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

Crystallization of *Chlorella* deoxyuridine triphosphatase

Laura Badalucco,[a] Ishwari Poudel,[b] Mamoru Yamanishi,[a] Chandrasekhar Natarajan[a] and Hideaki Moriyama[a]*

[a]School of Biological Sciences, University of Nebraska-Lincoln, Lincoln, NE 68588-0118, USA, and [b]Department of Engineering Mechanics, University of Nebraska-Lincoln, Lincoln, NE 68583-0526, USA

Correspondence e-mail: hmoriyama2@unl.edu

Received 7 July 2011
Accepted 18 September 2011

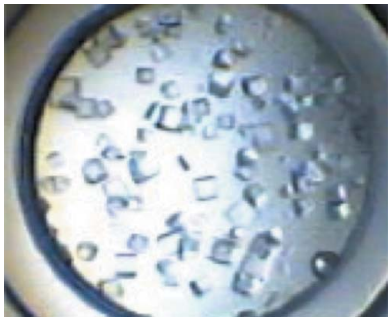
Deoxyuridine triphosphatase (dUTPase) is a ubiquitous enzyme that has been widely studied owing to its function and evolutionary significance. The gene coding for the dUTPase from the *Chlorella* alga was codon-optimized and synthesized. The synthetic gene was expressed in *Escherichia coli* and recombinant core *Chlorella* dUTPase (chdUTPase) was purified. Crystallization of chdUTPase was performed by the repetitive hanging-drop vapor-diffusion method at 298 K with ammonium sulfate as the precipitant. In the presence of 2′-deoxyuridine-5′-[(α,β)-imido]triphosphate and magnesium, the enzyme produced die-shaped hexagonal *R*3 crystals with unit-cell parameters $a = b = 66.9$, $c = 93.6$ Å, $\gamma = 120^\circ$. X-ray diffraction data for chdUTPase were collected to 1.6 Å resolution. The crystallization of chdUTPase with manganese resulted in very fragile clusters of needles.

1. Introduction

Chlorella is a unicellular photosynthetic eukaryote belonging to the Viridiplantae phylum. Recently, the entire genome of *C. variabilis* NC64A has been published and its co-evolution with *Chlorella* viruses has been demonstrated (Blanc *et al.*, 2010). *Chlorella* viruses are double-stranded DNA viruses with many genes (Van Etten *et al.*, 2010). For instance, the chlorovirus PBCV-1 has been predicted to have 366 coding sequences (Li *et al.*, 1997; Kutish *et al.*, 1996; Lu *et al.*, 1995, 1996). Curiously, *Chlorella* viruses have many genes encoded in their genome but very few of these genes have been functionally annotated (Van Etten, 2003). One way of addressing the host–virus interaction is to perform comparative studies. We chose *Chlorella* deoxyuridine triphosphatase enzyme as a model system to unravel the host–virus relationship and studies have been successfully conducted on the structure and function of this enzyme (Zhang *et al.*, 2005; Homma & Moriyama, 2009).

dUTPase is an enzyme that catalyzes the hydrolysis of deoxyuridine triphosphate (dUTP) to deoxyuridine monophosphate (dUMP) and pyrophosphate (Mol *et al.*, 1996). The enzyme belongs to the pyrimidine-biosynthesis pathway and provides a substrate, dUMP, to the downstream enzyme thymidylate synthetase (Carreras & Santi, 1995). dUTPase plays a key role in the production of deoxythymidine triphosphate (dTTP; Ladner *et al.*, 1996). In order to avoid uracil misincorporation into DNA (Tinkelenberg *et al.*, 2002), organisms facilitate higher dUTPase levels in dividing cells (Yan *et al.*, 2011; Traut, 1994). Viruses which replicate in environments rich in uracil encode dUTPase (Spatz *et al.*, 2007) and/or uracil-DNA glycosylase (Boyle *et al.*, 2011) to avoid the threat of misincorporation.

