Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Determination of the GH3.12 protein conformation through HPLC-integrated SAXS measurements combined with X-ray crystallography

Adam Round,[a,b]* Elizabeth Brown,[c] Romain Marcellin,[c] Ulrike Kapp,[c] Corey S. Westfall,[d] Joseph M. Jez[d] and Chloe Zubieta[c]*

[a]European Molecular Biology Laboratory, Grenoble Outstation, 6 Rue Jules Horowitz, 38042 Grenoble, France, [b]Unit for Virus Host-Cell Interactions, Université Grenoble Alpes–EMBL–CNRS, 6 Rue Jules Horowitz, 38042 Grenoble, France, [c]Structural Biology Group, European Synchrotron Radiation Facility, 6 Rue Jules Horowitz, 38000 Grenoble, France, and [d]Department of Biology, Washington University, St Louis, MO 63130, USA

Correspondence e-mail: around@embl.fr, zubieta@esrf.fr

Received 6 June 2013
Accepted 11 July 2013

PDB Reference: GH3.12–AMPCPP–salicylate complex, 4l39

The combination of protein crystallography and small-angle X-ray scattering (SAXS) provides a powerful method to investigate changes in protein conformation. These complementary structural techniques were used to probe the solution structure of the apo and the ligand-bound forms of the *Arabidopsis thaliana* acyl acid–amido synthetase GH3.12. This enzyme is part of the extensive GH3 family and plays a critical role in the regulation of plant hormones through the formation of amino-acid-conjugated hormone products *via* an ATP-dependent reaction mechanism. The enzyme adopts two distinct C-terminal domain orientations with 'open' and 'closed' active sites. Previous studies suggested that ATP only binds in the open orientation. Here, the X-ray crystal structure of GH3.12 is presented in the closed conformation in complex with the nonhydrolysable ATP analogue AMPCPP and the substrate salicylate. Using on-line HPLC purification combined with SAXS measurements, the most likely apo and ATP-bound protein conformations in solution were determined. These studies demonstrate that the C-terminal domain is flexible in the apo form and favours the closed conformation upon ATP binding. In addition, these data illustrate the efficacy of on-line HPLC purification integrated into the SAXS sample-handling environment to reliably monitor small changes in protein conformation through the collection of aggregate-free and highly redundant data.

1. Introduction

Small-angle X-ray scattering (SAXS) is rapidly becoming an important method to probe proteins and protein complexes in solution, thus helping to elucidate their function (Jacques & Trewhella, 2010; Lipfert *et al.*, 2007; Grant *et al.*, 2011; Hura *et al.*, 2009; Doniach & Lipfert, 2009; Mertens & Svergun, 2010). By examining proteins in solution, flexibility and thermodynamic ensemble properties are captured, in contrast to other structural techniques such as protein crystallography or electron microscopy, which are generally performed under cryogenic conditions (Rambo & Tainer, 2010). While atomic resolution data is often necessary for a detailed understanding of protein function and mechanism, high-resolution methods such as NMR have size limitations for the types of proteins and complexes that can be studied, EM is restricted to large complexes and generally a single highly stable conformation, and X-ray crystallography is limited by sampling a single generally low-energy conformation, providing a static snapshot of a specific protein conformation. By examining the solution state of either isolated proteins or macromolecular complexes and having few size limitations, SAXS experiments provide important complementary data to atomic-level studies. The protein can be studied under conditions which

more closely mimic the physiological environment, and data relating to the assembly state in solution, flexibility of domains and overall conformation can be derived from the relatively low-resolution SAXS measurements, data which are often not accessible by NMR, EM or X-ray crystallography alone (Schneidman-Duhovny *et al.*, 2012; Sibille & Bernadó, 2012).

Conformational changes, in particular domain rotations, are one method that enzymes have of altering the configuration of the active site, allowing multiple chemical reactions and/or binding events to take place using distinct sets of amino acids from the same protein. SAXS experiments provide a practical way to monitor and model domain movements in solution (Pelikan *et al.*, 2009; Bernadó, 2010; Hammel, 2012; Yang *et al.*, 2010). To fully exploit the utility of SAXS measurements, particularly for visualizing relatively small changes in conformation, high sample quality is of critical importance. Protein aggregation often occurs at high protein concentrations (Fink, 1998), and even small amounts of aggregate will impact SAXS data and fitting results. Aggregation is particularly noticeable at low resolution, where inter-particle effects are most apparent, and will greatly impact the determination of the radius of gyration (R_g) and the $I(0)$ value of the macromolecule. One method to address this that is becoming more common is the use of on-line sample-purification systems directly upstream of the SAXS sample environment (David & Pérez, 2009; http://www.saxier.org). By measuring the protein or complex immediately after chromatographic purification (*i.e. via* high-resolution size-exclusion chromatography), sample homogeneity is improved and can be verified through two independent techniques (UV absorbance and X-ray scattering measurements). We have recently commissioned an on-line HPLC purification system which is integrated into the sample-handling environment on BM29, the dedicated bioSAXS beamline at the European Synchrotron Radiation Facility (ESRF). Combined with automatic peak selection and data processing, this experimental setup has allowed us to

