crystallization communications

Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of YvoA from *Bacillus subtilis*

Marcus Resch,[a] Heide Marie Roth,[a] Mathias Kottmair,[a] Madhumati Sevvana,[a] Ralph Bertram,[b]‡ Fritz Titgemeyer[b]§ and Yves A. Muller[a]*

[a]Lehrstuhl für Biotechnik, Department of Biology, Friederich-Alexander University Erlangen-Nuremberg, Henkestrasse 91, D-91052 Erlangen, Germany, and [b]Lehrstuhl für Mikrobiologie, Department of Biology, Friedrich-Alexander University Erlangen-Nuremberg, Staudtstrasse 5, D-91052 Erlangen, Germany

‡ Present address: Mikrobielle Genetik, Eberhard Karls University Tübingen, Waldhäuser Strasse 70/8, D-72076 Tübingen, Germany. § Present address: Fachbereich Oecotrophologie, Fachhochschule Münster, Corrensstrasse 25, D-48149 Münster, Germany.

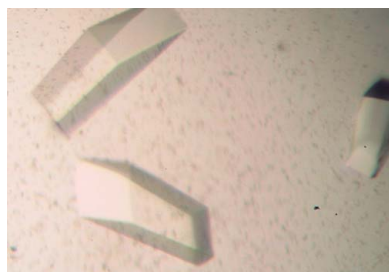Correspondence e-mail: ymuller@biologie.uni-erlangen.de

The putative transcriptional regulator protein YvoA (BSU35030) from *Bacillus subtilis* was cloned and heterologously expressed in *Escherichia coli*. The protein was purified by immobilized metal-affinity chromatography and size-exclusion chromatography and subsequently crystallized. A complete native data set was collected to 2.50 Å resolution. The crystals belonged to the monoclinic space group *C*2 and preliminary analysis of the diffraction data indicated the presence of approximately 12 molecules per asymmetric unit.

1. Introduction

Bacteria can grow on a variety of carbon sources. Their utilization is strictly regulated by globally and specifically acting transcription factors (Brueckner & Titgemeyer, 2002; Stuelke & Hillen, 2000). *N*-Acetylglucosamine (GlcNAc) is a preferred nutrient for many microorganisms as it is highly abundant and provides both carbon and nitrogen. For example, GlcNAc is the monomeric building block of the polysaccharide chitin on which many bacteria live. GlcNAc is also part of the bacterial cell wall and thus is of critical importance for cell growth and cell-wall turnover.

In the soil-dwelling bacterium *Bacillus subtilis*, the regulation of GlcNAc metabolism has been assumed to be under the control of the transcription regulator YvoA (Reizer *et al.*, 1999; Stuelke & Hillen, 2000). According to the current model and as inferred from studies in related bacteria (Alice *et al.*, 2003; Reizer *et al.*, 1999), GlcNAc is taken up in the absence of glucose by the *B. subtilis* cell *via* NagP (a GlcNAc-related permease) and is simultaneously phosphorylated, yielding GlcNAc-6-phosphate (GlcNAc-6-P; Mobley *et al.*, 1982; Vincent *et al.*, 2004). GlcNAc-6-P is further processed by the enzyme NagA (GlcNAc-6-P deacetylase), yielding glucosamine-6-phosphate (GlcN-6-P; Vincent *et al.*, 2004). GlcN-6-P can then either be channelled into peptidoglycan biosynthesis or deaminated to fructose-6-phosphate (Fru-6-P) by the enzyme NagB (GlcN-6-P deaminase; Vincent *et al.*, 2005). The genes encoding NagA and NagB are located within the same putative operon as the transcriptional regulator encoding the *yvoA* gene (Kunst *et al.*, 1997). YvoA has been shown to modulate the expression of a gene divergently transcribed from *nagP* (You *et al.*, 2005). Moreover, YvoA shares high sequence identity with DasR from *Streptomyces coelicolor*. The latter has been characterized as a pleiotropic transcriptional regulator and the DasR regulon comprises all of the abovementioned GlcNAc-related genes (Rigali *et al.*, 2006, 2008). In addition, the two regulators YvoA and DasR are likely to recognize identical operator DNA sequences, the so-called *dre* (DasR responsive element) sites (Rigali *et al.*, 2004). The *dre* sequence was first detected *in silico* by Rigali *et al.* (2004) and a consensus sequence containing the nucleotides 5′-TGG-TCTAGACCA-3′ was recently derived for *S. coelicolor* (Colson *et al.*, 2007).

