Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization of the catalytic domain of murine terminal deoxynucleotidyl transferase

N. Sukumar,[a]† J. B. Boulé,[b]† N. Expert-Bezançon,[a] N. Jourdan,[a] J. Lescar,[c] F. Rougeon,[b] C. Papanicolaou[b] and M. Delarue[a]*

[a]Unité de Biochimie Structurale, Institut Pasteur, 25 Rue du Dr Roux, 75015 Paris, France, [b]Unité de Génétique et Biochimie du Développement, Institut Pasteur, 25 Rue du Dr Roux, 75015 Paris, France, and [c]Joint Structural Biology Group, ESRF, Grenoble, France

† The first two authors contributed equally to this work.

Correspondence e-mail: delarue@pasteur.fr

Received 19 July 2000
Accepted 21 September 2000

The catalytic domain of murine terminal deoxynucleotidyl transferase (TdT) has been crystallized in the space group $P2_12_12_1$, with unit-cell parameters a = 47.1, b = 86.2, c = 111.7 Å. The crystals diffract to a resolution of 2.4 Å using synchrotron radiation and a full data set has been collected from the native crystals. The enzyme was shown to be active in the crystalline state.

1. Introduction

Terminal deoxynucleotidyl transferase (TdT; E.C. 2.7.7.31) is a template-independent DNA polymerase which carries out the 3′-OH extension of a DNA strand one nucleotide at a time (Kato *et al.*, 1967; Bollum, 1974, 1978). It cannot synthesize a strand *de novo* and the oligonucleotide primer must be at least three nucleotides long with a 5′-phosphate (Kato *et al.*, 1967). TdT can incorporate both ribonucleotides and deoxyribonucleotides, as well as several unnatural nucleoside triphosphates. *In vivo*, it is known to contribute to the generation of diversity of the immune repertoire by adding random nucleotides, called N regions, at the V(D)J recombination junction sites of immunoglobulins and T-cell receptors (Kallenbach *et al.*, 1990; Komori *et al.*, 1993; Gilfillan *et al.*, 1993).

TdT belongs to an ancient nucleotidyltransferase (NT) superfamily, as shown by sequence database searches as well as by structure-inspired delineation of functionally important sequence motifs (Holm & Sander, 1995). This family includes DNA polymerase β (pol β), poly(A) polymerase, CCA:tRNA nucleotidyltransferase, 2′-5′ oligoadenylate synthetase and several other prokaryotic and eukaryotic nucleotidyltransferases (for a recent review, see Aravind & Koonin, 1999). Recent data indicate that *Escherichia coli* DNA polymerase III could also belong to the NT family (Pritchard & McHenry, 1999).

Only one member of this family, DNA polymerase β, is known in detail, at atomic resolution, both as an isolated protein and in various binary and ternary complexes with natural substrates (Davis *et al.*, 1994; Pelletier *et al.*, 1994*a*,*b*).

The topology of the catalytic core of pol β is very different from that of polymerases from the DNA pol I, reverse transcriptases and DNA polymerase α families. These three families share a common fold represented by the topology of the Klenow fragment (Ollis *et al.*, 1985; Delarue *et al.*, 1990; Wang *et al.*, 1997; Doublié *et al.*, 1998). *Thermus aquaticus* DNA polymerase I, a member of the pol I family, has been shown to adopt a closed and an open conformation during catalysis (Li *et al.*, 1998). Pol β also exists in at least two conformations, namely the closed and open complexes (Sawaya *et al.*, 1997). Since TdT does not use a template DNA strand, it would be interesting to know whether or not this transition between an open and closed complex also occurs.

Significant sequence similarity with pol β can be detected using *BLASTP* ($E = 1 \times 10^{-23}$ between rat pol β and murine TdT sequences). This similarity encompasses both the catalytic domain and the so-called finger and thumb regions, which are involved in the binding of both primer and template strands. Therefore, in principle TdT has all the elements to be a replicase. The structure should explain why TdT cannot accommodate a template strand and suggest possible mutations to transform it into a replicative polymerase.