The structures of dUTPases can be divided into three categories: trimers, trimer-mimicking monomers and nucleocapsid dUTPases (Vértessy & Tóth, 2009). Most belong to the trimer form and have five important motifs. In human dUTPase (Mol *et al.*, 1996), the motif sequences are motif 1, AGYDL; motif 2, PRSG; motif 3, GVIDE-DYRG; motif 4, RIAQ; and motif 5, RGSGGFGST. In the plant *Arabidopsis thaliana* (PDB entry 2p9o; Bajaj & Moriyama, 2007) and in *Saccharomyces cerevisiae* (baker's yeast; PDB entry 3p48; Tchigvintsev *et al.*, 2011) the five motifs are conserved, with some



replacements (Supplementary Material Fig. S1*a*[1]). However, the key residues forming the reactive intermediates are identical to those in human dUTPase. Although the evolutionary and phylogenic relationships between baker's yeast and *Chlorella* virus are unclear, they have sequence variations in motifs 3 and 5 and share the same level of kinetic parameters, including K_m and V_{max}. The putative *Chlorella* dUTPase has 171 amino acids and the sequence starting at position 28 can be aligned well with the human dUTPase sequence starting at position 22. The five motifs were also well conserved in *Chlorella* dUTPases, except for the differences Tyr (human 48)/Phe (*Chlorella* 29) in motif 1 and Phe (human 158)/Tyr (*Chlorella* 139) in motif 5. Since the Phe in motif 5 is critical, a potential compensatory mutation may take place in motif 1 in *Chlorella* dUTPase. In order to identify the quaternary structure and address the structural effects of the potential compensatory mutations, we have initiated crystallographic studies of *Chlorella* dUTPase.

2. Methods and results

2.1. Construction of the gene-expression system

We synthesized two versions of the dUTPase gene, encoding long and core dUTPases. The longer version contains 60 amino acids as leading residues which the core dUTPase lacks. The genome of *C. variabilis* NC64A has a GC content of 67% and that of the putative dUTPase-coding region is 66%. This GC content is higher than those of the *Escherichia coli* genome (51%) and its dUTPase-coding region (55%). There were two reasons for converting to bacterial codon utilization for synthesis of the dUTPase gene. DNA synthesis with high GC content is usually difficult and expensive (Jensen *et al.*, 2010). A significant mismatch in codon utilization would limit protein expression (Holm, 1986). Therefore, we converted *Chlorella* codons to the optimum codon set of *E. coli* (Sharp & Li, 1987; Natarajan *et al.*, 2011). The GC contents of the final products were 58 and 55% for the long and core dUTPases, respectively. The DNAs synthesized by GenScript (Piscataway, New Jersey, USA) were cloned into the pGEX-2T expression vector (GE Healthcare, Piscataway, New Jersey, USA), which contains a removable glutathione *S*-transferase (GST) tag, using restriction sites *Bam*HI and *Eco*RI.

2.2. Production of *Chlorella* dUTPase

The expressed dUTPases were purified by GST affinity chromatography in batch mode (Fig. 1) and subjected to thrombin cleavage as described previously (Zhang *et al.*, 2005; Homma & Moriyama, 2009). After a second GST affinity chromatography step to remove unwanted minor protein fragments, the dUTPases were passed through SP resin at pH 7.4. The proteins were purified to homogeneity. The gene construction was such that a glycylserine (GS) peptide remained in the product. The long dUTPase contained 233 amino acids, consisting of GS, the 60 amino acids of the leading peptide deduced from the preliminary genome annotation and the 171 amino acids of the entire dUTPase, and was unstable. Much protein was lost during concentration and the dUTPase had a tendency to aggregate. The short dUTPase contained 146 amino acids, consisting of GS plus those of a truncated dUTPase that lacks 27 residues at the amino-terminus. The core dUTPase remained active for over a year in buffer at 273 K in an ice–water bath.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: FW5325).