The primary sequences identify YvoA and DasR as members of the GntR superfamily of repressor proteins. GntR-family members are characterized by an N-terminal winged-helix–turn–helix domain (wHTH; Fig. 1*a*) which is responsible for DNA binding and recog-

nition (Haydon & Guest, 1991; Rigali *et al.*, 2004; van Aalten *et al.*, 2000). The C-terminal moieties of the GntR proteins are highly diverse and allow further division of the GntR family into six sub-families, namely AraR, FadR, YtrA, MocR, PlmA and HutC (Lee *et al.*, 2003; Rigali *et al.*, 2002). YvoA belongs to the HutC subfamily (Rigali *et al.*, 2002) and like all HutC members contains a chorismate lyase fold, also termed an UTRA domain (Aravind & Anantharaman, 2003; Fig. 1*a*), at the C-terminus. Members of the HutC family of transcriptional regulators are widely distributed in bacteria and control various biological processes including antibiotic production, sensing of nutritional status, growth, proliferation and development (Rigali *et al.*, 2006, 2008). They are thought to recognize a large variety of effector molecules *via* their variable C-terminal domain (Aravind & Anantharaman, 2003) and, as is the case for many bacterial transcription regulators, binding of small-molecule effectors alters their DNA-binding affinities. The crystal structures of several HutC-family proteins have been determined to date (Dong *et al.*, 2007; Gorelik *et al.*, 2006; Rezacova *et al.*, 2007), but detailed mechanistic insight is currently hampered by a paucity of full-length structures of HutC-family members and the absence of the structure of any HutC-family member in complex with DNA or effector molecules.

In this report, we describe the cloning of full-length *yvoA* from genomic DNA, the heterologous overexpression of the protein in *Escherichia coli*, its purification to homogeneity and successful crystallization. These results lay the foundations for future studies aiming towards understanding the molecular mechanism by which YvoA regulates gene expression.

2. Experimental methods

2.1. Cloning

Genomic *yvoA* DNA (BSU35030) from *B. subtilis* 168 was used as a template for the polymerase chain reaction (PCR). The DNA fragment was amplified using the forward and reverse primers yvoA_ov_fw (5′-GGAACATGCTGACATATGAATATCAAT-3′) and yvoA_ov_rev (5′-CATATCGTCCGGATCCAGACCAGT-3′), respectively, and first cloned into the pET-3c vector (Novagen, EMD Chemicals Inc., Darmstadt, Germany) using the restriction enzymes *Nde*I and *Bam*HI. To improve the protein-expression levels and to simplify purification, the *yvoA* gene was subsequently transferred into the pET-15b vector (Novagen, EMD Chemicals Inc., Darmstadt, Germany). The gene was cloned in fusion with an N-terminal hexa-histidine tag followed by a thrombin cleavage site *via* *Nde*I/*Bam*HI. Successful cloning was confirmed by DNA-sequence analysis.

