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

Cloning, expression, purification, crystallization and preliminary X-ray diffraction data of the *Pyrococcus horikoshii* RadA intein

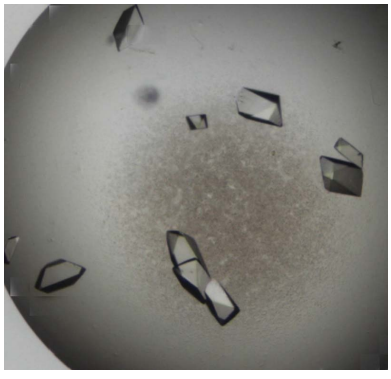
Andrzej Lyskowski,[a]‡§ Jesper S. Oeemig,[a]‡ Anniina Jaakkonen,[a] Katariina Rommi,[a] Frank DiMaio,[b] Dongwen Zhou,[c] Tommi Kajander,[a] David Baker,[b] Alexander Wlodawer,[c] Adrian Goldman[a]* and Hideo Iwaï[a]*

[a]Research Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki, PO Box 65, Helsinki FIN-00014, Finland, [b]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA, and [c]Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, Frederick, MD 21702, USA

‡ These authors contributed equally to this work.
§ Current address: ACIB GmbH c/o IMB, University of Graz, Humboldtstrasse 50/III, Graz 8010, Austria.

Correspondence e-mail: adrian.goldman@helsinki.fi, hideo.iwai@helsinki.fi

Received 22 November 2010
Accepted 5 March 2011



The RadA intein from the hyperthermophilic archaebacterium *Pyrococcus horikoshii* was cloned, expressed and purified for subsequent structure determination. The protein crystallized rapidly in several conditions. The best crystals, which diffracted to 1.75 Å resolution, were harvested from drops consisting of 0.1 *M* HEPES pH 7.5, 3.0 *M* NaCl and were cryoprotected with Paratone-N before flash-cooling. The collected data were processed in the orthorhombic space group $P2_12_12_1$, with unit-cell parameters a = 58.1, b = 67.4, c = 82.9 Å. Molecular replacement with *Rosetta* using energy- and density-guided structure optimization provided the initial solution, which is currently under refinement.

1. Introduction

Protein splicing is a post-translational modification in which an intervening sequence catalyzes self-excision from an immature precursor and ligates the two flanking polypeptide sequences to form a mature functional host protein (Paulus, 2000). The intervening sequence is termed an intein (*in*ternal pro*tein*) and the two flanking sequences are termed N- and C-exteins (*ex*ternal pro*teins*). The protein-splicing process catalysed by inteins does not require any external energy equivalent such as ATP or cofactors. Since a self-splicing intervening sequence was first reported in 1990, more than 500 putative inteins have been identified in all three kingdoms of life (archaea, bacteria and eukarya; Hirata *et al.*, 1990; Kane *et al.*, 1990; Perler, 2002). The unique concerted reactions catalyzed by inteins have led to a multitude of biotechnological applications such as protein-purification tags, protein semi-synthesis, protein cyclization, segmental isotope labelling and site-specific chemical modification (Chong *et al.*, 1997; Saleh & Perler, 2006; Iwai & Plückthun, 1999; Volkmann & Iwaï, 2010; Mootz, 2009). However, the biological functions of protein splicing still remain unclear because inteins do not seem to provide any advantages to host organisms. Therefore, inteins are often considered to be selfish genetic elements or parasitic proteins (Belfort *et al.*, 1995).

Inteins have been identified in several proteins belonging to a superfamily of DNA strand-exchange proteins (recombinases), including bacterial RecA, archaeal RadA and eukaryotic Rad51 proteins. RecA/RadA/Rad51 proteins play critical roles in homologous recombination. *Mycobacterium tuberculosis* RecA protein is interrupted by an intein which must be excised from the inactive precursor to produce a functional RecA recombinase protein (Davis *et al.*, 1992). Therefore, it has been suggested that protein splicing in RecA could be a novel target for antituberculosis drugs and that inhibitors of protein splicing could potentially become a novel category of drugs (Paulus, 2003). In archaea, RadA is a RecA equivalent that plays a pivotal role in homologous recombination. The RadA protein from *Pyrococcus horikoshii* (*Pho*) also contains a 172-amino-acid intein insertion. The intein is located within the Walker A ATPase motif of the RadA protein (Fig. 1), which is essential for the ATPase function of RadA. *Pho*RadA with the intein insertion is likely to remain inactive until the RadA intein is self-excised to produce a mature RadA protein. Thus, excision of the RadA intein could in principle control the function of the host RadA protein.

