3′–5′ exonuclease that degrades single-stranded DNA. The enzyme has been implicated as primarily being involved in repairing frameshift mutations. The structure of the enzyme has previously been determined in a hexagonal space group at 2.4 Å resolution. Here, the structure of ExoI in complex with a nucleotide product, thymidine 5′-monophosphate, is described in an orthorhombic space group at 1.5 Å resolution. This new high-resolution structure provides some insight into the interactions involved in binding a nucleotide product. The conserved active site contains a unique metal-binding position when compared with orthologous sites in the Klenow fragment, T4 DNA polymerase and dnaQ. This unique difference is proposed to be a consequence of the repositioning of an important histidine, His181, away from the active site.

## 1. Introduction

DNA mismatch repair (MMR) is an important pathway that organisms use to maintain genome fidelity (Modrich, 1989). All organisms have a similar conserved MMR pathway (Schofield & Hsieh, 2003). These repair mechanisms ensure that insertion/deletion loops that result in frameshifts and base–base mismatches are minimized. Methyl-directed mismatch repair in *Escherichia coli* increases the fidelity of DNA replication by approximately 1000-fold (Modrich & Lahue, 1996). Deficiency in MMR results in increased mutation rates and reduces the barrier to cross-species recombination events. Additionally, defects in human MMR predispose individuals to hereditary nonpolyposis colorectal cancer (Lynch & de la Chapelle, 1999). In *E. coli*, the exonucleases involved in MMR are RecJ, exonuclease VII, exonuclease X and exonuclease I (Burdett *et al.*, 2001).

The N-terminal domain of exonuclease I (ExoI) is a member of the dnaQ superfamily. This family consists of related 3′–5′ DNA-processing and RNA-processing enzymes (Moser *et al.*, 1997). These proofreading enzymes share a similar active site defined by exonuclease-sequence (EXO) motifs (Shevelev & Hübscher, 2002). Four of these motifs contain conserved acidic residues that are critical to exonuclease activity. These conserved residues are important for coordinating the divalent cations involved in a phosphoryl-transfer mechanism utilized by a variety of enzymes (Derbyshire *et al.*, 1988, 1991; Hamdan *et al.*, 2002). Typically, these metals are two cations that are ∼3.9 Å apart. In the case of T4 DNA polymerase the two ions are at a greater distance (4.6 Å). This change may occur because the coordinating

**Table 1**
Data-collection and refinement statistics.

Values in parentheses are for the outer shell.

| Data collection | |
| --- | --- |
| Resolution range (Å) | 30–1.5 (1.54–1.50) |
| Space group | $P2_12_12_1$ |
| Unit-cell parameters (Å) | $a = 52.63$, $b = 91.86$, $c = 102.75$ |
| No. of observed reflections | 548887 |
| Completeness (%) | 99.8 (99.5) |
| $R_{merge}$† (%) | 4.0 (37.1) |
| $I/\sigma(I)$ | 47.5 (3.9) |
| Refinement statistics | |
| Resolution range (Å) | 30–1.5 (1.54–1.50) |
| Reflections in working set | 76179 |
| Reflections in test set | 4023 |
| $R$ factor‡ (%) | 20.5 (24.1) |
| $R_{free}$‡ (%) | 22.8 (29.0) |
| Deviations from ideal geometry | |
|   Bond lengths (Å) | 0.019 |
|   Bond angles (°) | 1.8 |
| Distribution of $(\varphi, \psi)$§ | |
|   Most favoured (%) | 99.2 |
|   Additionally allowed (%) | 0.8 |
| No. of protein atoms | 3731 |
| No. of water molecules | 544 |
| No. of TMP molecules | 1 |
| No. of $Mg^{2+}$ atoms | 2 |
| No. of $Na^+$ atoms | 1 |

† $R_{merge} = 100 \times \sum_{hkl} \sum_i |I_i(hkl) - \overline{I(hkl)}|/\sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and $\overline{I(hkl)}$ is the average intensity. ‡ $R$ factor $= 100 \times \sum |F_o| - |F_c|/\sum |F_o|$, where $F_o$ and $F_c$ are the observed and calculated structure factors, respectively. $R_{free}$ is the cross-validation $R$ factor calculated for 5% of the reflections omitted during the refinement process. § Calculated with *MolProbity* (Davis *et al.*, 2007).

tyrosine is oriented away from the active site. In ExoI, the two metals are separated by an even greater distance (5.8 Å). This difference may be a consequence of the coordinating histidine being oriented outside the active site. This structure is unique among dnaQ-family enzymes in the large distance between the metal ions and in the coordinating histidine being oriented away from the active site.

