Acta Cryst. (1996). D**52**, 226–227

Crystallization and preliminary X-ray analysis of bacteriophage T4 deoxynucleotide kinase

PAULA SEBASTIAO,[a] GALYA OBMOLOVA,[a] GEORGE S. BRUSH,[b] MAURICE J. BESSMAN[b] AND ALEXEI TEPLYAKOV[a]* *at* [a]*European Molecular Biology Laboratory, c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany, and* [b]*Department of Biology and the McCollum-Pratt Institute, The Johns Hopkins University, Baltimore, MD 21218, USA*

(Received 21 March 1995; accepted 1 June 1995)

Abstract

T4 deoxynucleotide kinase catalyzes the phosphorylation of 5-hydroxymethyldeoxycytidylate, dTMP and dGMP while excluding dCMP and dAMP. In order to understand the mechanism of this remarkable specificity, the enzyme was over-expressed in *Escherichia coli*, purified and crystallized for X-ray diffraction analysis. The crystals belong to the monoclinic space group $C2$ with cell dimensions $a = 155.2$, $b = 58.5$, $c = 75.7$ Å, $\beta = 108.1°$. There are two protein monomers in the asymmetric unit related by a twofold axis. Diffraction data to 2.0 Å resolution have been collected.

1. Introduction

In most species, four distinct and highly specific enzymes catalyze the phosphorylation of the canonical deoxynucleotides to the corresponding diphosphates in the presence of ATP and Mg^{2+}. However, the selective pressure for efficient DNA precursor synthesis in phage-infected cells has resulted in the unusually broad specificity of bacteriophage T4 deoxynucleotide kinase (T4DNK) which recognizes three structurally dissimilar nucleotides: hmdCMP (5-hydroxymethyldeoxycytidylate), dTMP and dGMP while excluding dCMP and dAMP (Bello & Bessman, 1963; Duckworth & Bessman, 1967). Only deoxynucleotides are phosphorylated, indicating that the enzyme recognizes the sugar of the substrate as well as the base. Kinetic experiments and the chemical modification studies (Brush & Bessman, 1993) have shown that one active site is responsible for all three activities. The mechanism that allows for this remarkable specificity has been the focus of current research on T4DNK. An understanding of this discrimination may provide an insight into principles of specific protein–nucleotide and protein–nucleic acid interactions.

2. Results

T4DNK was over-expressed in *Escherichia coli* and purified to homogeneity as described (Brush, Bhatnagar & Bessman, 1990). In the absence of dithiothreitol (DTT), T4DNK forms high molecular weight aggregates. In the presence of 10 m*M* DTT, 90% of the protein exists as homodimers which is the native form of the enzyme. Other oligomeric forms of T4DNK were removed using fast protein liquid chromatography with a Mono Q anion-exchange column.

T4DNK has been crystallized by vapour diffusion in hanging drops at 277 K. The initial screening of crystallization conditions was performed with the Hampton Research crystallization kit. Small well shaped crystals were obtained under conditions No. 18: 20% PEG 8000 in sodium cacodylate buffer, pH 6.5, with 0.2 *M* magnesium acetate. As magnesium was known to play a functional role in kinases, various concentrations of magnesium acetate were tried in order to improve the quality of the crystals. It appeared that the crystals grew only in the range of 0.2–0.4 *M* of magnesium acetate. Most of these crystals were twinned. Out of many additives tried, dioxane gave the best improvement in diffraction ability of the crystals and to avoid twinning. Finally, 3 µl of the 10 mg ml^{-1} protein solution in 20 m*M* Tris–HCl buffer (pH 7.6) with 7 m*M* DTT and 5 m*M* dGMP were mixed with 3 ml of the reservoir solution which contained 11–14% PEG 8000, 0.2 *M* magnesium acetate, 7 m*M* DTT, 2% dioxane and 0.1 *M* cacodylate buffer (pH 6.5). Crystals appeared after a few days and reached a size of 0.5 × 0.3 × 0.2 mm. They belong to the monoclinic space group $C2$ with cell dimensions $a = 155.2$, $b = 58.5$, $c = 75.7$ Å, $\beta = 108.1°$. Assuming a 54 kDa T4DNK dimer in the asymmetric unit gives a solvent content of 59% and a V_M of 3.0 Å^3 Da^{-1} (Matthews, 1968).