2.2. Expression and purification

His-tagged YvoA was overexpressed in *E. coli* BL21 (DE3) pREP4-*groESL* (Novagen). Cells were grown in LB medium in the presence of 100 µg ml^{-1} ampicillin and 10 µg ml^{-1} kanamycin at 310 K to an OD$_{600}$ of 0.8 before the temperature was lowered to 293 K and protein expression was induced with 1 m*M* isopropyl β-D-thiogalactopyranoside (IPTG). After 6 h, the cells were harvested by centrifugation and the pellet was stored at 193 K. For protein puri-fication, 3 g of the pellet was thawed and resuspended in 50 m*M* phosphate buffer pH 8.0 containing 300 m*M* NaCl, 1 m*M* phenyl-methanesulfonylfluoride (PMSF) and 1 mg ml^{-1} lysozyme prior to disruption by sonication. After 1 h centrifugation at 95 000*g* and 277 K the supernatant was filtered through a 0.45 µm filter (Millipore, Schwalbach/Ts., Germany) and subsequently loaded onto a 2 ml HisTrap column (GE Healthcare, Munich, Germany). YvoA was eluted with a linear gradient of 0–500 m*M* imidazole in 50 m*M* phosphate buffer pH 8.0 supplemented with 300 m*M* NaCl. Fractions containing the target protein were pooled and the His$_6$ tag was cleaved off overnight at 277 K using thrombin at a final ratio of 5 NIH units per milligram of YvoA. The sample was subjected to a size-exclusion chromatography column (HiLoad Superdex 200 26/60, GE Healthcare, Munich, Germany) pre-equilibrated in 20 m*M* Tris–HCl, 150 m*M* NaCl pH 7.5 in order to remove thrombin and the cleaved His$_6$ tag. The eluted protein was concentrated to 4.7–15.3 mg ml^{-1} using Vivaspin concentrators (Vivascience, Hanover, Germany). Protein that was >95% pure as judged by SDS–PAGE (Laemmli, 1970) was flash-frozen in liquid nitrogen and stored at 193 K until further use. Protein concentrations were determined spectro-photometrically at 280 nm employing an extinction coefficient of 16 390 M^{-1} cm^{-1} and a molecular weight of 27 790 Da. The mono-dispersity of the purified sample was confirmed by dynamic light scattering (DynaPro Titan, Wyatt Technology Corporation, Cali-fornia, USA).

2.3. Protein crystallization

Initial screening for crystallization conditions was performed using the sitting-drop vapour-diffusion method by mixing 400 nl YvoA (15.3 mg ml^{-1} in 20 m*M* Tris–HCl, 150 m*M* NaCl pH 8.0) with 200 nl reservoir solution and equilibrating the droplets against 50 µl reser-voir solution at 292 K at Hamburg HT-X facility, EMBL Hamburg (Mueller-Dieckmann, 2006). Microcrystals were obtained in 1.6 *M* ammonium sulfate, 2%(*w*/*v*) PEG 1000, 100 m*M* HEPES pH 7.5. These conditions were then optimized using the hanging-drop vapour-diffusion method. Diffraction-quality crystals were obtained by mixing 2 µl protein solution (4.7 mg ml^{-1} YvoA in 20 m*M* Tris–HCl, 150 m*M* NaCl pH 7.5) with 1 µl reservoir solution [100 m*M* HEPES pH 6.5, 1.26 *M* ammonium sulfate, 111 m*M* nondetergent
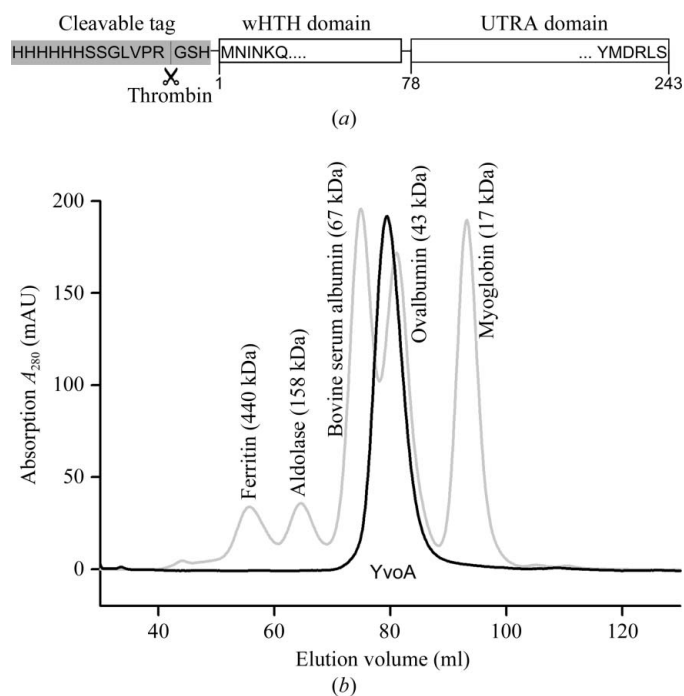
Figure 1
Domain organization (*a*) and size-exclusion chromatography (*b*) of YvoA from *B. subtilis* (black line) together with reference proteins (grey line) on a HiLoad Superdex 200 16/60 column. The apparent molecular weight of YvoA is 51 kDa, which corresponds to 1.8 times the molecular weight of the monomer.