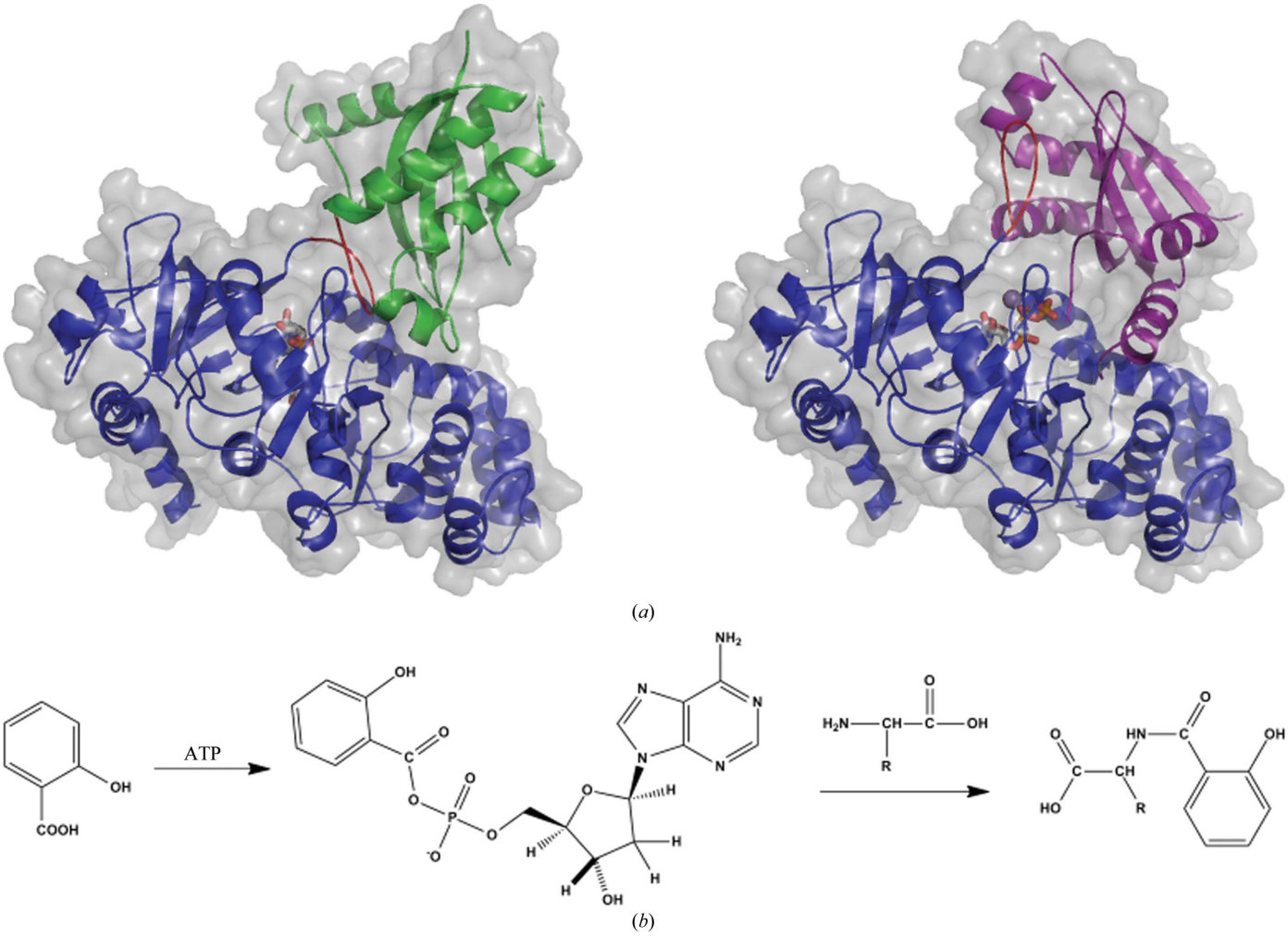

Figure 1
Structures of GH3.12 (*a*) and the two-step reaction catalyzed by the protein (*b*). (*a*) Left, GH3.12 in the active-site closed conformation. The protein is in complex with AMPCPP and salicylate (PDB entry 4l39) with the N-terminal domain coloured blue, the C-terminal domain green and the hinge loop red. Right, GH3.12 in complex with AMPCPP in the active-site open conformation (PDB entry 4ewv). The N-terminal domain is coloured blue, the C-terminal domain purple and the hinge loop red. The orientation is the same for both structures and is based on the N-terminal domain. (*b*) The reaction catalyzed by GH3.12 with salicylic acid as the carboxylate substrate. The AMP–salicylate conjugate is formed in the first reaction. This activated intermediate is then conjugated to an amino acid in the second reaction.