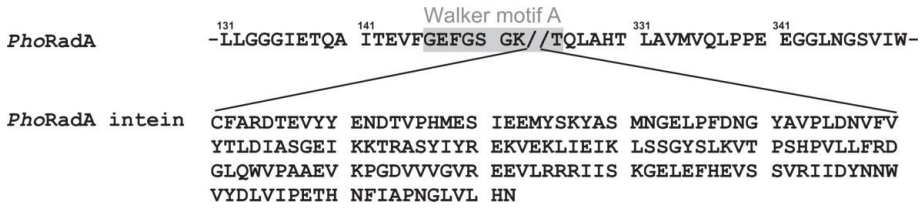


Figure 1
The insertion site of the *Pho*RadA intein within the *Pho*RadA protein. The primary structure of the *Pho*RadA intein is also shown. The Walker A motif is highlighted in grey.

Several three-dimensional structures of inteins have been determined by X-ray crystallography (Klabunde *et al.*, 1998; Hu *et al.*, 2000; Ichiyanagi *et al.*, 2000; Mizutani *et al.*, 2002; Sun *et al.*, 2005) and NMR spectroscopy (Johnson *et al.*, 2007; Oeemig *et al.*, 2009). However, all of the currently available structures are of inactive splicing domains either without any host-protein sequence or with only a few residues of the host protein. To better understand the possible regulatory function of inteins, it is of interest to elucidate how the insertion of inteins affects the structure of host proteins such as RecA/RadA. The structure of the RadA intein could also provide a structural basis for engineering new split inteins that could be used in protein ligation through protein *trans*-splicing (Aranko *et al.*, 2009; Oeemig *et al.*, 2009). As a first step towards a better understanding of the biological functions of inteins and their biotechnological applications, we have cloned, expressed, purified and crystallized an inactive variant of the RadA intein consisting of 174 amino acids from the hyperthermophilic archaebacterium *P. horikoshii* (*Pho*RadA intein).

2. Methods

2.1. Cloning, protein expression and purification of the *Pho*RadA intein

The gene for the *Pho*RadA intein was cloned from genomic DNA of *P. horikoshii* (ATCC 700860) using polymerase chain reaction (PCR) with the primers HK376 (5′-TTGGTACCTAGCTGAGTATTATGGAGAACAAGTC) and HK377 (5′-TATTATGAAAACGATACTGTACCACACATGGAATC). The PCR product was further amplified with HK359 (5′-GCTAGGGATACCGAAGTTTATTTAGAAAACGATAC) and HK376 to eliminate the *Kpn*I site in the RadA intein. The PCR product was again amplified using the primers HK390 (5′-GCATATGGCCTTTGCTAGGGATACC) and HK391 (5′-GGACTTGTTCTCCATAATGCTCAGTAAGCTTGTG) in which a C1A mutation in the *Pho*RadA intein and a T+1A mutation in the two-residue C-extein (TQ) were introduced to produce the inactive intein bearing 174 amino acids. The PCR product was subsequently digested with *Nde*I and *Hin*dIII restriction enzymes and ligated into pRSET-A (Invitrogen).

The *Pho*RadA intein was expressed in *Escherichia coli* strain ER2566 in LB medium with 100 µg ml^{-1} ampicillin at 310 K. At an OD_{600} of ∼0.6 the cells were induced with a final concentration of 0.1 m*M* isopropyl *β*-D-1-thiogalactoside (IPTG) and harvested by centrifugation at 8900*g* and 277 K for 10 min after 3 h induction. Subsequently, the cell pellet was resuspended in 50 m*M* Tris pH 7.9, 1 m*M* EDTA, 10 m*M* NaCl (buffer *A*) and flash-frozen in liquid nitrogen.