Diffraction data to 2.0 Å resolution were collected from one of these crystals at 277 K using synchrotron beamline BW7B (EMBL, DESY, Hamburg) equipped with a MAR Research imaging-plate scanner. The crystal was mounted with the a^*

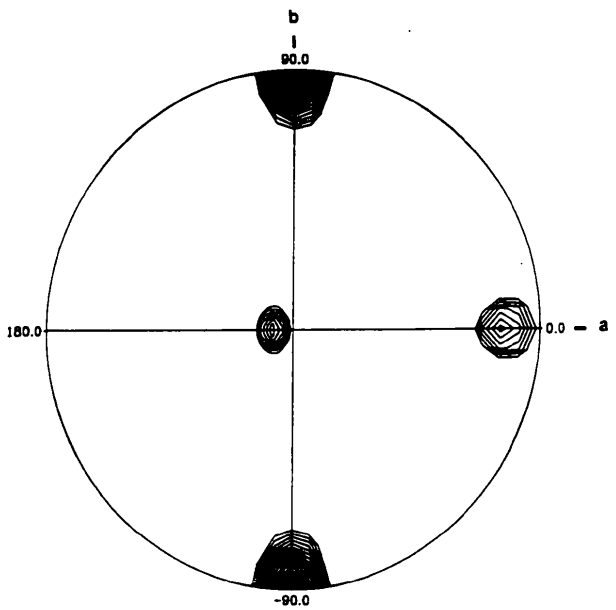

Fig. 1. Stereographic projection of the section $\kappa = 180°$ of the self-rotation function of T4DNK. Integration radius in Patterson space is 17 Å and data are included from 15 to 3.0 Å. Two peaks at $\omega = 90°$ and $\varphi = 90°$ (with the relative height of 100) correspond to the crystallographic twofold b axis and two peaks at $\omega = 81$, $\varphi = 0$ and $\omega = 9$, $\varphi = 180°$ (with the relative height of 64) correspond to the non-crystallographic twofold axes lying in the ac plane.

© 1996 International Union of Crystallography
Printed in Great Britain – all rights reserved

axis approximately parallel to the spindle axis and c^* initially perpendicular to the direction of the X-ray beam. In this orientation, the complete data set could be obtained after 90° rotation using 1.2° oscillations. Two data sets, at 2.0 and 3.8 Å resolution, were collected at different exposure times (4 and 1 min per image, respectively) to record the whole range of intensities. The data were processed using the programs *DENZO* and *SCALEPACK* (Otwinowski, 1993). A total of 11 8991 independent measurements were merged to obtain 41 280 unique reflections (R_{merge} = 5.9%) which constitute 97% of theoretically possible data.

The self-rotation function calculated on these data using the *CCP*4 package (Collaborative Computational Project, Number 4, 1994) revealed two strong peaks representing non-crystallographic twofold axes lying in the *ac* plane (Fig. 1). This confirms the presence of the protein dimer in the asymmetric unit.

Sequence alignment studies using the algorithm of Higgins, Bleasby & Fuchs (1992) showed a very low percentage of identity (12–18%) to adenylate and guanylate kinases with known three-dimensional structures (Diederichs & Schulz, 1990; Dreusicke, Karplus & Schulz, 1988; Müller & Schulz, 1992; Stehle & Schulz, 1990). However, the mononucleotide-binding consensus sequence Gly-*X*-*X*-Gly-*X*-Gly-Lys (where *X* represents any amino acid) observed in the kinases is also present in T4DNK although the second Gly is replaced by Arg. This 'Gly-loop' forms a giant anion hole which accommodates phosphates of a mononucleotide (Dreusicke & Schulz, 1986). It is located between a β-strand of the central parallel pleated sheet and an α-helix. The existence of the 'Gly-loop' in T4DNK implies the similar topology of the mononucleotide-binding motif in T4DNK and adenylate and guanylate kinases.

The work was supported by the Commission of the European Communities fellowship No. ERB4001GT941738 to PS, and by a grant (GM-18649) to MJB from the National Institutes of Health.