Table 1
Crystallographic statistics.

Values in parentheses are for the highest resolution shell.

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 62.36, b = 114.1, c = 158.0
Data collection	
Beamline	ID23-2, ESRF
Wavelength (Å)	0.873
Resolution (Å)	58.0–2.81 (3.00–2.81)
Reflections (total/unique)	201836/28172
Completeness (%)	99.7 (99.5)
$\langle I/\sigma(I)\rangle$	8.0 (2.7)
$R_{\rm merge}$ (%)	26.4 (70.5)
Model and refinement	
$R_{\rm cryst}/R_{\rm free}$ (%)	19.5/26.6
No. of protein atoms	8412
No. of waters	231
No. of ligand atoms	83
R.m.s. deviation, bond lengths (Å)	0.0169
R.m.s. deviation, bond angles (°)	1.578
Average B factor (Å^2)	
Protein	43.5
Water	20.1
Ligand	32.9
Stereochemistry, residues in (%)	
Most favoured regions	93.41
Allowed regions	5.94
Disallowed regions	0.65

investigate small changes in protein conformation which were not previously possible.

In order to test the efficacy of HPLC purification for downstream SAXS measurements, we selected an enzyme which (i) displayed multiple conformations that were crystallographically characterized, (ii) exhibited differences in calculated $R_{\rm g}$ values owing to different domain conformations and (iii) had domain conformations determined by ligand binding. GH3.12, a recently characterized acyl acid–amido synthetase, putatively undergoes domain rotation during catalysis, with the domain rotation attributed to ligand binding (Westfall *et al.*, 2012). GH3.12 is a member of the GH3 enzyme family that is widespread in plants and critical for the modulation of plant hormone activity by forming amino-acid conjugates of the carboxylic acid hormone substrate (Okrent *et al.*, 2009). These enzymes are part of the ANL superfamily, which comprises acyl-CoA synthetases, NRPS adenylation domains and luciferase enzymes (Gulick, 2009). The GH3 family shares the typical ANL tertiary structure consisting of a large N-terminal domain and a smaller C-terminal domain. The large N-terminal domain comprises a β-barrel and two β-sheets flanked by α-helices, and the C-terminal domain consists of a single four-stranded β-sheet bracketed by two α-helices on each side, with the active site located at the domain interface. A flexible hinge loop connects the two domains and putatively pivots the C-terminal domain during catalysis, altering the topology of the active site. This domain rotation is postulated to be important for the two-step reaction catalyzed by the GH3 proteins.

In the first reaction, the GH3 enzyme binds the carboxylate substrate and a molecule of ATP, forming an activated acyl-adenylate intermediate. The enzyme then selectively couples an amino acid to the acyl-adenylate, forming a stable hormone–amino acid conjugate (Fig. 1; Westfall *et al.*, 2010; Chen *et al.*, 2010; Jagadeeswaran *et al.*, 2007; Kumar *et al.*, 2012; Nobuta *et al.*, 2007; Okrent & Wildermuth, 2011). A C-terminal domain rotation is thought to be necessary to direct each reaction and to present different active-site amino acids for catalysis. Based on the available structural information for three members of this family, ATP binding was postulated to determine the C-terminal domain orientation (Westfall *et al.*, 2012; Peat *et al.*, 2012). The rotations of the C-terminal domain based on available X-ray crystal structures give rise to small differences in the overall size and shape of the macromolecule, suggesting the feasibility of SAXS experiments for monitoring the protein conformation in the presence of various substrates and products.

Here, we report the X-ray crystal structure of GH3.12 in complex with AMPCPP and salicylate and the fit of two different crystallographically determined GH3.12 structures to SAXS data obtained from on-line HPLC-purified protein in the apo, AMP-bound and ATP-bound forms. The quality of fit was used to determine the most likely protein conformation in solution with different ligands, providing insight into the catalytic mechanism. These data provide a probable catalytic mechanism for the two-step reaction performed by the GH3 enzymes.