The thawed cells were lysed by heating to 348 K for 20 min and the cell debris was removed by centrifugation at 34 000*g* and 277 K for 40 min. DNase I was added to the supernatant to a concentration of 42 µg ml^{-1} and the supernatant was incubated at 310 K for 2.5 h. Digestion was analysed on a 1.2% agarose gel. The sample was then loaded onto a DEAE Sepharose FF 5 ml column (GE Healthcare) and washed with buffer *A*. The bound protein was eluted with a 25 ml linear gradient from 0 to 50% 2.5 *M* NaCl. A sample containing the *Pho*RadA intein was dialysed overnight against 10 m*M* sodium phosphate buffer pH 8.2. The dialysed sample was further purified by loading it onto a 1 ml MonoQ 5/50 GL column (GE Healthcare) and washing with 10 m*M* sodium phosphate pH 8.2. The bound protein was eluted with a 22 ml linear gradient from 0 to 20% 2.5 *M* NaCl. The fractions containing the *Pho*RadA intein were pooled and concentrated using a microcentrifugal device (GE Healthcare Vivaspin 6, 3 kDa molecular-weight cutoff).

2.2. Crystallization and optimization

The *Pho*RadA intein was concentrated to 7.1 mg ml^{-1} and the buffer was exchanged to Milli-Q grade water. The sitting-drop vapour-diffusion method was used with drops consisting of 100 nl protein solution and 100 nl reservoir solution from the Index HT screen (Hampton Research) produced using a Cartesian MicroSys pipetting robot in an Innovadyne SD2 96-well plate. The plate was stored at 293 K. To improve crystal quality, four crystal-producing conditions from the initial screen were selected for optimization. The conditions for further optimization were (i) 0.1 *M* HEPES pH 7.0,

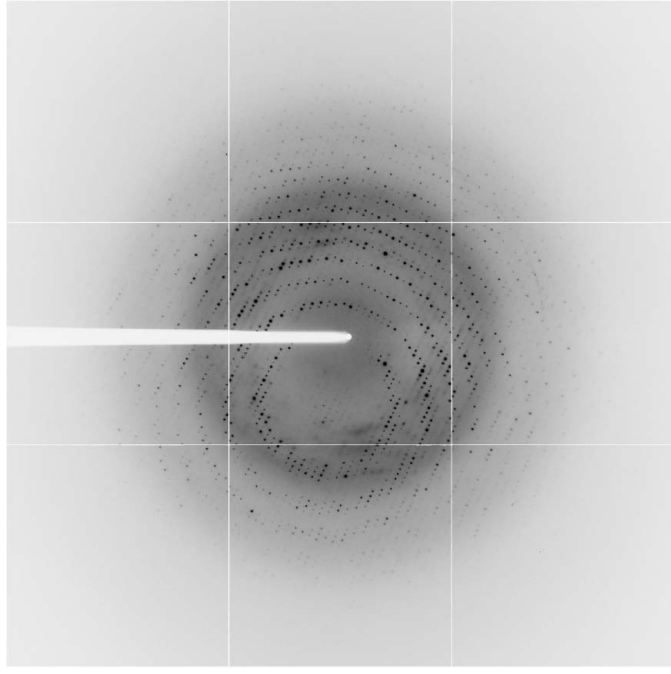


Figure 2
Diffraction image of a *Pho*RadA intein crystal. The image was prepared using the *ADXV* program (http://www.scripps.edu/~arvai/adxv.html). The resolution at the edge of the image is 1.75 Å. The data were collected on the ESRF ID23-1 beamline using an ADSC Quantum Q315R detector.

15% Tacsimate pH 7.0 and 2% PEG 3350, (ii) 0.1 *M* HEPES pH 7.5 and 0.5 *M* magnesium formate, (iii) 0.1 *M* Bis-Tris pH 5.5 and 2 *M* ammonium sulfate and (iv) 0.1 *M* HEPES pH 7.5 and 2 *M* ammonium sulfate. Grids varying selected components [8–20% Tacsimate pH 7.0 and 0–5% PEG 3350, 0.1–0.8 *M* magnesium formate, 1.5–2.5 *M* ammonium sulfate, PEG molecular weight (PEG 400, PEG 1000, PEG 3350 and PEG 8000) at PEG concentrations of 10 and 30% varying the pH from 7.0 (0.1 *M* HEPES) to 9.5 (0.1 *M* Tris)] were prepared. Two Innovadyne SD2 plates were set up for grid optimization screens with a drop size of 200 + 200 nl. One plate containing the grid optimization (96 drops) was prepared in which the protein concentration was constant and equal to that in the initial screen (7.1 mg ml^{-1}). This plate was stored and imaged at 293 K. The other plate with the same grid optimization was stored and imaged at 277 K. In this plate, two protein concentrations (7.1 and 3.6 mg ml^{-1}) were tested using two drop sites for grid optimization (a total of 2 × 96 drops).