## 2. Materials and methods

### 2.1. Protein expression and purification

Exonuclease I was overexpressed utilizing a pET22b vector (Novagen) with a C-terminal His tag consisting of six histidine residues and no linker. The protein was purified using the following procedure adapted from Breyer & Matthews (2000). The protein was overexpressed in a fermenter (750 rev min$^{-1}$, 7.1 l s$^{-1}$ air) with 4 l LB broth and 0.27 µM ampicillin. The cells were induced at an OD of ~1.2 with 1 mM IPTG for 4 h at 289 K (225 rev min$^{-1}$, 7.5 l min$^{-1}$ air). The cells were harvested and resuspended in 50 mM Tris–HCl pH 8.0, 5 mM imidazole, 500 mM NaCl, 0.1% Triton X-100, 10% glycerol (+TG buffer) and sonicated for 10 min. The cell lysate was centrifuged (17 000 rev min$^{-1}$, Beckman JA-20 rotor) for 30 min and the supernatant was applied onto a column of charged Ni–NTA agarose resin (Qiagen). The column was washed with +TG buffer and −TG buffer (50 mM Tris–HCl pH 8.0, 5 mM imidazole, 500 mM NaCl) until the absorbance

was zeroed. The protein was eluted with a 5–200 mM imidazole gradient. Fractions were collected and analyzed on 20% SDS–PAGE gels (Phastgel). Fractions containing the ExoI protein (54 kDa band) were pooled. Purified ExoI was dialyzed into 50 mM Tris–HCl pH 8.0, 150 mM NaCl, 1 mM EDTA and 1 mM dithiothreitol (DTT). Approximately 200 mg purified protein was obtained from this procedure.

### 2.2. Crystallization

The sequence of the DNA used for cocrystallization was 5′-TTTTTTTTTTT-3′-PHO, where PHO designates a 3′-phosphoryl group at the 3′-terminus. The single-stranded oligonucleotides (cartridge-purified, Invitrogen) were dissolved in water. The protein solution consisted of 20 mg ml$^{-1}$ (0.37 mM) ExoI in buffer (150 mM NaCl, 20 mM Tris–HCl pH 7.9, 1 mM EDTA) and the DNA solution consisted of 0.55 mM ssDNA in water. The protein and DNA solutions were mixed together in equal volumes prior to conducting crystallization setups. The final concentration of protein was 10 mg ml$^{-1}$ (0.19 mM) with 0.28 mM ssDNA. The precipitant solution consisted of 200 mM NaCl, 100 mM Tris–HCl pH 8.5 and 20% PEG 8000 added in an equal volume to the protein–DNA mixture. The reservoir contained 500 µl precipitant solution and the mother liquor consisted of 5 µl protein–DNA mixture and 5 µl well solution. Crystals formed in approximately one week utilizing the hanging-drop vapour-diffusion method at room temperature. The crystals typically measured 0.05 × 0.05 × 0.15 mm. Crystallization experiments using ExoI without the DNA did not form crystals but spherical blobs.

### 2.3. Data collection and structure determination

Diffraction data were collected at the Advanced Light Source, Berkeley, California (beamline 8.2.2, λ = 0.97 Å at 100 K). The crystal was frozen in a solution consisting of precipitant solution and 25% PEG 400 as a cryoprotectant. Data were integrated and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). The structure was solved by molecular replacement with *EPMR* (Kissinger *et al.*, 1999) using the ExoI apo structure (PDB code 1fxx; Breyer & Matthews, 2000) as a search model. A solution with a correlation coefficient of 0.625 was found in space group $P2_12_12_1$ after a single run using default parameters. Atomic coordinates were refined using the *EPMR* solution as the starting model. The model was refined using *TNT* (Tronrud, 1997) and *REFMAC*5 (Collaborative Computational Project, Number 4, 1994). Rigid-body refinement using a 4 Å cutoff was initially used to place the protein in the cell. Rigid-body refinement was then carried out for the three domains. Rigid-body refinement of the helices and $\beta$-sheets was performed using a 3 Å cutoff. Finally, positional and $B$-factor refinement were started with a 2.0 Å cutoff, finishing with a 1.5 Å resolution cutoff.