2. Materials and methods

2.1. Protein expression and purification

GH3.12 was cloned into pET28a vector with an N-terminal thrombin-cleavable hexahistidine tag as described previously (Westfall *et al.*, 2012) and expressed in *Escherichia coli* cells (Rosetta 2, Novagen). The cells were grown at 310 K in LB broth to an optical density (A_{600}) of 0.8, after which the temperature was lowered to 293 K and 0.5 m*M* IPTG was added. The cells were grown overnight and then pelleted at 6000*g* for 15 min. The cells were resuspended in lysis buffer consisting of 25 m*M* Tris pH 7.5, 100 m*M* NaCl, 2 m*M* $MgCl_2$, 20 m*M* imidazole, 1× protease inhibitors (Roche, EDTA-free), 10 m*M* β-mercaptoethanol (BME) and lysed *via* sonication. Cell debris was pelleted at 40 000*g* and the supernatant was applied onto a 2 ml Ni–NTA (Qiagen) gravity column. The immobilized protein was washed with 10 column volumes (CV) of lysis buffer, 10 CV of lysis buffer plus 500 m*M* NaCl and 10 CV of lysis buffer and was then eluted with 10 ml lysis buffer plus 200 m*M* imidazole. The purified protein was dialyzed against dialysis buffer (25 m*M* Tris pH 7.5, 100 m*M* NaCl, 2 m*M* $MgCl_2$, 10 m*M* BME) and incubated overnight with thrombin protease to remove the hexahistidine tag. After depletion of thrombin and uncleaved protein over benzamidine Sepharose and Ni–NTA resin, respectively, purified GH3.12 was concentrated to approximately 10 mg ml^{-1} and applied onto a Superdex 200 (S200 GL 10/300, GE Healthcare) size-exclusion column for use in crystallization experiments or applied directly to the on-line HPLC/size-exclusion column for bioSAXS measurements using the same buffer as used in dialysis, with the exception of the 10 m*M* BME being

replaced by 1 m*M* tris(2-carboxyethyl)phosphine (TCEP) as a reducing agent and the addition of 1 m*M* AMP or ATP and 1 m*M* salicylic acid.

2.2. Crystallization and data collection

Salicylic acid (5 m*M*) and α,β-methyleneadenosine 5′-triphosphate (AMPCPP; 1 m*M*) were added to purified GH3.12 and the protein–ligand complex was concentrated to ∼10–15 mg ml^{-1} prior to crystallization. Crystals of the GH3.12–AMPCPP–salicylate complex grew at 288 K over one week from a solution consisting of 20% PEG 3350, 0.25 *M* ammonium acetate, 0.1 *M* sodium acetate pH 4.5, 5 m*M* TCEP. Crystals were cryoprotected in a solution consisting of 30% PEG 3350, 0.25 *M* ammonium acetate, 0.1 *M* sodium acetate pH 4.5 prior to cryocooling in liquid nitrogen. All data collection was performed at 100 K on beamline ID23-2 of the ESRF using the *EDNA* pipeline (Incardona *et al.*, 2009). Data were indexed and scaled using *XDS*/*XSCALE* (Kabsch, 2010*a*,*b*) and the structure was solved by molecular replacement (PDB entry 4eql; Westfall *et al.*, 2012) with the program *Phaser* (McCoy *et al.*, 2007). The model was built and refined using *Coot* (Emsley *et al.*, 2010) and *PHENIX* (Adams *et al.*, 2010), respectively (Table 1).