The active site of all structurally known polymerases always contains crucial aspartate side chains in the vicinity of a catalytically important Mg^{2+} ion and of the incoming nucleoside triphosphate. In TdT, Co^{2+} and Mg^{2+} are the most efficient divalent cations for the incorporation of pyrimidine and purine nucleotides, respectively (Bollum, 1978). The three-dimensional structure of TdT should explain why this is the case, as well as provide a limit to the tolerance of the enzyme towards incorporation of modified nucleotides.

In conclusion, structural comparison of polymerase members of the NT family should give insights into both the mechanisms of nucleic acid synthesis which are shared by all polymerases and those which underlie substrate specificities.

2. Materials and methods

2.1. Protein purification

The protein was overexpressed and purified in *E. coli* as described in Boulé *et al.* (1998).

The purity was estimated to be more than 98% based on SDS gel electrophoresis stained by Coomassie blue.

2.2. Crystallization

The protein was kept frozen at 253 K in 20 m*M* Tris–HCl pH 6.8, 200 m*M* NaCl, 50 m*M* $Mg(OAc)_2$ and 100 m*M* $(NH_4)_2SO_4$ at a concentration of 10 mg ml^{-1}. Aliquots of the protein stock solution were defrosted just before setting up the drops. Crystallization trials were performed using the hanging-drop method.

2.3. Crystal handling and flash-cooling

In a first step, crystals were isolated from their original drop using cryoloops from Hampton Research Corp. (Laguna Niguel, CA, USA) and were then transferred to drops containing only the reservoir solution, where their optical quality and their thickness were evaluated. In particular, surface irregularities on the largest face were indicative of generation of non-unique crystal blocks during crystal growth and generally resulted in too high mosaicity in the diffraction pattern; therefore, these crystals were discarded.

Crystals were too small to be tested in the laboratory and were frozen in liquid nitrogen just before synchrotron trips in a solution containing the reservoir and 25% glycerol as a cryoprotectant.

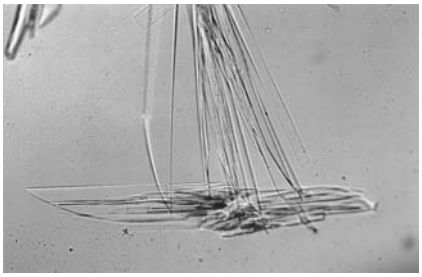

Figure 1
Photograph of TdT crystals: the knife-shaped crystal can be seen in the upper part of the figure. Typical dimensions are 500 µm in the long dimension and 80–100 µm in the other visible dimension.

Figure 2
Analysis of TdT activity in the crystals. Crystals were soaked in 400 n*M* [γ-^{32}P]-$(dA)_{10}$ and 20 m*M* dATP; the polymerization products were analyzed by PAGE.

2.4. TdT activity in the crystalline state

Carefully washed crystals were soaked for 12 h at 298 K in a solution containing 20% PEG 6000, 100 m*M* MES pH 6.0, 1 *M* LiCl, 16 m*M* Mg^{2+}, 16 µ*M* Zn^{2+}, 4 m*M* dATP and 400 n*M* [γ-^{32}P]-$(dA)_{10}$. It was verified under the microscope that this treatment left the crystals intact. 200 m*M* EDTA was added to prevent any further reaction occurring before the crystals were dissolved in a formamide dye solution (10 m*M* NaOH, 95% formamide, 0.05% bromophenol blue, 0.05% xylene cyanol). The mixture was then analysed using polyacrylamide gel electrophoresis.

2.5. Data collection

All data collection was performed at the ESRF in Grenoble at beamlines ID2, ID14-EH3 or ID14-EH2.