References

Bello, L. J. & Bessman, M. J. (1963). *J. Biol. Chem.* **238**, 1777–1787.
Brush, G. S. & Bessman, M. J. (1993). *J. Biol. Chem.* **268**, 1603–1609.
Brush, G. S., Bhatnagar, S. K. & Bessman, M. J. (1990). *J. Bacteriol.* **172**, 2935–2939.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Diederichs, K. & Schulz, G. E. (1990). *Biochemistry*, **29**, 8138–8144.
Dreusicke, D., Karplus, P. A. & Schulz, G. E. (1988). *J. Mol. Biol.* **199**, 8138–8144.
Dreusicke, D. & Schulz, G. E. (1986). *FEBS Lett.* **208**, 301–304.
Duckworth, D. H. & Bessman, M. J. (1967). *J. Biol. Chem.* **242**, 2877–2885.
Higgins, D., Bleasby, A. & Fuchs, R. (1992). *CABIOS*, **8**, 189–191.
Matthews, B. W. (1968). *J. Mol. Biol.* **3**, 491–497.
Müller, C. W. & Schulz, G. E. (1992). *J. Mol. Biol.* **224**, 159–167.
Otwinowski, Z. (1993). *Proceedings of the CCP*4 *Study Weekend. Data Collection and Processing*, pp. 56–62. Warrington, England: Daresbury Laboratory.
Stehle, T. & Schulz, G. E. (1990). *J. Mol. Biol.* **211**, 249–254.

MEETING REPORT

Acta Cryst. (1996). D52, 228–234

Report of a Workshop on the Use of Statistical Validators in Protein X-ray Crystallography*

ELEANOR DODSON,[a] GERARD J. KLEYWEGT[b] AND KEITH WILSON[a]

[a]*Department of Chemistry, University of York, York YO1 5DD, England, and* [b]*Department of Molecular Biology, Biomedical Center, University of Uppsala, Box 590, S-75124 Uppsala, Sweden*

(Received 20 July 1995; accepted 16 August 1995)

1. Aim of the meeting

Investigators from many of the major groups involved in developing structure and phase refinement packages were present. There were also representatives from the Brookhaven Protein Data Bank (PDB) and the International Union of Crystallography (IUCr), as well as other interested crystallographers. The program is given in the *Appendix*. The objectives were to review methods of refinement, with particular reference to new techniques and to evaluate methods for assessing the correctness and accuracy of refined atomic parameters. In addition efforts to assess density-modification routines and their power of phase improvement were discussed: articles by the three major contributors in this area are being published in this volume and are not addressed in detail here.

2. Questions addressed

2.1. *Why is refinement important?*

Reliable structural information and a proper awareness of the significance of the atomic parameters is needed to extract physical, chemical and biological information (energy, bond lengths, folding pathways, binding and enzymatic mechanisms and interactions) sensibly from the structure. All of these are distorted by errors in coordinates. An X-ray experiment is complicated and time consuming and it is unlikely that the results will be repeated independently by other investigators. This places an unavoidable responsibility on the crystallographer carrying out the research to ensure that the final model is a realistic representation of the available data.

2.2. *Why is protein refinement difficult?*

Small-molecule crystallographers manage analysis and refinement of their structures with very few problems. Why are proteins different? Macromolecular crystals present several particular problems in refinement.

For macromolecular crystals, the unit cell is big, and there is a very large number of X-ray data to collect, all of which are very weak. The signal-to-noise ratio is low. It is, therefore, not usually possible to collect data to atomic resolution as is normal for small molecules. The data available often suffer from both systematic and random errors. These arise because of the crystal size, problems of mounting, absorption, crystal decay and sometimes non-isomorphism between different crystals.

Protein crystals have an additional problem. There is usually a high solvent content, and the crystal forces are weak. Some parts of the chain may not be crystalline at all, and others may have high thermal motion. This means that not all the unit cell can be properly parameterized. This is true for almost all proteins, not just those which diffract to lower resolution. This problem particularly reduces the intensity of the high-resolution data. In addition it leads to severe effects of radiation damage. The distribution of amplitudes is not Wilsonian, and, therefore, scaling is more difficult.

These two problems mean that experimental data extend to limited resolution, typically to a maximum limit in the range 3–2 Å. Some structures generate data to 1.5 Å or better and allow more detailed analysis.

This means that the ratio of observations to parameters to be fitted is too low for conventional least-squares minimization to converge.

3. Improvement in refinement possibilities for proteins

3.1. *Data quality*

There has been a general improvement in the quality of diffraction data obtained over the past five years. Several factors have contributed.

Two-dimensional detectors, especially image plates, are now routinely used, they have a greater dynamic range than film, and are less prone to some of the systematic errors.

Cryogenic freezing techniques mean that crystal lifetime is now no longer a problem, and the quality of the diffraction does not degenerate significantly during the experiment. They often extend the limiting resolution obtainable.

* Meeting title: *Integrated procedures for recording and validating results of 3-D structural studies of biological macromolecules.* Held in York, April 1995, under the auspices of the EU Validation Network with additional support from the Collaborative Computational Project, Number 4. EU Contract number: BIO2CT-920524.

© 1996 International Union of Crystallography
Printed in Great Britain – all rights reserved