2.3. Crystallization of *Chlorella* dUTPase

The crystallization of the core dUTPase was performed by the hanging-drop vapor-diffusion method (McPherson, 1990) using the NeXtal crystallization apparatus (Qiagen, Valencia, California, USA; Bajaj & Moriyama, 2007; Homma & Moriyama, 2009), which has movable lids and fixed cups on a base plate. Protein solutions consisting of 20 mg ml^{-1} dUTPase (approximately 0.4 m*M* as the trimer) in 50 m*M* Tris–HCl buffer at various pH values (8.2, 8.5 and 8.9) were prepared using Amicon Ultra-0.5 centrifugal filter devices (Millipore, Billerica, Massachusetts, USA). From the original buffer, a dialysis ratio of greater than 1:250 was achieved by repeated operations. Drops were made up of 3 µl protein solution, 1 µl 5 m*M* $MgCl_2$ or $MnCl_2$ and 1 µl 5 m*M* 2′-deoxyuridine-5′-[(α,β)-imido]triphosphate (dUMPNPP; Jena Bioscience, Jena, Germany). The drops were equilibrated against 2.2 *M* ammonium sulfate in reverse-osmosis water (RiOs, Millipore) at 298 K. There was no immediate precipitation under any of the conditions.

After 3 d there was no precipitation in the crystallization drop with magnesium ions. Therefore, the concentration of ammonium sulfate was increased by moving the lid with the protein drop to another crystallization cup containing a higher concentration of ammonium sulfate. After a further 5 d, a 2 µl drop of 20%(*w*/*v*) polyethylene glycol (PEG) 8000 (Hampton Research, Aliso Viejo, California, USA) was added to the protein drop. Initially, rough layered crystals were found in the pH 8.3 bin after two weeks (crystal type 1; Fig. 2*a*). A type 1 crystal was transferred into a droplet of 1:100 diluted Izit Crystal Dye (methylene blue trihydrate; Hampton Research) in 3 *M* ammonium sulfate. The crystal absorbed the dye, confirming that it was a protein crystal. A type 1 crystal was mounted in a capillary and exposed to X-rays on a SMART Apex CCD system using a Mo tube (wavelength 0.711 Å; Bruker AXS, Madison, Wisconsin, USA). Diffraction was not observed from the type 1 crystal.

In order to improve the crystal quality of core *Chlorella* dUTPase, repetitive crystallizations were performed at pH 8.3. All lids with hanging protein drops were moved to their original crystallization bins containing 2.2 *M* ammonium sulfate and the crystals were allowed to dissolve for 3 d. Because some crystals had been used in the diffraction experiments, the concentration of protein was lower.