Crystals were harvested from both the initial screen hits and the optimization grids. Crystals were vitrified with and without cryoprotectant. The cryoprotectants tested were PEG 400, Paratone-N, paraffin and 2-methyl-1,3-propanediol. The best diffracting crystal was harvested from the initial screen hit containing 0.1 *M* HEPES pH 7.5 and 3 *M* NaCl and was cryoprotected with Paratone-N. The best crystallization condition was not included in the initial optimization attempts because crystals grew significantly more slowly under this condition compared with the optimized conditions.

2.3. X-ray data collection

The best native data sets were collected at the European Synchrotron Radiation Facility (ESRF), Grenoble. One data set was collected to 2.2 Å resolution on beamline ID23-2 using a MAR 225 detector and another was collected to 1.75 Å resolution on ID23-1 using an ADSC Quantum Q315R detector at 100 K (Fig. 2).

The diffraction images were integrated and scaled using the *XDS* package (Kabsch, 2010). The automatic space-group assignment was further verified with the program *POINTLESS* (Evans, 2006). The results of data collection and processing for the best diffracting crystal are presented in Table 1.

3. Results and discussion

Initial screening using Hampton Research Index HT screen resulted in several crystals of the *Pho*RadA intein appearing within hours of setup (Fig. 3). Further crystals appeared after a few days or weeks. The crystals obtained initially diffracted to 2.2 Å resolution. In order