sulfobetaine (NDSB) 256, 10 mM Na-EDTA and 1%(w/v) PEG 1000] and equilibrating against 700 µl reservoir solution at 292 K. Crystals grew to average dimensions of $0.2 \times 0.1 \times 0.1$ mm within two weeks.
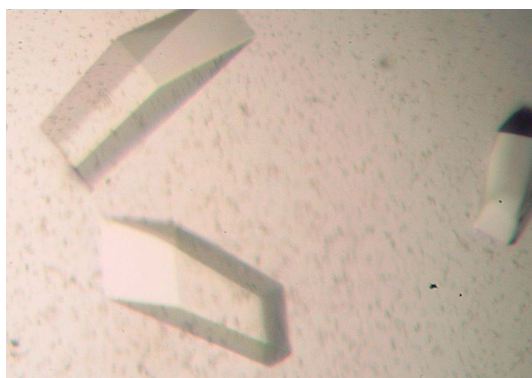
2.4. Diffraction data collection and processing

Before flash-freezing, the crystals were soaked for 30 s in mother liquor that consisted of reservoir solution to which 20%(v/v) glycerol had been added. A native diffraction data set was collected from a single YvoA crystal at 100 K on synchrotron beamline BL14.2 of the Free University Berlin at BESSY (Berlin, Germany). Data were collected at a wavelength of 0.91841 Å in 0.5° oscillations steps covering a total rotation range of 120°. Data were indexed and integrated using *XDS* and scaled using *XSCALE* (Kabsch, 1993). A self-rotation function was calculated using *GLRF* (Tong & Rossmann, 1997).

3. Results and discussion

Heterologous expression of full-length YvoA from *B. subtilis* in *E. coli* BL21 (DE3) pREP4-*groESL* cells only resulted in predominantly soluble protein when the culture was grown at 293 K. Recombinant protein expressed from pET-15b-*yvoA* was purified in a two-step procedure applying immobilized metal (Ni^{2+}) affinity and size-exclusion chromatography to give a final yield of approximately 50 mg highly pure protein per litre of culture medium. In order to investigate the monodispersity and the oligomeric state of YvoA, the purified sample was analyzed by dynamic light scattering. The sample gave a single peak with an apparent molecular weight of 64 kDa, corresponding to 2.3 times the molecular weight of the monomer. This finding indicates a dimeric assembly of YvoA in solution and is in accordance with the results from size-exclusion chromatography (Fig. 1*b*). Using ferritin (440 kDa), aldolase (158 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa) and myoglobin (17 kDa) as reference proteins, the apparent molecular weight of YvoA was calculated to be 51 kDa, which corresponds to 1.8 times the molecular weight of the monomer.

High-throughput screening for crystallization resulted in a single hit. Upon extensive optimization of the initial crystallization condition, high-quality protein crystals (Fig. 2) were obtained from ammonium sulfate with the additives NDSB 256 and Na-EDTA. Crystals were analyzed by SDS–PAGE and showed a band corresponding to full-length YvoA (data not shown). Diffraction data were collected from a single flash-frozen crystal to a resolution of 2.50 Å. The relevant data-collection and processing parameters are given in Table 1. The diffraction data of YvoA could be indexed in the

Figure 2
Crystals of YvoA (BSU35030) from *B. subtilis*.