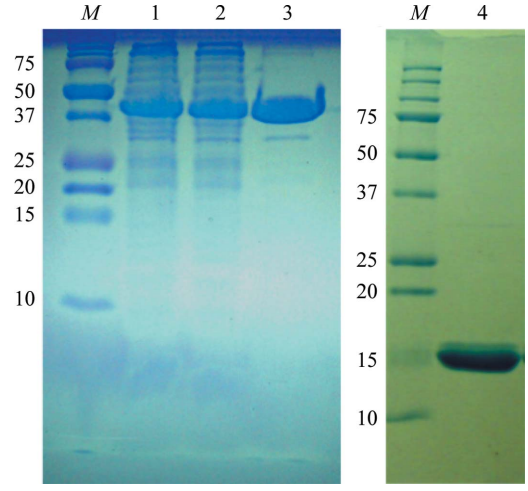
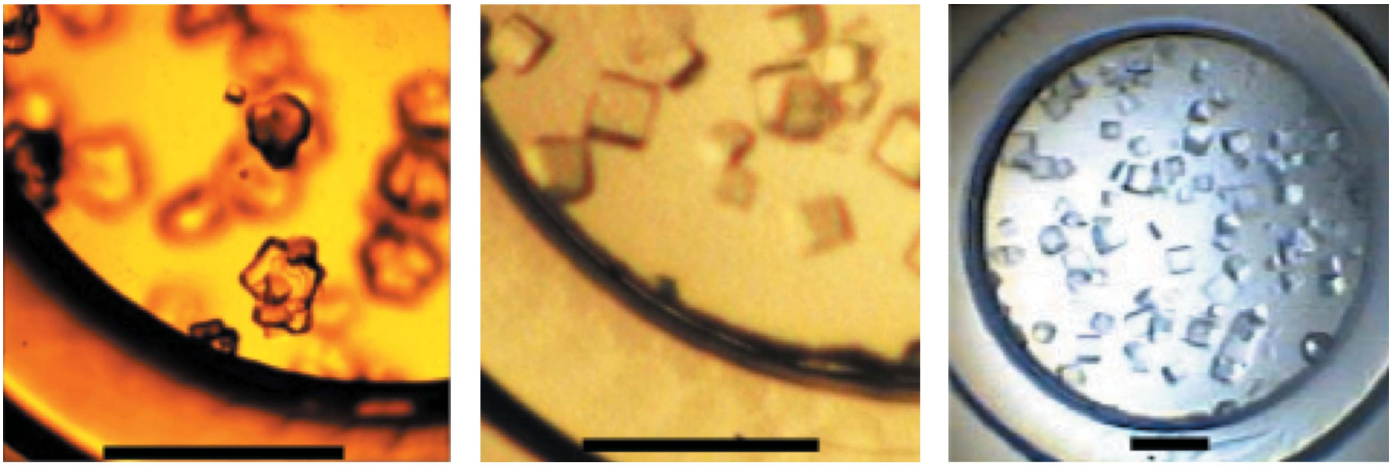


Figure 1
SDS–PAGE images indicating the progress of purification of *Chlorella* core dUTPase. Lane *M* in both gels contain molecular-weight markers (labeled in kDa). Lane 1, total cell lysate. Lane 2, cell supernatant. Lane 3, the glutathione *S*-transferase (GST; 26 kDa) fusion dUTPase (41 kDa) after batch glutathione column chromatography. Lane 4, purified *Chlorella* core dUTPase (15 kDa).



(*a*) (*b*) (*c*)

Figure 2
Two types of core *Chlorella* dUTPase crystals. (*a*) Type 1 crystals found in the initial crystallization using 2.5 *M* ammonium sulfate and a drop of 20% PEG 8000. (*b*, *c*) Type 2 crystals in the same crystallization well after repetitive crystallizations. The scale bars indicate 1 mm.

Each drop was moved over a crystallization solution containing 3.0 *M* ammonium acetate and the solution was confirmed to be turbid with microcrystals after overnight incubation. The drops were returned to the bin containing 2.2 *M* ammonium sulfate and the concentration of ammonium sulfate was increased to 2.7 *M* over 3 d in 0.1 *M* increments. As phase separation took place in the drop, die-shaped crystals formed (type 2; Fig. 2*b* and 2*c*).

2.4. Diffraction data collection

X-ray diffraction experiments were performed using a MicroMax-007 rotating-anode generator (Cu $K\alpha$, 1.542 Å) with Blue optics and an R-AXIS VI^{++} imaging plate with cryoequipment (Rigaku, The Woodlands, Texas, USA). Although 20%(*w*/*v*) trehalose and 3 *M* ammonium sulfate did not protect crystal quality during flash-cooing, the crystals were successfully mounted using 20%(*w*/*v*) glycerol and 3 *M* ammonium sulfate as a cryoprotectant. The crystal diffracted to 1.6 Å resolution. A complete data set was collected from a single crystal. The recorded diffraction images were indexed, integrated and scaled using *HKL*-2000 (HKL Research Inc.; Otwinowski & Minor, 1997). Initially, the crystal was treated as belonging to the trigonal space group *R*3. However, after evaluating the self-rotation function using the *MOLREP* program from the *CCP*4 program suite (Winn *et al.*, 2011), the space group was designated as hexagonal *R*3 (Fig. 3) owing to the additional twofold axis. Analysis of the unit-cell content assuming the hexagonal *R*3 space group and a monomer molecular mass of 15 360 Da gave a Matthews coefficient of 2.63 Å^3 Da^{-1} with a solvent content of 53.19% when the number of molecules in the asymmetric unit was assumed to be one. Thus, the type 2 crystals belonged to the hexagonal space group *R*3, with unit-cell parameters $a = b = 66.9$, $c = 93.6$ Å, $\gamma = 120^\circ$. Crystal parameters and data-collection and structure-refinement statistics are summarized in Table 1 (see also Supplementary Fig. 2).