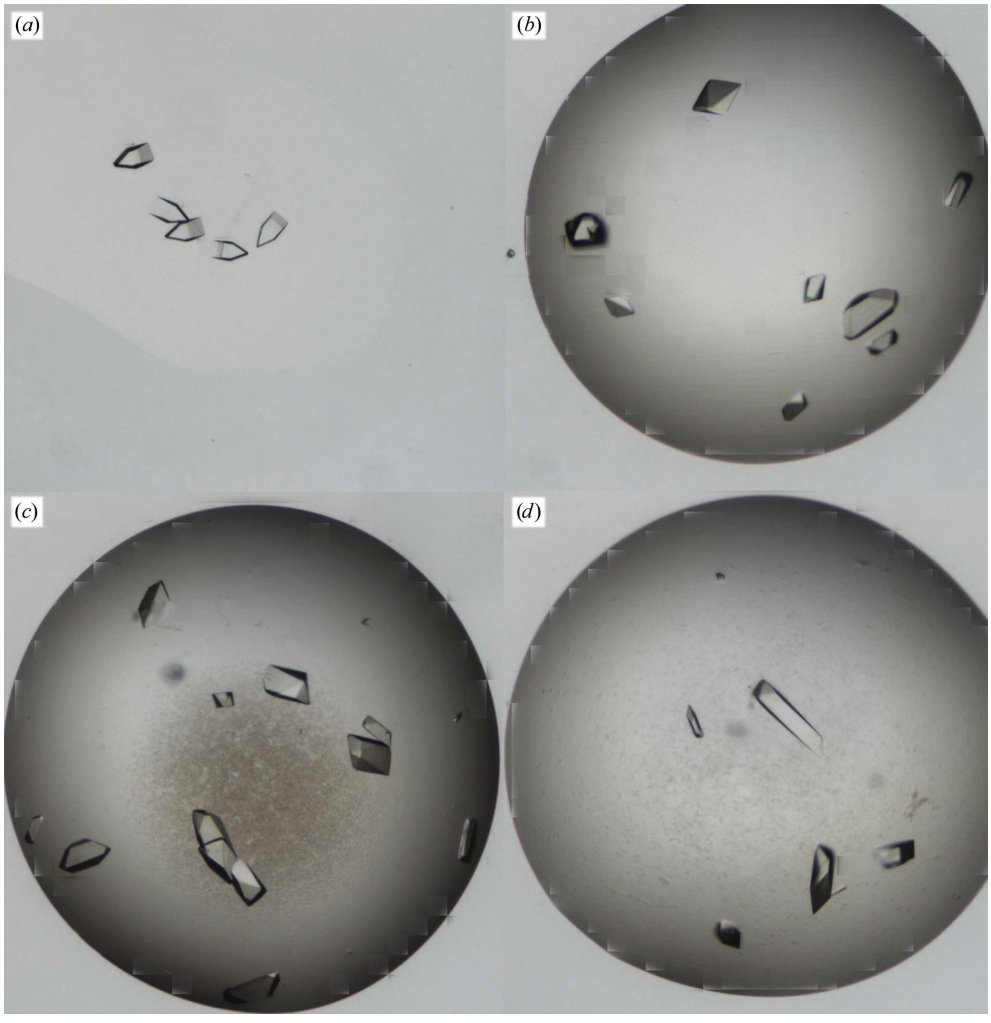


Figure 3
Crystals of the *Pho*RadA intein. (*a*, *b*) Crystals from the initial screen grown using (*a*) 0.1 *M* HEPES pH 7.5 and 3 *M* NaCl and (*b*) 0.1 *M* HEPES pH 7.0, 15% Tacsimate pH 7.0 and 2% PEG 3350. (*c*, *d*) Crystals grown from the grid optimization conditions. (*c*) 0.1 *M* HEPES pH 7.0, 15.2% Tacsimate pH 7.0. (*d*) 0.1 *M* HEPES pH 7.0, 10.4% Tacsimate pH 7.0 and 3.0% PEG 3350.

Table 1
Data-collection statistics.

Values in parentheses are for the outer shell.

Diffraction source	ESRF beamline ID23-1
Crystal system	Orthorhombic
Space group	$P2_12_12_1$
Unit-cell parameters (Å, °)	a = 58.1, b = 67.4, c = 82.9, $\alpha = \beta = \gamma = 90$
No. of molecules in unit cell, Z	2
Wavelength (Å)	0.97240
Detector	ADSC Quantum Q315R
Temperature (K)	100
Resolution range (Å)	33.7–1.75 (1.80–1.75)
No. of unique reflections	33297 (2662)
No. of observed reflections	234686 (19194)
Completeness (%)	99.3 (99.8)
Multiplicity	7.1 (7.2)
$\langle I/\sigma(I)\rangle$	17.3 (2.5)
$R_{\rm meas}$† (%)	7.6 (102.7)
$R_{\rm merge}$‡ (%)	7.0 (95.3)
Data-processing software	*XDS*

† $R_{\rm meas} = \sum_{hkl}[N/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$.

to improve the diffraction limits, we optimized selected crystallization conditions (see §2.2).

Unit-cell content analysis (Kantardjieff & Rupp, 2003) showed that the asymmetric unit could contain one or two molecules with ∼69% or ∼38% solvent content, respectively. Preliminary attempts to solve the structure of the *Pho*RadA intein using automated and manual molecular replacement were unsuccessful. The *BALBES* molecular-replacement pipeline (Long *et al.*, 2008) identified 11 similar structures, with the highest sequence identity being 21.5%. The *MrBUMP* (Keegan & Winn, 2008) automated pipeline for molecular replacement identified three templates, with the top one having 27.6% sequence identity. Both pipelines were able to place two copies of the selected template in the asymmetric unit using their respective molecular-replacement engines *MOLREP* (Vagin & Teplyakov, 2010) and *Phaser* (McCoy *et al.*, 2007). However, the quality of the solution was poor in both cases: the resulting electron density only partially covered the input templates and could not be interpreted, nor could the model be refined. Manually prepared templates for molecular replacement were derived from the available NMR structures (PDB code 2jmz; Johnson *et al.*, 2007), but these did not yield any promising solutions. We also attempted to use a three-dimensional envelope generated by the program *DAMAVER* (Volkov & Svergun, 2003) based on these structures as the input template.

Next, we used a preliminary NMR structure of the identical construct of the *Pho*RadA intein as a search model in molecular replacement using the program *Phaser*. Again, none of the attempts using the molecular-replacement approaches mentioned above using *MOLREP* or *Phaser* showed any promising results, which might be a consequence of the deviations (the r.m.s.d. was 3.6 ± 0.3 Å for the backbone atoms of residues 1–172) between the preliminary NMR structures used and the crystal structure of the *Pho*RadA intein. Finally, we used the same preliminary NMR structures as a starting model in a novel molecular-replacement methodology implemented in *Rosetta* with an energy- and electron-density-guided model-building and refinement protocol (Qian *et al.*, 2007; DiMaio *et al.*, 2011) to successfully refine the structure. The complete structure determination and refinement are in progress and will be reported elsewhere.