There are three main regions of disorder. Residues 179–181, which are near the active site, define one region and are discussed in further detail later in this article. The second disordered region includes residues 276–294. Although there is difference density in this region, it was not possible to model

these amino acids. Residues 275–285 form helix *K*, which is found in the SH3-like domain, and residues 286–294 form a flexible linker from the helix to the next *β*-strand. Some of these residues (291–294) are disordered in the apo structure. Helix *K* includes residues that are involved in crystal contacts in the apo structure. Thermal factors at crystal-packing contacts have been shown to decrease in proportion to the area of the surface patch involved in forming contacts (Carugo & Argos, 1997). *CryCo* (Sobolev *et al.*, 1999) determined that the ExoI surface area involved in crystal contacts is 391 Å$^2$ when utilizing a 4.5 Å cutoff for the contact distance. This area is quite small compared with the total surface area (21 775 Å$^2$) exposed to solvent. A majority of the crystal contacts in the ExoI molecule are found on the 'back' of the protein. These contacts possibly play a role in stabilizing helix *K* in the apo structure. There are no crystal contacts in this region of the ExoI–TMP structure, which may explain the increased disorder.

The third site of disorder is the loop spanning the groove. Residues 355–358 were disordered in the apo structure. In the ExoI–TMP structure the disorder extends over an additional nine residues (348–360).

The first six residues of the protein are disordered, which is similar to the apo structure, which is missing residues 1–7. The C-terminal His tag is also not visible in the ExoI–thymidine 5′-monophosphate (TMP) structure, as in apo ExoI.

A short region consisting of residues 157–160 is disordered in the apo structure. In the ExoI–TMP structure these amino acids are ordered and form a turn in the loop connecting helices *E* and *F*. These residues are also involved in crystal contacts, which may account for their visibility in this structure.

A TMP molecule was unambiguously identified in $2F_o − F_c$ and $F_o − F_c$ OMIT electron-density maps (Fig. 2). The programs *XFIT* (McRee, 1999) and *Coot* (Emsley & Cowtan, 2004) were employed for manual building of the model. A summary of the data-processing and refinement statistics is given in Table 1.

## 3. Results and discussion

### 3.1. Structure determination

Crystals of a complex between ExoI and thymidine 5′-monophosphate (ExoI–TMP) were obtained in the orthorhombic space group $P2_12_12_1$, with unit-cell parameters $a = 52.63$, $b = 91.86$, $c = 102.75$ Å (see §2). They are non-isomorphous with the apoprotein (Breyer & Matthews, 2000) and diffract to higher resolution (1.5 Å compared with 2.4 Å). The structure, which was determined by molecular replacement, has overall excellent stereochemistry as judged from the Ramachandran plot, which shows that 100% of the residues are in the allowed region. The final refinement of ExoI–TMP yielded a protein model with an *R* value of 20.5% ($R_{free}$ = 22.8%). Data-collection and refinement statistics are summarized in Table 1.