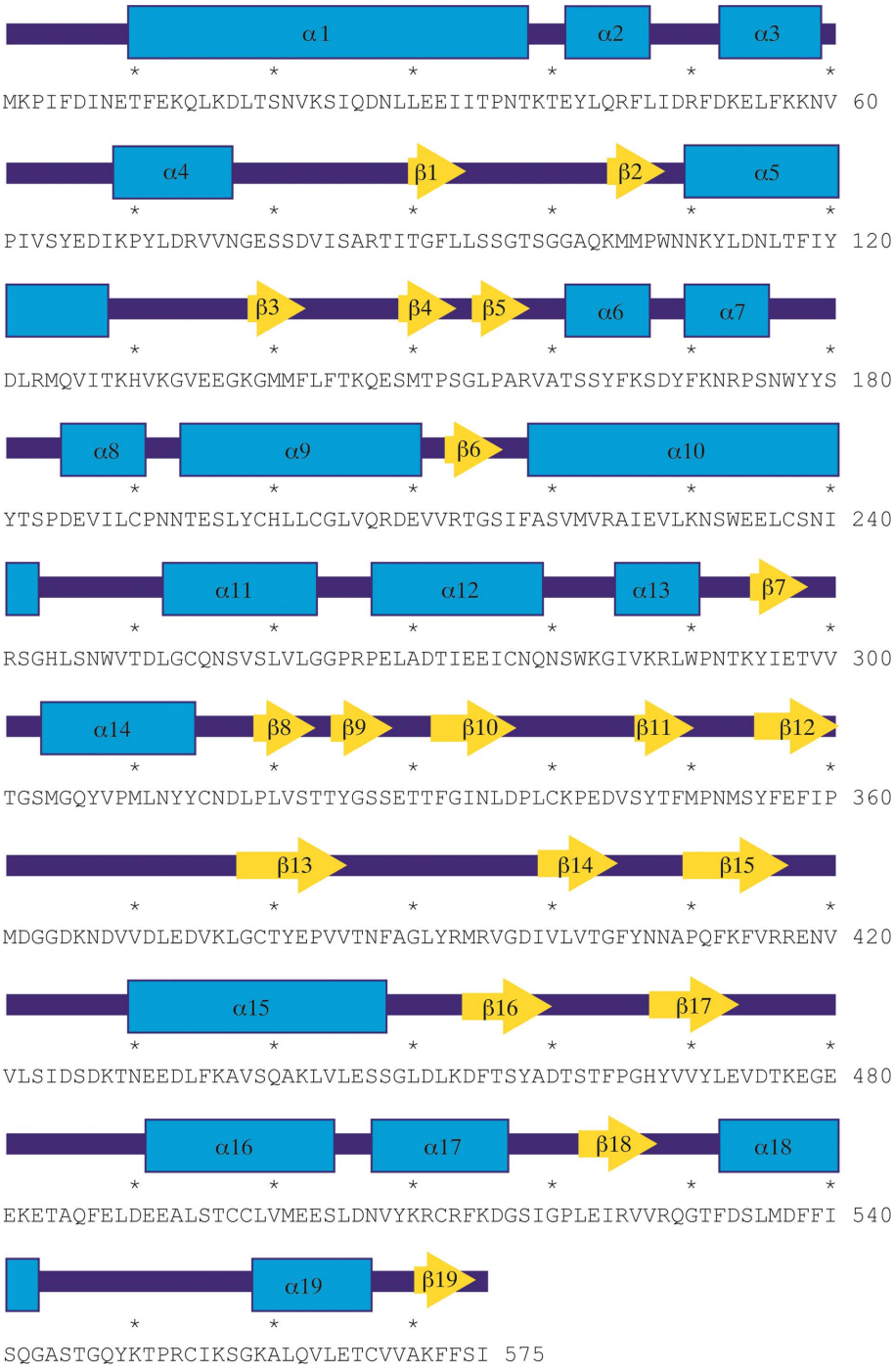

Figure 2
Sequence of GH3.12 with secondary-structure elements labelled. Helices are coloured light blue, coils dark blue and β-strands yellow. Every tenth amino acid is marked by a star.

2.3. BioSAXS measurements and experimental setup

An on-line HPLC system (Viscotek GPCmax, Malvern Instruments) was attached directly to the sample-inlet valve of the BM29 sample changer. Protein samples were loaded into vials and automatically injected onto the column (Superdex 200 10/300 GL, GE Healthcare) *via* an integrated syringe system. Buffers were degassed on-line and a flow rate of 0.5 ml min^{-1} at room temperature was used for all sample runs. The buffers used were as described above. Prior to each run, the column was equilibrated with 1.5 column volumes of buffer and the baseline was monitored. All data from the run were collected at a wavelength of 0.9919 Å using a sample-to-detector (PILATUS 1M, DECTRIS) distance of 2.81 m corresponding to an s range of 0.08–4.5 nm^{-1}. Approximately 3600 frames (1 frame s^{-1}) were collected per 60 min sample run. Initial data processing was performed automatically using the *EDNA* pipeline (Incardona *et al.*, 2009), generating radially integrated, calibrated and normalized one-dimensional profiles for each frame. All frames were compared with the initial frame and matching frames were merged to create the reference buffer. Any subsequent frames which differed from the reference buffer were subtracted and then processed within the *EDNA* pipeline using tools from the EMBL-HH *ATSAS*

2.5.1 suite. The invariants calculated by the *ATSAS* tool (*AUTORG*) were used to select a subset of frames from the peak scattering intensity. Frames with a consistent R_g from the peak scattering intensity were automatically merged to yield a single averaged frame corresponding to the scattering of an individual SEC purified species. The peaks of interest were reprocessed manually to maximize the signal-to-noise ratio. Specifically, the 31 frames corresponding to the highest protein concentration [based on UV absorbance and $I(0)$ values] were merged and used for all further data processing and model fitting. Porod–Debye plots (Supplementary Fig. S1[1]) were used to determine the flexibility of the protein according to Rambo & Tainer (2011).

2.4. Model fitting

Structures corresponding to the open form (PDB entry 4ewv; Northeast Structural Genomics Consortium, unpublished work) and closed form (this study; PDB entry 4l39) were used to calculate theoretical scattering curves. These curves were compared with the experimental data using *CRYSOL* (Svergun *et al.*, 1995). χ values for apo (3.45 and 2.96), ATP-bound (3.72 and 2.65) and AMP-bound (3.45 and 2.22) protein were calculated using *CRYSOL* for the open and closed forms, respectively. D_{max} values were calculated using *GNOM*, which is part of the *ATSAS* 2.5.1 suite.