2.5. Preliminary structure analysis

Preliminary structural analysis was performed using the *PHENIX* software package (Adams *et al.*, 2010). *HKL*-2000 and *phenix.xtriage* indicated a lower $I/\sigma(I)$ at 2.2 Å resolution owing to ice formation. However, we judged that the diffraction data were still usable. To solve the structure using the molecular-replacement method, a search model was constructed using the *SWISS-MODEL* server (Arnold *et al.*, 2006) using chain *A* of human dUTPase (PDB entry 3ehw; Takács *et al.*, 2009) as a template in the alignment mode. After locating the search model in the crystal lattice using *AutoMR* and assuming two monomers per asymmetric unit, a new model was built using *AutoBuild*. The best model had a correlation coefficient of 0.81 and a crystallographic *R* factor of 0.33. With a combination of repetitive *phenix.refine* runs and visual inspection, the crystallographic $R_{\rm free}$ factor reached 0.18. The current model explains the electron density that covers residues 7–126 of the 146 residues in the resolution range 27.0–1.7 Å. Although the N- and C-termini were not found in the electron-density map, the molecule formed a trimeric quaternary structure (Supplementary Fig. 1*b*). This structure has been deposited in the Protein Data Bank (Berman *et al.*, 2000) as entry 3so2.

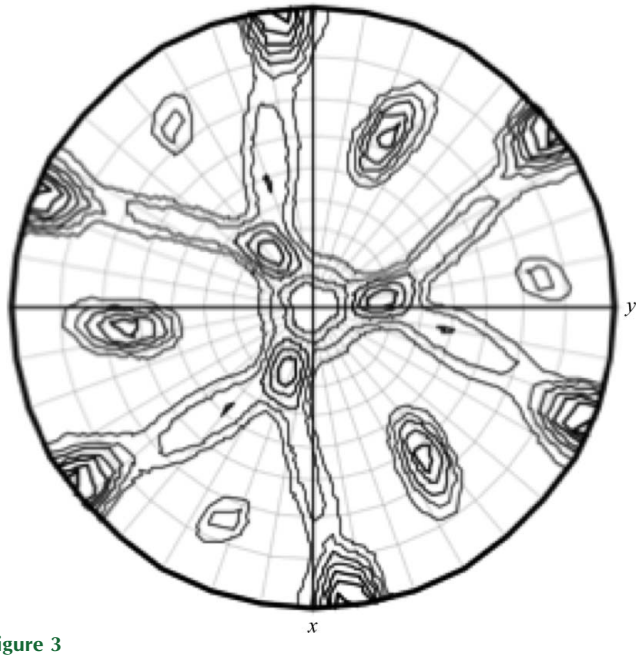


Figure 3
Self-rotation function: stereographic projection of the $\kappa = 180$ section for *Chlorella* core dUTPase assuming the trigonal space group *R*3 calculated with reflection data from 28 to 10 Å resolution and a sphere of 23 Å using the *MOLREP* program (Winn *et al.*, 2011). Spherical polar angles are designated as follows: φ, the angle from the Cartesian *x* axis (*a*) on the *xy* plane (*ab*); ω, the angle from the *z* axis (*c*); κ, the rotation around the axis defined by $\varphi\omega$.

Table 1
Summary of crystallographic data for *Chlorella* core dUTPase.

The scale log from *HKL*-2000 and the wwPDB validation report (PDB entry 3so2) were used.