### 3.2. Overview of the structure

ExoI–TMP assumes a roughly spherical shape with a large groove through the centre of the molecule (Fig. 1). The r.m.s. difference of 403 C$^α$ positions relative to the apoprotein is 1.03 Å using a nonlinear least-squares fit in *Coot* (Emsley & Cowtan, 2004). The first 201 residues consist of the exonuclease domain, which shares homology to the proofreading domains of DNA polymerases. The SH3-like domain, residues 202–354, and the C-terminal *α*-helical domain, residues 359–475, maintain a similar fold to the apo structure.
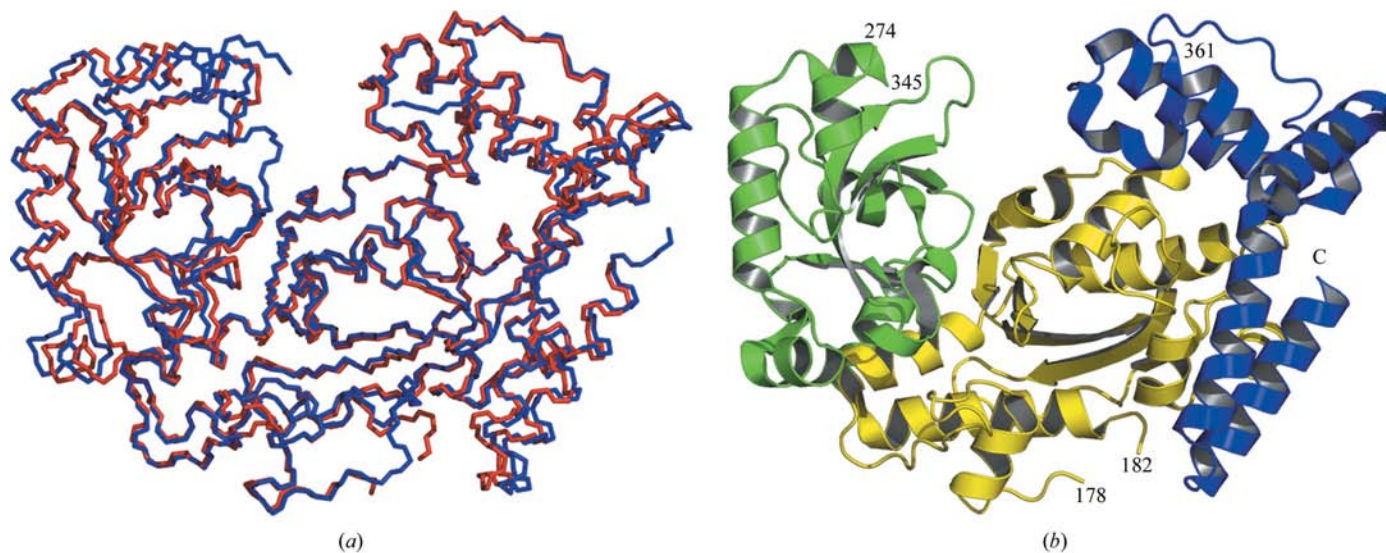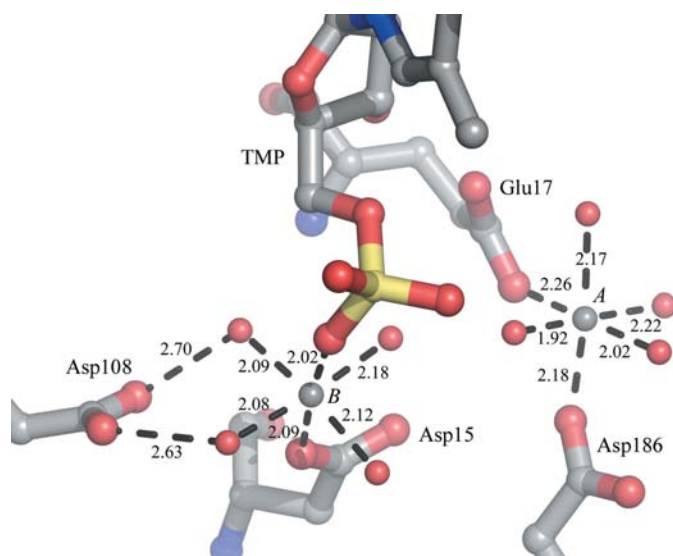


**Figure 1**
ExoI–TMP structure. (*a*) C$^α$ structural alignment of ExoI (blue) with ExoI–TMP (red) utilizing *SSM* (Krissinel & Henrick, 2004). The TMP-bound structure is highly similar to the apo form. (*b*) High-resolution structure showing the N-terminal exonuclease domain (yellow), the SH3-like domain (green) and the C-terminal domain (blue). Disordered regions include residues 1–6, 179–181, 276–294, 348–360 and the C-terminal His tag. The overall fold is highly similar to the apo structure. Numbers indicate primary amino-acid sequence locations near disordered regions.
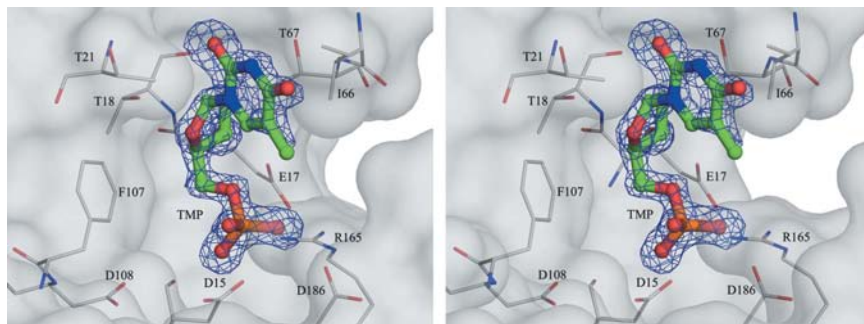
**Figure 2**
Stereoview of an OMIT map of the TMP molecule bound in the active site. Electron density is contoured at $3\sigma$. The resolution was 1.5 Å. Phases from the refined model with the TMP coordinates deleted are shown. Asp15, Glu17, Asp108 and Asp186 are active-site residues. Thr21, Ile66, Thr67 and Phe107 are involved in hydrophobic contacts with TMP. Arg165 shows hydrogen bonding of the guanidinium group to the 5′-phosphate of TMP.