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

Beamline	BESSY-MX BL14.2
Wavelength (Å)	0.91841
Temperature (K)	100
No. of crystals	1
Detector	MX-225 CCD
Crystal-to-detector distance (mm)	280
Rotation range per image (°)	0.5
Total rotation range (°)	120
Resolution (Å)	45.0–2.5 (2.6–2.5)
Space group	$C2$
Unit-cell parameters (Å, °)	$a = 207.98$, $b = 135.90$, $c = 118.58$, $\beta = 94.72$
Mean $I/\sigma(I)$	18.8 (2.9)
R_{int}†	7.6 (32.9)
R_{meas}‡	8.3 (42.0)
R_{mrgd-F}§	9.4 (49.6)
No. of observed reflections	590411
No. of unique reflections	111150
Redundancy	5.3
Completeness (%)	97.9 (97.2)
Wilson B value (Å^2)	59.5
Estimated No. of molecules in ASU	Between 8 and 15
Solvent content (%)	Between 67.3 and 34.6

† $R_{int} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity measurement for reflection hkl and $\langle I(hkl)\rangle$ is the mean intensity of symmetry-related reflections and replicants. ‡ R_{meas} is defined as $\sum_h [n_h/(n_h-1)]^{1/2} \sum_i^{n_h} |\langle I_h\rangle - I_{h,i}| / \sum_h \sum_i^{n_h} I_{h,i}$ (Diederichs & Karplus, 1997). § R_{mrgd-F} is defined as $\sum |A_{I_{h,P}} - A_{I_{h,Q}}| / 0.5 \sum (A_{I_{h,P}} + A_{I_{h,Q}})$ (Diederichs & Karplus, 1997).

monoclinic space group $C2$, with unit-cell parameters $a = 208.0$, $b = 135.9$, $c = 118.6$ Å, $\beta = 94.7°$. Based on the total molecular weight of 27 790 Da and the unit-cell parameters, calculation of the Matthews parameter (Matthews, 1968) suggested between eight and 15 molecules per asymmetric unit, corresponding to a solvent content of between 67.3% ($V_M = 3.67$ Å^3 Da^{-1}) and 34.6% ($V_M = 1.88$ Å^3 Da^{-1}). These values lie within the range of solvent contents usually observed for protein crystals (Matthews, 1968).

To further confine the number of molecules in the asymmetric unit, a self-rotation function was calculated. The self-rotation function for $\kappa = 180°$ displays a number of twofold axes oriented in a single plane and one additional twofold axis oriented at a right angle to this plane (coordinates $\varphi = 30°$, $\psi = 90°$; Fig. 3). The latter also reoccurs at the same position on the sections $\kappa = 40$, 70 and 110°, suggesting that this peak represents a higher symmetry axis such as, for example, a sixfold or even a 12-fold symmetry axis. This symmetry is however far from perfect since a perfect 12-fold symmetry axis would generate peaks of identical heights at $\kappa = 30$, 60, 120 and 180°. Nevertheless, the self rotation could be interpreted as originating from a pseudo D_6 or D_{12} point-group symmetry arrangement of the molecules in the asymmetric unit. Taking into account the crystallographic dyad, such an arrangement could be explained by the presence of six or 12 molecules in the asymmetric unit.