Space group	*R*3 (*H*) (No. 146)
Unit-cell parameters (Å, °)	a = 66.91, c = 93.63, γ = 120
Resolution (Å)	50–1.6
Total reflections	18497
Completeness (%)	97.1
R_{merge}	0.10
Refinement resolution (Å)	27–1.7
R_{work}	0.18
R_{free}	0.19
$B_{average}$ (Å^2)	27.0
No. of subunits in the asymmetric unit	1
Residues determined	7–126

3. Discussion

From the preliminary structure analysis, we identified that the core *Chlorella* dUTPase adopts a trimeric quaternary structure. We found that Phe29 of *Chlorella* dUTPase in motif 1 is in the same location as Tyr48 of human dUTPase. However, limited electron density was obtained in the C-terminal region, including the active site. Therefore, alternative crystallization is necessary in order to address the structural implication of the Phe (human 158)/Tyr (*Chlorella* 139) subsititution in motif 5. While the current model is limited, one way to produce better crystals would be by the addition of an inhibitor, such as the 3′-azido derivative of dideoxy-UTP (Palmén & Kvassman, 2010) or a β-branched acyclic nucleoside analogue (Baragaña *et al.*, 2011), and/or by causing a mutation in Tyr/Phe139 (Recio *et al.*, 2011). These modifications would not only stabilize the structure of the C-terminus but would also provide insight into the role of Tyr139.

We have been crystallizing a series of dUTPases from a plant (Bajaj & Moriyama, 2007), an alga (this study) and algal viruses (Homma & Moriyama, 2009). The sequence similarity among them was approximately 53%. Of the four dUTPases above, the plant and algal dUTPases were crystallized using ammonium sulfate as the major precipitant. *Chlorella* virus dUTPase and its mutated derivatives were crystallized using polyethylene glycol as the major precipitant. This difference in the precipitant corresponded to the grand average of hydropathy (GRAVY; Tobias *et al.*, 1991) values. The GRAVY values calculated using the *ProtParam* server (Gasteiger *et al.*, 2005) were −0.1 for the plant and algal dUTPases and 0.2 for the *Chlorella* virus dUTPases.

The concentration of taurine in the crystallization of the plant dUTPase was 250 times that of the protein. The addition of dUMP-NPP to the crystallization of *Arabidopsis* dUTPase required more ammonium sulfate. This observation was consistent with the crystallization of *Chlorella* dUTPase, which required 2.7 *M* ammonium sulfate. However, the crystal provided a very weak electron-density map at the termini. Further crystallization efforts will be necessary.