3. Results

Recombinant GH3.12 was purified as a monomeric soluble protein. During purification, substrate or product (AMP, ATP or AMPCPP and salicylate) was added to the dialysis buffer at 1 m*M* to help to ensure that the protein was in a single conformation during all downstream crystallization and HPLC experiments. As described previously, GH3.12 adopts a two-domain structure typical of the ANL superfamily with a large N-terminal domain (residues 1–419) and a small C-terminal domain (residues 433–575), with the active site located at the domain interface. A flexible hinge loop (Val420–Glu432) between β15 and α15 connects the two domains and is postulated to pivot the C-terminal domain during the multi-step catalytic reaction (Westfall *et al.*, 2012; Figs. 1 and 2).

We crystallized GH3.12 in the presence of AMPCPP and salicylate to supplement the crystallographic data from previous studies in which structures of GH3.12 with AMP, with AMP and salicylate and with AMPCPP alone were determined (Westfall *et al.*, 2012). Based on the results of previous work and the current study, the protein was observed in two different conformations (open and closed, respectively) in complexes with AMPCPP (Figs. 1 and 3). It had been postulated that AMPCPP/ATP binding was only consistent with the open conformation owing to potential steric clashes between the protein and the β- and γ-phosphates of ATP in the closed conformation. Pyrophosphate release was hypothesized to trigger domain rotation from open to closed by removing these steric clashes (Westfall *et al.*, 2012). However, the structure of GH3.12 with AMPCPP and salicylate demonstrates an alternate positioning of the β- and γ-phosphates which allows complexation of the AMPCPP substrate in the closed conformation (Fig. 3).

Based on sequence alignments and structural studies, all GH3 enzymes have a conserved adenosine triphosphate/monophosphate (ATP/AMP) binding site defined by three sequence motifs (Westfall *et al.*, 2012; Peat *et al.*, 2012; Staswick *et al.*, 2005; Gulick, 2009). The first motif, a canonical P-loop (phosphate-binding loop), comprises residues 95–104 (S_{95}SGTSGGAQK$_{104}$), with the Ser95 and Ser96 hydroxyls putatively involved in contacts with the α-phosphate. However, the P-loop was disordered in all GH3.12 structures, with no interactions observed with the nucleotide. A second motif clearly visible in the electron density and involved in phosphate binding and positioning (S_{322}TTYGSSE$_{329}$) forms a β–turn–β structure, with Ser328 hydrogen-bonding the

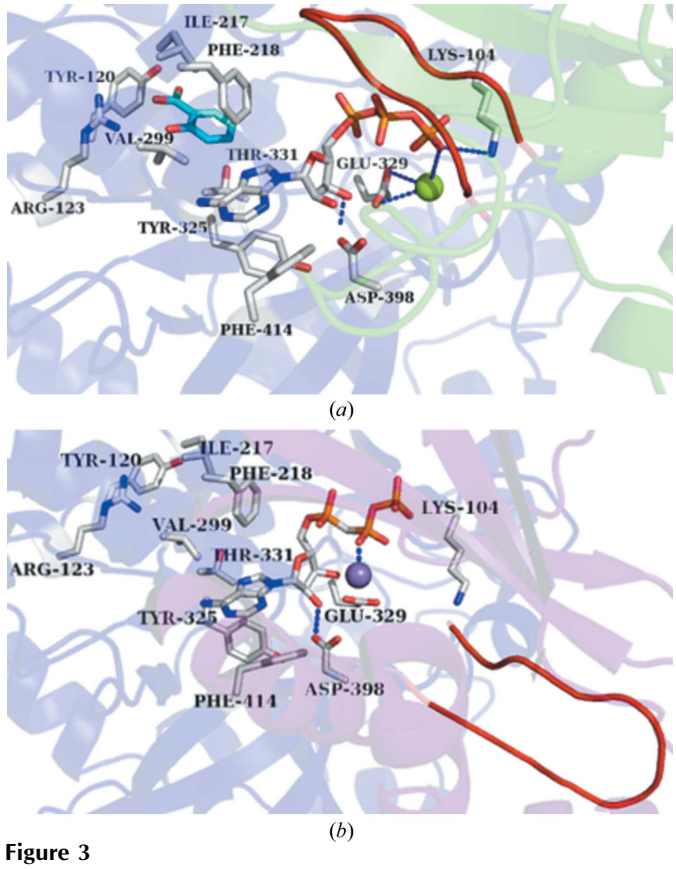

Figure 3
Close-up views of the active site of GH3.12 in complex with AMPCPP. (*a*) Active-site closed conformation with AMPCPP depicted as sticks and coloured by atom with C atoms in grey, salicylate C atoms in cyan and side-chain C atoms in grey. Residues directly contacting AMPCPP are shown. Lys104 is involved in coordinating a magnesium ion (green) as well as the γ-phosphate group. Glu329 is within coordinating distance of the Mg^{2+} ion, as shown. (*b*) Active-site open conformation with AMPCPP coloured by atom and residues involved in AMPCPP binding depicted. Lys104 is disordered and is too far away to contact the AMPCPP phosphates. An Mn^{2+} ion was observed in the active site within coordinating distance of the β-phosphate of AMPCPP. Lys550 forms an electrostatic interaction with the β-phosphate of AMPCPP.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: DW5060). Services for accessing this material are described at the back of the journal.