The self-rotation function does not reveal whether the molecules really adopt point-group symmetry, since identical peaks could also originate from screw translational symmetry axes (Sawaya, 2007). For completeness, we also calculated a native Patterson (data not shown). We did not find any hints of pure translational NCS. A definitive answer to the actual symmetry arrangement will have to await structure solution. The structure of the recently solved prokaryotic transcriptional regulator YydK from *B. subtilis* (PDB code 3bwg; K. Tan, M. Zhou, J. Abdullah & A. Joachimiak, unpublished work) shows 27% sequence identity to YvoA. Structure solution will therefore be attempted by molecular replacement. If this fails, the phase problem will instead be solved by multiwavelength anomalous dispersion using selenomethionine-substituted protein.
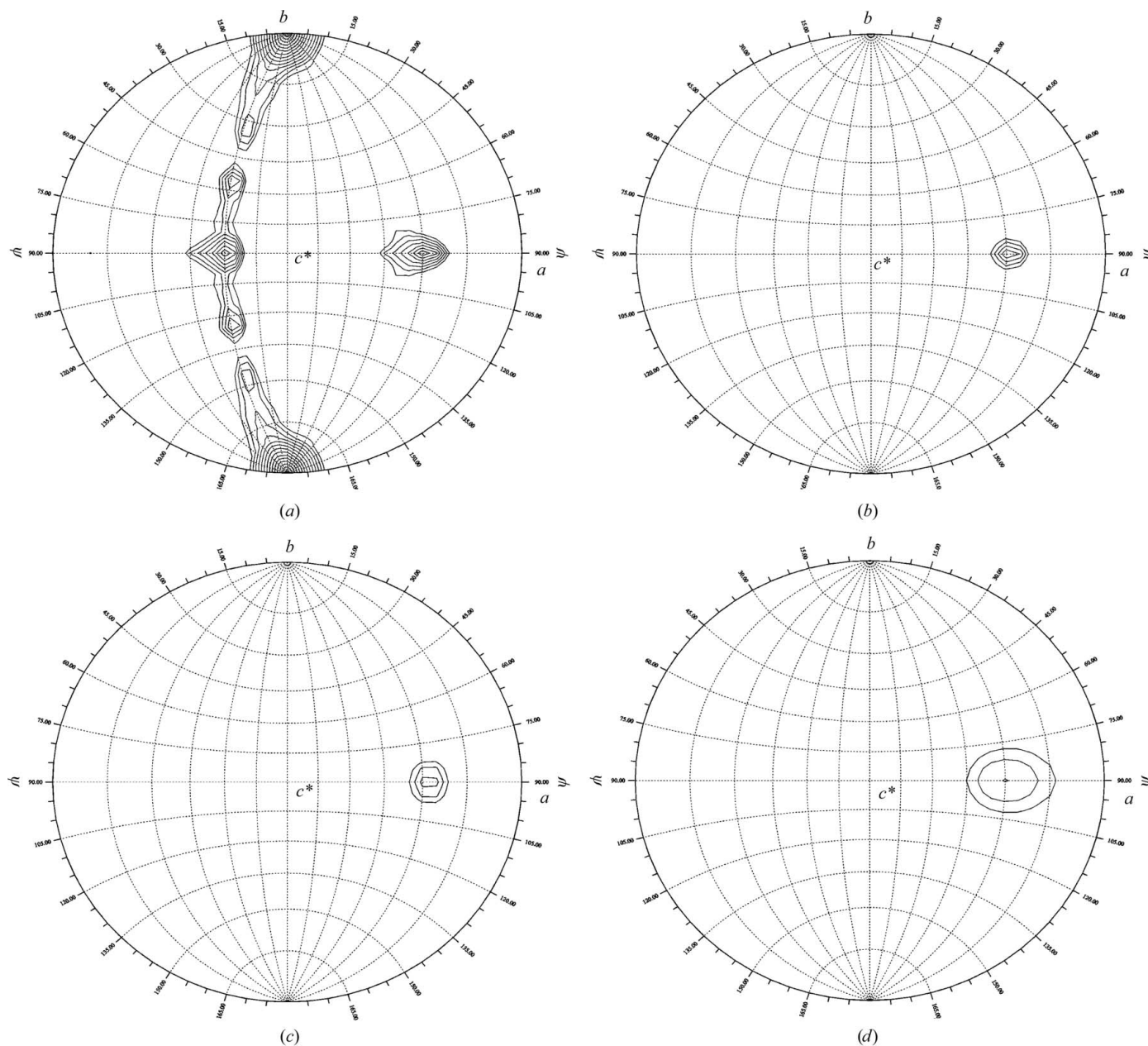
Figure 3
Self-rotation function based on data collected from a monoclinic crystal of YvoA. Shown are the major peak maxima observed for $\kappa = 180°$ (a), $\kappa = 110°$ (b), $\kappa = 70°$ (c) and $\kappa = 40°$ (d). The c^* axis is oriented perpendicular to the plane of the page.