This work was supported in part by a grant from the Academy of Finland (1131413) and by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. The crystallization facility is supported by Biocenter Finland. JSO acknowledges the National Doctoral Programme in Informational and Structural Biology (ISB) and the Centre for International Mobility (CIMO) for financial support. AL was supported by Marie Curie Fellowship under FP7 (Project TrimBAT, No. 219889).

References

Aranko, A. S., Züger, S., Buchinger, E. & Iwaï, H. (2009). *PLoS One*, **4**, e5185.
Belfort, M., Reaban, M. E., Coetzee, T. & Dalgaard, J. Z. (1995). *J. Bacteriol.* **177**, 3897–3903.
Chong, S., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Vence, L. M., Perler, F. B., Benner, J., Kucera, R. B., Hirvonen, C. A., Pelletier, J. J., Paulus, H. & Xu, M.-Q. (1997). *Gene*, **192**, 271–281.
Davis, E. O., Jenner, P. J., Brooks, P. C., Colston, M. J. & Sedgwick, S. G. (1992). *Cell*, **71**, 201–210.
DiMaio, F., Terwilliger, T. C., Read, R. J., Wlodawer, A., Oberdorfer, G., Wagner, U., Valkov, E., Alon, A., Axelrod, H. L., Fass, D., Das, D., Vorobiev, S. M., Iwaï, H., Pokkuluri, P. R. & Baker, D. (2011). *Nature (London)*. doi:10.1038/nature09964.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Hirata, R., Ohsumi, Y., Nakano, A., Kawasaki, H., Suzuki, K. & Anraku, Y. (1990). *J. Biol. Chem.* **265**, 6726–6733.
Hu, D., Crist, M., Duan, X., Quiocho, F. A. & Gimble, F. S. (2000). *J. Biol. Chem.* **275**, 2705–2712.
Ichiyanagi, K., Ishino, Y., Ariyoshi, M., Komori, K. & Morikawa, K. (2000). *J. Mol. Biol.* **300**, 889–901.
Iwai, H. & Plückthun, A. (1999). *FEBS Lett.* **459**, 166–172.
Johnson, M. A., Southworth, M. W., Herrmann, T., Brace, L., Perler, F. B. & Wüthrich, K. (2007). *Protein Sci.* **16**, 1316–1328.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M. & Stevens, T. H. (1990). *Science*, **250**, 651–657.
Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
Keegan, R. M. & Winn, M. D. (2008). *Acta Cryst.* D**64**, 119–124.
Klabunde, T., Sharma, S., Telenti, A., Jacobs, W. R. & Sacchettini, J. C. (1998). *Nature Struct. Biol.* **5**, 31–36.
Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* D**64**, 125–132.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Mizutani, R., Nogami, S., Kawasaki, M., Ohya, Y., Anraku, Y. & Satow, Y. (2002). *J. Mol. Biol.* **316**, 919–929.
Mootz, H. D. (2009). *Chembiochem*, **10**, 2579–2589.
Oeemig, J. S., Aranko, A. S., Djupsjöbacka, J., Heinämäki, K. & Iwaï, H. (2009). *FEBS Lett.* **583**, 1451–1456.
Paulus, H. (2000). *Annu. Rev. Biochem.* **69**, 447–496.
Paulus, H. (2003). *Front. Biosci.* **8**, s1157–s1165.
Perler, F. B. (2002). *Nucleic Acids Res.* **30**, 383–384.
Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J. & Baker, D. (2007). *Nature (London)*, **450**, 259–264.
Saleh, L. & Perler, F. B. (2006). *Chem. Rec.* **6**, 183–193.
Sun, P., Ye, S., Ferrandon, S., Evans, T. C., Xu, M.-Q. & Rao, Z. (2005). *J. Mol. Biol.* **353**, 1093–1105.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Volkmann, G. & Iwaï, H. (2010). *Mol. Biosyst.* **6**, 2110–2121.
Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 860–864.