**Figure 3**
Hydrogen-bonding and metal-binding network in the active site. The red spheres are water molecules and the grey spheres (*A* and *B*) are magnesium ions. The acidic residues Asp15, Glu17, Asp108 and Asp186 are involved in coordinating these metals. The distances are in Å.

### 3.3. Disordered region near active site

In the ExoI–TMP structure there are three regions that are disordered in addition to the amino acids at the N- and C-termini. Some of these regions are described in §2. An important region of disorder is residues 179–181. In the apo structure, these residues form a flexible loop that joins helices *F* and *G*. They are also mobile, with an average *B* factor of ~75 Å². In ExoI His181 is oriented toward the active site and is proposed to play a role in coordinating the water for attack on the phosphate group (Hamdan *et al.*, 2002). In the ExoI–TMP structure this residue is disordered. *ARP/wARP* v.6.1 (Morris *et al.*, 2003) was utilized to determine whether the loop could be modelled based solely on experimental phases or using an existing model. In both cases, residues 179–181 could not be modelled in a satisfactory manner. An attempt to model the loop was also made using *XPLEO* (van den Bedem

*et al.*, 2005). In this case, the solution resulted in a loop in which His181 was oriented away from the active site. In both $2F_o - F_c$ and $F_o - F_c$ maps there is almost no significant electron density for these residues. Therefore, the location of His181 in this structure remains uncertain. However, this residue is not likely to be oriented towards the active site based on the positions of visible downstream residues.

### 3.4. TMP product bound in the active site

The ssDNA used for crystallization (see §2) has a 3′-phosphoryl group added. It has previously been shown that such substrates are inactive in ExoI and the Klenow fragment (KF; Lehman & Nussbaum, 1964; Livingston & Richardson, 1975). However, the full-length ssDNA substrate [p(dT)$_{10}$] is not observed in the crystal. Only a single TMP product appears to be bound in the active site (Fig. 2). This suggests that the enzyme is capable of hydrolyzing substrates with 3′-phosphoryl groups during the period of days required to grow the crystals. *In vitro*, the enzyme is active at the crystallization pH of 8.5.

The TMP molecule is oriented within the active site similarly to DNA substrates complexed with the KF (Ollis *et al.*, 1985; Derbyshire *et al.*, 1988; Brautigam & Steitz, 1998; Brautigam *et al.*, 1999), T4 DNA polymerase (Wang *et al.*, 1996) and dnaQ (Hamdan *et al.*, 2002). In the ExoI–TMP structure the TMP molecule is bound in the active site through hydrogen bonds and hydrophobic interactions (Fig. 2). The residues involved in hydrophobic contacts are Thr21, Ile66, Thr67 and Phe107. The $C^\gamma$ of Thr21 interacts with the thymine ring on one side and with Ile66 on the other, forming a hydrophobic sandwich. Thr67 and Phe107 are involved in hydrophobic contacts with the sugar moiety. The backbone N atom of Thr18 forms a hydrogen bond to the 3′ O of the sugar. Interestingly, Arg165 shows hydrogen bonding of the guanidinium group to the 5′-phosphate of TMP by rotating ~90° in relation to the apo structure. This residue may play a role in orienting the substrate within the active site.

### 3.5. Active-site bonding network

In addition to the hydrophobic and hydrogen-bond interactions with the main-chain atoms, there is an extensive hydrogen-bonding network coordinating active-site residues with metal ions and phosphate O atoms (Fig. 3). All the conserved acidic residues are involved in the hydrogen-bond network. Of particular interest is the coordination of two Mg$^{2+}$ ions, *A* and *B*. Metal *B* is coordinated to a phosphate O atom through a monodentate linkage. The carboxyl group of Asp15 coordinates a water and metal *B*. Metal *B* is also liganded by four waters found in the same plane. These waters are also involved in coordination to additional water and side-chain carboxyl O atoms in the active-site region. Metal *A* is coordinated by four water molecules and the carboxyl O atoms