Diffraction data were collected from a crystal mounted in a cryoloop placed in a stream of cold nitrogen at 100–110 K. Exposure times were typically 3 × 5 s at beamline ID14-EH3 equipped with a MAR CCD detector. The wavelength was 0.9402 Å. The crystal-to-detector distance was set to 150 mm and the oscillation angle for each frame was 1°, collected in three passes. The speed of data collection using a CCD detector was instrumental in collecting a full data set at the highest possible resolution. A low-resolution pass at 3.5 Å was performed after the high-resolution pass, with exposure time 3 × 1 s. The crystals were stable in the beam for about 90–120 min. Data processing was performed with the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

3. Results and discussion

TdT contains two functionally independent regions (Chang *et al.*, 1982). The N-terminal region (130 amino acids) contains a BRCT motif involved in protein–protein interactions (Koonin *et al.*, 1996; Caillebaut & Mornon, 1997; Bork *et al.*, 1997). The BRCT domain of XRCC1 is believed to be functionally important in bringing different partners into physical proximity through heterodimerization (Rice, 1999). The 3.2 Å X-ray structure of this domain has recently been solved (Zhang *et al.*, 1998). The C-terminal region of TdT contains the catalytic core of the protein.

Of the TdT protein variants that were overexpressed in *E. coli* and purified, only the so-called Δ129 construct, in which the first 129 amino acids were removed, gave crystals diffracting to high resolution. It contains 399 amino acids (six histidines were added at the N-terminus to facilitate purification, as well as 14 other amino acids arising from the cloning itself) and is catalytically fully active (Boulé & Papanicolaou, data not shown).

Crystals of the Δ129 construct were obtained after less than 24 h of equilibration of the drop against a reservoir containing 20–23% PEG 6000, 100 m*M* MES pH 6.0, 1 *M* LiCl. The drop was set up with one volume of the protein stock solution and one volume of the reservoir, although in some cases better results were obtained with an unequal ratio (*e.g.* 3:2, 3.5:2.5) of these two components. This parameter was systematically screened each time crystallization experiments were set up.

Crystals tended to appear as clusters or showers of long needles which were difficult to handle, 300–500 µm long, sometimes with a width of as much as 80–100 µm, but of unmeasurable thickness (less than 5 µm). In some cases, crystals with a knife-like appearance were observed and it was possible to see that these crystals had a measurable thickness (10 µm) at the cutting edge of the knife (Fig. 1). These were not only the crystals with the best morphology but also the best crystals in terms of diffraction. However, appearance of this type of knife-shaped crystals was rare and a number of crystallization drops (typically 12–24) had to be set up each time in order to see only one or two of these crystals.

The mass spectrometry of dissolved crystals revealed a major peak at $m/z = 45505 \pm 4.5$, whereas the expected m/z is 45504.9 without the first (methionine) residue. N-terminal chemical sequencing was performed for the first 22 amino acids of the purified protein. They were found to be in accordance with the expected sequence, starting just after the N-terminus methionine. In addition to this major peak in the mass spectrometry experiment, an additional peak was observed at $m/z = 45811 \pm 1$, which was consistently found in four separate experiments of different protein

preparations; this was interpreted as an adduct between a cysteine residue of the protein and a glutathione molecule.

To assess the presence of TdT activity in the crystals, crystals were first washed several times in the reservoir solution and were then incubated overnight in a solution containing a radioactive $(dA)_{10}$ oligonucleotide primer and dATP. PAGE analysis of the reaction products clearly showed primers elongated by several nucleotides. This result suggests that the enzyme is active in the crystal form (Fig. 2).

Using synchrotron radiation, a full data set was collected from a single native crystal at beamline ID2 at 2.8 Å resolution; subsequently, a full data set at 2.4 Å was colelcted at beamline ID14-EH3. A total of 110 frames were collected from a single crystal. The unit-cell parameters are $a = 47.1, b = 86.2, c = 111.7$ Å and the space group is $P2_12_12_1$. An overall total of 99 454 observations leading to 19 351 independent reflections were recorded (99.8% complete for all data and 98.8% complete for the last resolution shell), with an average R_{sym} of 5.9% (23% in the last resolution shell, 2.46–2.40 Å resolution). The percentage of data with $I > 3\sigma(I)$ is 84 and 61.2% for all data and for the last resolution shell, respectively. The mean redundancy is 5.2 for all reflections and 4.1 for the last resolution shell.