We thank Dr James Van Etten and Ms Kathleen Williams at the University of Nebraska-Lincoln (UNL) for their generous support. We would like to thank Dr Vivian Stojanoff and her team members at the National Synchrotron Light Source, Brookhaven National Laboratory for technical contributions. This work was supported in part by the Nebraska Tobacco Settlement Biomedical Research Development Funds (HM), The Institutional Development Award Program (IDeA) Networks for Biomedical Research Excellence (INBRE) Scholars Program of Nebraska (IP) and The Undergraduate Creative Activities and Research Experiences (UCARE) Program of UNL (LB). A portion of the X-ray diffraction studies were carried out on the X6A beamline of the National Synchrotron Light Source, Brookhaven National Laboratory.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). *Bioinformatics*, **22**, 195–201.
Bajaj, M. & Moriyama, H. (2007). *Acta Cryst.* F**63**, 409–411.
Baragaña, B., McCarthy, O., Sánchez, P., Bosch-Navarrete, C., Kaiser, M., Brun, R., Whittingham, J. L., Roberts, S. M., Zhou, X.-X., Wilson, K. S., Johansson, N. G., González-Pacanowska, D. & Gilbert, I. H. (2011). *Bioorg. Med. Chem.* **19**, 2378–2391.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Blanc, G. *et al.* (2010). *Plant Cell*, **22**, 2943–2955.
Boyle, K. A., Stanitsa, E. S., Greseth, M. D., Lindgren, J. K. & Traktman, P. (2011). *J. Biol. Chem.* **286**, 24702–24713.
Carreras, C. W. & Santi, D. V. (1995). *Annu. Rev. Biochem.* **64**, 721–762.
Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, pp. 571–607. Totowa: Humana Press.
Holm, L. (1986). *Nucleic Acids Res.* **14**, 3075–3087.
Homma, K. & Moriyama, H. (2009). *Acta Cryst.* F**65**, 1030–1034.
Jensen, M. A., Fukushima, M. & Davis, R. W. (2010). *PLoS One*, **5**, e11024.
Kutish, G. F., Li, Y., Lu, Z., Furuta, M., Rock, D. L. & Van Etten, J. L. (1996). *Virology*, **223**, 303–317.
Ladner, R. D., McNulty, D. E., Carr, S. A., Roberts, G. D. & Caradonna, S. J. (1996). *J. Biol. Chem.* **271**, 7745–7751.
Li, Y., Lu, Z., Sun, L., Ropp, S., Kutish, G. F., Rock, D. L. & Van Etten, J. L. (1997). *Virology*, **237**, 360–377.
Lu, Z., Li, Y., Que, Q., Kutish, G. F., Rock, D. L. & Van Etten, J. L. (1996). *Virology*, **216**, 102–123.
Lu, Z., Li, Y., Zhang, Y., Kutish, G. F., Rock, D. L. & Van Etten, J. L. (1995). *Virology*, **206**, 339–352.
McPherson, A. (1990). *Eur. J. Biochem.* **189**, 1–23.
Mol, C. D., Harris, J. M., McIntosh, E. M. & Tainer, J. A. (1996). *Structure*, **4**, 1077–1092.
Natarajan, C., Jiang, X., Fago, A., Weber, R. E., Moriyama, H. & Storz, J. F. (2011). *PLoS One*, **6**, e20176.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **4**, 307–326.
Palmén, L. G. & Kvassman, J. O. (2010). *J. Enzyme Inhib. Med. Chem.* **25**, 146–150.
Recio, E., Musso-Buendía, A., Vidal, A. E., Ruda, G. F., Kasinathan, G., Nguyen, C., Ruiz-Pérez, L. M., Gilbert, I. H. & González-Pacanowska, D. (2011). *Eur. J. Med. Chem.* **46**, 3309–3314.
Sharp, P. M. & Li, W.-H. (1987). *Nucleic Acids Res.* **15**, 1281–1295.
Spatz, S. J., Petherbridge, L., Zhao, Y. & Nair, V. (2007). *J. Gen. Virol.* **88**, 1080–1096.
Takács, E., Barabás, O., Petoukhov, M. V., Svergun, D. I. & Vértessy, B. G. (2009). *FEBS Lett.* **583**, 865–871.
Tchigvintsev, A., Singer, A. U., Flick, R., Petit, P., Brown, G., Evdokimova, E., Savchenko, A. & Yakunin, A. F. (2011). *Biochem. J.* **437**, 243–253.
Tinkelenberg, B. A., Hansbury, M. J. & Ladner, R. D. (2002). *Cancer Res.* **62**, 4909–4915.
Tobias, J. W., Shrader, T. E., Rocap, G. & Varshavsky, A. (1991). *Science*, **254**, 1374–1377.
Traut, T. W. (1994). *Mol. Cell. Biochem.* **140**, 1–22.
Van Etten, J. L. (2003). *Annu. Rev. Genet.* **37**, 153–195.
Van Etten, J. L., Lane, L. C. & Dunigan, D. D. (2010). *Annu. Rev. Microbiol.* **64**, 83–99.
Vértessy, B. G. & Tóth, J. (2009). *Acc. Chem. Res.* **42**, 97–106.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yan, N., O'Day, E., Wheeler, L. A., Engelman, A. & Lieberman, J. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 9244–9249.
Zhang, Y., Moriyama, H., Homma, K. & Van Etten, J. L. (2005). *J. Virol.* **79**, 9945–9953.