# research papers

of Glu17 and Asp186. While metal *B* is in a highly similar position as observed in KF and dnaQ structures, metal *A* is in a different position. The distances between metals *A* and *B* in the KF and dnaQ are 3.85 and 3.74 Å, respectively. In the ExoI–TMP structure these metals are 5.66 Å apart. The reason for this increased distance between the metals is not entirely clear. In dnaQ, His162, which corresponds to His181 in ExoI, occupies a position that would result in a steric clash with metal *A*. Differences in the way the TMP molecule was bound were observed in the $\varepsilon$ subunit of polymerase III at pH 5.8 and pH 8.5 (Hamdan *et al.*, 2002). These changes involve the coordination of His162 and not the binuclear metal centres. Therefore, the differences in metal coordination observed in the ExoI–TMP structure are likely to be a consequence of structural rearrangement rather than pH. Movement of His181 from the active site may allow the shift in metal position. In addition to the $Mg^{2+}$ ions, an $Na^+$ ion is also involved in the bonding network. Glu17, Asp186 and four waters coordinate this ion. To assess whether the metal assignment is correct, the coordination and bond distances of these ions were measured by the calcium bond-valence sum (CBVS) method (Müller *et al.*, 2003). The results of these calculations on the three metal ions are consistent with CBVS values, suggesting that the metal assignment is correct (data not shown).

In summary, the ExoI–TMP structure is very similar to the previously solved ExoI structure. A product of the exonuclease reaction, TMP, is bound in the active site. The TMP is bound through hydrophobic and hydrogen-bond interactions. Arg165 assists in the coordination of the phosphate *via* its guanidinium group. His181, which is presumed to play a role in coordination and deprotonation of a metal-coordinated water, is not visible in the structure. Based on the position of downstream residues, it appears to have flipped away from the active site. This movement may allow metal *A* to shift position further than observed in previous dnaQ-family structures. Whether these observed changes in metal position are relevant to the mechanism of hydrolysis remains to be seen.

## References

Bedem, H. van den, Lotan, I., Latombe, J.-C. & Deacon, A. M. (2005). *Acta Cryst.* D**61**, 2–13.
Brautigam, C. A. & Steitz, T. A. (1998). *J. Mol. Biol.* **277**, 363–377.
Brautigam, C. A., Sun, S., Piccirilli, J. A. & Steitz, T. A. (1999). *Biochemistry*, **38**, 696–704.
Breyer, W. A. & Matthews, B. W. (2000). *Nature Struct. Biol.* **7**, 1125–1128.
Burdett, V., Baitinger, C., Viswanathan, M., Lovett, S. T. & Modrich, P. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 6765–6770.
Carugo, O. & Argos, P. (1997). *Protein Sci.* **6**, 2261–2263.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B. III, Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). *Nucleic Acids Res.* **35**, W375–W383.
Derbyshire, V., Freemont, P. S., Sanderson, M. R., Beese, L., Friedman, J. M., Joyce, C. M. & Steitz, T. A. (1988). *Science*, **240**, 199–201.
Derbyshire, V., Grindley, N. D. & Joyce, C. M. (1991). *EMBO J.* **10**, 17–24.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Hamdan, S., Carr, P. D., Brown, S. E., Ollis, D. L. & Dixon, N. E. (2002). *Structure*, **10**, 535–546.
Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Lehman, I. R. & Nussbaum, A. L. (1964). *J. Biol. Chem.* **239**, 2628–2636.
Livingston, D. M. & Richardson, C. C. (1975). *J. Biol. Chem.* **250**, 470–478.
Lynch, H. T. & de la Chapelle, A. (1999). *J. Med. Genet.* **36**, 801–818.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Modrich, P. (1989). *J. Biol. Chem.* **264**, 6597–6600.
Modrich, P. & Lahue, R. (1996). *Annu. Rev. Biochem.* **65**, 101–133.
Morris, R. J., Perrakis, A. & Lamzin, V. S. (2003). *Methods Enzymol.* **374**, 229–244.
Moser, M. J., Holley, W. R., Chatterjee, A. & Mian, I. S. (1997). *Nucleic Acids Res.* **25**, 5110–5118.
Müller, P., Köpke, S. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 32–37.
Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. (1985). *Nature (London)*, **313**, 762–766.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Schofield, M. J. & Hsieh, P. (2003). *Annu. Rev. Microbiol.* **57**, 579–608.
Shevelev, I. V. & Hübscher, U. (2002). *Nature Rev. Mol. Cell Biol.* **3**, 364–376.
Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E. & Edelman, M. (1999). *Bioinformatics*, **15**, 327–332.
Tronrud, D. E. (1997). *Methods Enzymol.* **277**, 306–319.
Wang, J., Yu, P., Lin, T. C., Konigsberg, W. H. & Steitz, T. A. (1996). *Biochemistry*, **35**, 8110–8119.