Molecular replacement using DNA polymerase β has so far been unsuccessful. This may be because of the relatively low sequence identity between the two sequences (27.5% for 332 aligned positions). Heavy-atom derivatives are currently being screened in order to solve the phase problem experimentally.

We thank A. Namane for performing the mass-spectrometry measurements and J. D'Alayer for chemical N-terminal sequencing. It is a pleasure to thank S. Arzt and W. Burmeister for help during data collection at beamline ID14-EH3, as well as E. Mitchell at beamline ID14-EH2, at the ESRF in Grenoble, France. This work was partly supported by a grant from the Association de la Recherche sur le Cancer (ARC). NJ was supported in part by the Fondation pour la Recherche Medicale (FRM) and by a fellowship from the Institut Pasteur (Bourse Cantarini). NS was also supported by the latter fellowship.

References

Aravind, L. & Koonin, E. V. (1999). *Nucleic Acids Res.* **27**, 1609–1618.
Bollum, F. J. (1974). *The Enzymes*, edited by P. D. Boyer, Vol. 10, pp. 145–171. New York: Academic Press.
Bollum, F. J. (1978). *Adv. Enzymol.* **47**, 347–372.
Bork, P., Hofmann, K., Bucher, P., Neuwald, A. F., Altschul, S. F. & Koonin, E. V. (1997). *FASEB J.* **11**, 68–76.
Boulé, J. B., Johnson, E., Rougeon, F. & Papanicolaou, C. (1998). *Mol. Biotechnol.* **10**, 199–208.
Caillebaut, I. & Mornon, J. P. (1997). *FEBS Lett.* **400**, 5–30.
Chang, L. M. S., Plevani, P. & Bollum, F. (1982). *J. Biol. Chem.* **257**, 5700–5706.
Davis, J. F., Almassy, R. J., Hostomska, Z., Ferre, R. & Hostomsky, Z. (1994). *Cell*, **76**, 1123–1133.
Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. (1990). *Protein Eng.* **3**, 461–467.
Doublié, S., Tabor, S., Long, A. M., Richardson, C. R. & Ellenberger, T. (1998). *Nature (London)*, **391**, 251–258.
Gilfillan, S., Dierich, A. & Lemeur, M. (1993). *Science*, **261**, 1175–1178.
Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 345–347.
Kallenbach, S., Goodhardt, M. & Rougeon, F. (1990). *Nucleic Acids Res.* **18**, 6730–6737.
Kato, K. I., Gonçalves, J. M., Houts, G. E. & Bollum, F. J. (1967). *J. Biol. Chem.* **242**, 2780–2789.
Komori, T., Okada, A., Stewart, V. & Alt, F. W. (1993). *Science*, **261**, 1171–1175.
Koonin, E. V., Altschul, S. F. & Bork, P. (1996). *Nature Genet.* **13**, 266–267.
Li, Y., Korolev, S. & Waksman, G. (1998). *EMBO J.* **17**, 7514–7525.
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.
Pelletier, H., Sawaya, M. R., Wolfe, W. & Kraut, S. H. (1994*a*). *Science*, **264**, 1891–1903.
Pelletier, H., Sawaya, M. R., Wolfe, W. & Kraut, S. H. (1994*b*). *Science*, **264**, 1930–1935.
Pritchard, A. E. & McHenry, C. S. (1999). *J. Mol. Biol.* **285**, 1067–1080.
Rice, P. A. (1999). *Nature Struct. Biol.* **6**, 805–807.
Sawaya, M. R., Prasad, R., Wilson, S. H., Kraut, J. & Pelletier, H. (1997). *Biochemistry*, **36**, 11205–11215.
Wang, J., Sattar, A. K. M. A., Wang, C. C., Karam, J. D., Konigsberg, W. H. & Steitz, T. A. (1997). *Cell*, **89**, 1087–1099.
Zhang, X., Morera, S., Bates, P. A., Whitehead, P. C., Coffer, A. I., Hainbucher, K., Nash, R. A., Sternberg, M. J., Lindahl, T. & Freemont, P. S. (1998). *EMBO J.* **17**, 6404–6416.