References

Aalten, D. M. van, DiRusso, C. C., Knudsen, J. & Wierenga, R. K. (2000). *EMBO J.* **19**, 5167–5177.

Alice, A. F., Perez-Martinez, G. & Sanchez-Rivas, C. (2003). *Microbiology*, **149**, 1687–1698.

Aravind, L. & Anantharaman, V. (2003). *FEMS Microbiol. Lett.* **222**, 17–23.

Brueckner, R. & Titgemeyer, F. (2002). *FEMS Microbiol. Lett.* **209**, 141–148.

Colson, S., Stephan, J., Hertrich, T., Saito, A., van Wezel, G. P., Titgemeyer, F. & Rigali, S. (2007). *J. Mol. Microbiol. Biotechnol.* **12**, 60–66.

Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.

Dong, A. *et al.* (2007). *Nature Methods*, **4**, 1019–1021.

Gorelik, M., Lunin, V. V., Skarina, T. & Savchenko, A. (2006). *Protein Sci.* **15**, 1506–1511.

Haydon, D. J. & Guest, J. R. (1991). *FEMS Microbiol. Lett.* **63**, 291–295.

Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.

Kunst, F. *et al.* (1997). *Nature (London)*, **390**, 249–256.

Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.

Lee, M. H., Scherer, M., Rigali, S. & Golden, J. W. (2003). *J. Bacteriol.* **185**, 4315–4325.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Mobley, H. L., Doyle, R. J., Streips, U. N. & Langemeier, S. O. (1982). *J. Bacteriol.* **150**, 8–15.

Mueller-Dieckmann, J. (2006). *Acta Cryst.* D**62**, 1446–1452.

Reizer, J., Bachem, S., Reizer, A., Arnaud, M., Saier, M. H. Jr & Stuelke, J. (1999). *Microbiology*, **145**, 3419–3429.

Rezacova, P., Krejcirikova, V., Borek, D., Moy, S. F., Joachimiak, A. & Otwinowski, Z. (2007). *Proteins*, **69**, 679–682.

Rigali, S., Derouaux, A., Giannotta, F. & Dusart, J. (2002). *J. Biol. Chem.* **277**, 12507–12515.

Rigali, S., Nothaft, H., Noens, E. E., Schlicht, M., Colson, S., Muller, M., Joris, B., Koerten, H. K., Hopwood, D. A., Titgemeyer, F. & van Wezel, G. P. (2006). *Mol. Microbiol.* **61**, 1237–1251.

Rigali, S., Schlicht, M., Hoskisson, P., Nothaft, H., Merzbacher, M., Joris, B. & Titgemeyer, F. (2004). *Nucleic Acids Res.* **32**, 3418–3426.

Rigali, S., Titgemeyer, F., Barends, S., Mulder, S., Thomae, A. W., Hopwood, D. A. & van Wezel, G. P. (2008). *EMBO Rep.* **9**, 670–675.

Sawaya, M. R. (2007). *Methods Mol. Biol.* **364**, 95–120.

Stuelke, J. & Hillen, W. (2000). *Annu. Rev. Microbiol.* **54**, 849–880.

Tong, L. & Rossmann, M. G. (1997). *Methods Enzymol.* **276**, 594–611.

Vincent, F., Davies, G. J. & Brannigan, J. A. (2005). *J. Biol. Chem.* **280**, 19649–19655.

Vincent, F., Yates, D., Garman, E., Davies, G. J. & Brannigan, J. A. (2004). *J. Biol. Chem.* **279**, 2809–2816.

You, C., Lu, H., Sekowska, A., Fang, G., Wang, Y., Gilles, A. M. & Danchin, A. (2005). *BMC Microbiol.* **5**, 57.