α-phosphate group and Tyr325 stacking with the adenine ring. Previous studies of ANL proteins and mutagenesis studies of GH3-family proteins indicate that the glutamate residue (Glu329) in this motif is critical for Mg^{2+} binding and orientation of ATP for catalysis. This can be seen in the AMPCPP/salicylate structure of the current study, in which a magnesium ion was observed in the electron density within coordination distance of the γ-phosphate and the Glu329 side chain. In contrast, the structure with AMPCPP alone (PDB entry 4ewv) has an Mn^{2+} ion co-coordinating the β-phosphate with no interaction with Glu329 and an additional interaction with Lys550. The third motif (G_{391}LYLYRGD$_{398}$) contains a conserved aspartate (Asp398) that binds the nucleotide ribose hydroxyls, further orienting the nucleotide in the active site (Fig. 3).

To better understand the domain movements of GH3.12 in the solution state, SAXS experiments were performed using different combinations of substrates and products to complement the crystallographic studies. The R_g values calculated for the open and closed conformations differ by 0.35 nm, a relatively small but significant difference measureable through SAXS experiments using the HPLC system. Theoretical scattering curves and R_g values were generated based on the open- and closed-conformation crystal structures. Data from the measured SAXS data were fitted to each model using *CRYSOL* (Fig. 4). The apo structure had a slightly larger R_g (3.04 $\pm$ 0.04 nm) and a poorer fit to the open and closed conformation scattering curves, which is clearly indicative of domain flexibility. Indeed, no apo crystal structures of the protein are available, which is likely to be owing to the inherent flexibility of the two domains. Only in the presence of substrate or product is the domain orientation 'locked' into place, allowing crystallization to occur. The protein in solution in the presence of AMP or ATP gave smaller R_g values (2.75 $\pm$ 0.05 and 2.75 $\pm$ 0.05 nm, respectively) which, owing to the highly redundant data, are well above the noise level, implying that these small differences are real. The SAXS AMP data most closely resembled the closed conformation, as was predicted. The ATP structure fitted less well to the AMPCPP structure in the open conformation and fitted better to the AMPCPP structure in the closed conformation, which is virtually identical to the AMP crystal structure, suggesting that the enzyme quickly adopts a closed conformation in solution when bound to nucleotide.

4. Discussion

Probing the conformational changes of a protein during catalysis necessitates new approaches and the use of complementary techniques. High-resolution structural data are critical for detailed mechanistic studies, but they are often insufficient to fully understand enzymatic function, particularly during multi-step reactions which require remodelling of the active site, as not all relevant conformations are crystallizable. The combination of SAXS and X-ray crystallography has been used to probe gross changes in protein oligomerization state and large domain movements during protein DNA-binding events, RNA binding and catalysis (Pretto *et al.*, 2010; Rambo & Tainer, 2010; Putnam *et al.*, 2007; Hammel, 2012). More subtle domain movements have been, and remain, challenging to probe through SAXS experiments; however, the use of on-line purification directly upstream of the X-ray sample environment facilitates these types of studies (Jensen *et al.*, 2010).

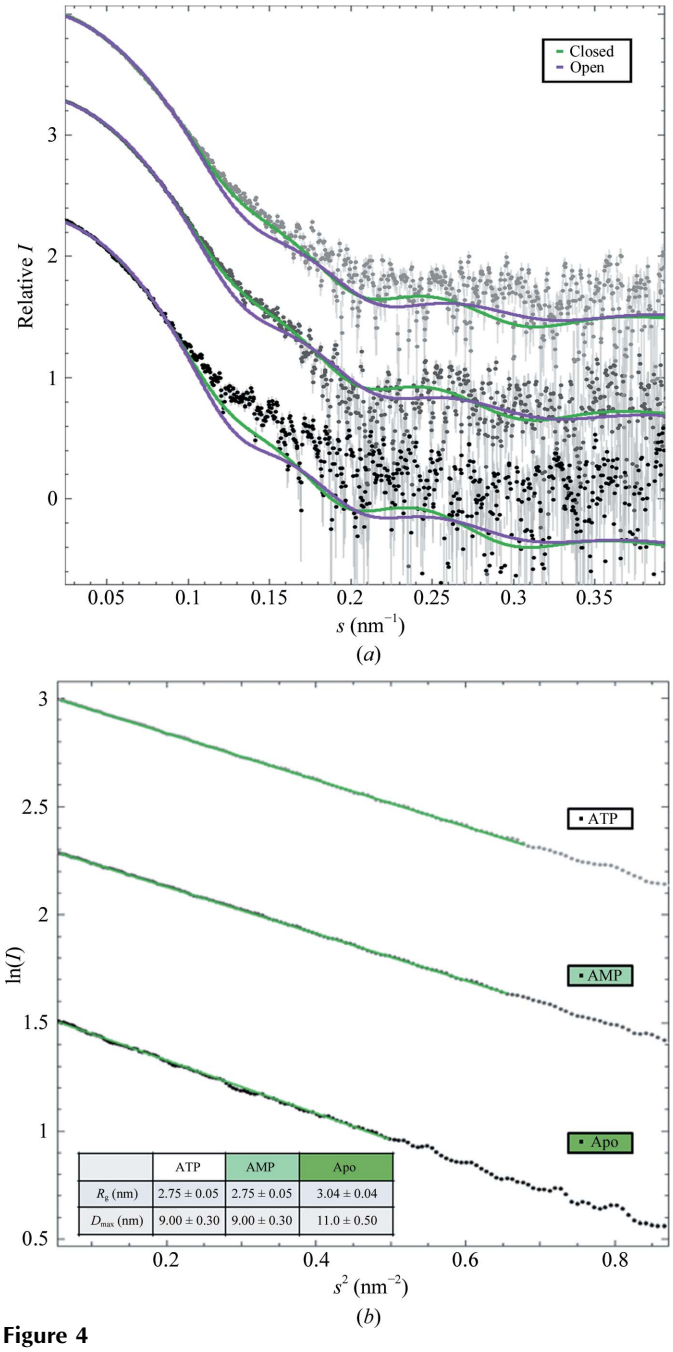

Figure 4
SAXS data for GH3.12 in the apo, AMP-bound and ATP-bound forms. (*a*) Scattering curves for the apo (bottom curves), AMP-bound (middle curves) and ATP-bound (top curves) forms of GH3.12. Theoretical scattering curves for the closed (green) and open (purple) forms were overlaid on the experimental data using *CRYSOL*. (*b*) Guinier plots of the apo (bottom), AMP-bound (middle) and ATP-bound (top) forms of GH3.12. R_g and D_{max} values are shown. R_g values were calculated from the linear Guinier region.

Based on structural data for GH3.12 and related enzymes which show two different C-terminal domain conformations owing to a pivot about the hinge loop, hypotheses correlating C-terminal domain position with nucleotide binding and catalysis have been made (Peat *et al.*, 2012; Westfall *et al.*, 2012). In the open conformation (PDB entry 4ewv), AMPCPP is bound in the active site with the hinge loop directed away from the ligand; the α18/α19 loop is positioned towards the N-terminal domain to cap the nucleotide β- and γ-phosphates, and Lys550 and an Mn^{2+} ion interact with the β-phosphate, helping to position the nucleotide (Fig. 3). This conformation was postulated to be the dominant form for the adenylation reaction, allowing facile access to the active site for substrate binding as well as providing a solvent-accessible channel for pyrophosphate release upon formation of the acyl-adenylate. In contrast, in AMP-bound structures (PDB entries 4eql and 4epm) and that with AMPCPP/salicylate (this study), the hinge and α18/α19 loops swap positions, giving rise to a less compact conformation (the closed conformation). In the AMPCPP/salicylate structure a magnesium ion is coordinated by Lys104, Glu329 and the γ-phosphate of the AMPCPP ligand, resulting in an alternate conformation of the β- and γ-phosphates and the loss of the electrostatic interaction with Lys550. This conformation would engender steric clashes that are incompatible with the open conformation. Indeed, Lys104 does not interact with the nucleotide or the Mn^{2+} ion in the open conformation, as the Lys104 side chain points away from the ATP-binding cavity and is disordered in the structure. Thus, the two crystal structures of GH3.12 with AMPCPP show distinct C-terminal domain conformations with similar substrates crystallized under different conditions. Based on X-ray crystallographic studies alone, the C-terminal domain conformation which is likely to exist in solution under physiological conditions is unclear and it is not known which conformation is catalytically active. In order to understand the catalytic mechanism, SAXS solution-state studies were used to determine the protein conformation in solution. Small-angle scattering data from protein in complex with nucleotide substrate (ATP) and product (AMP) and in the apo form were measured and compared with known crystal structures in order to determine protein conformations.

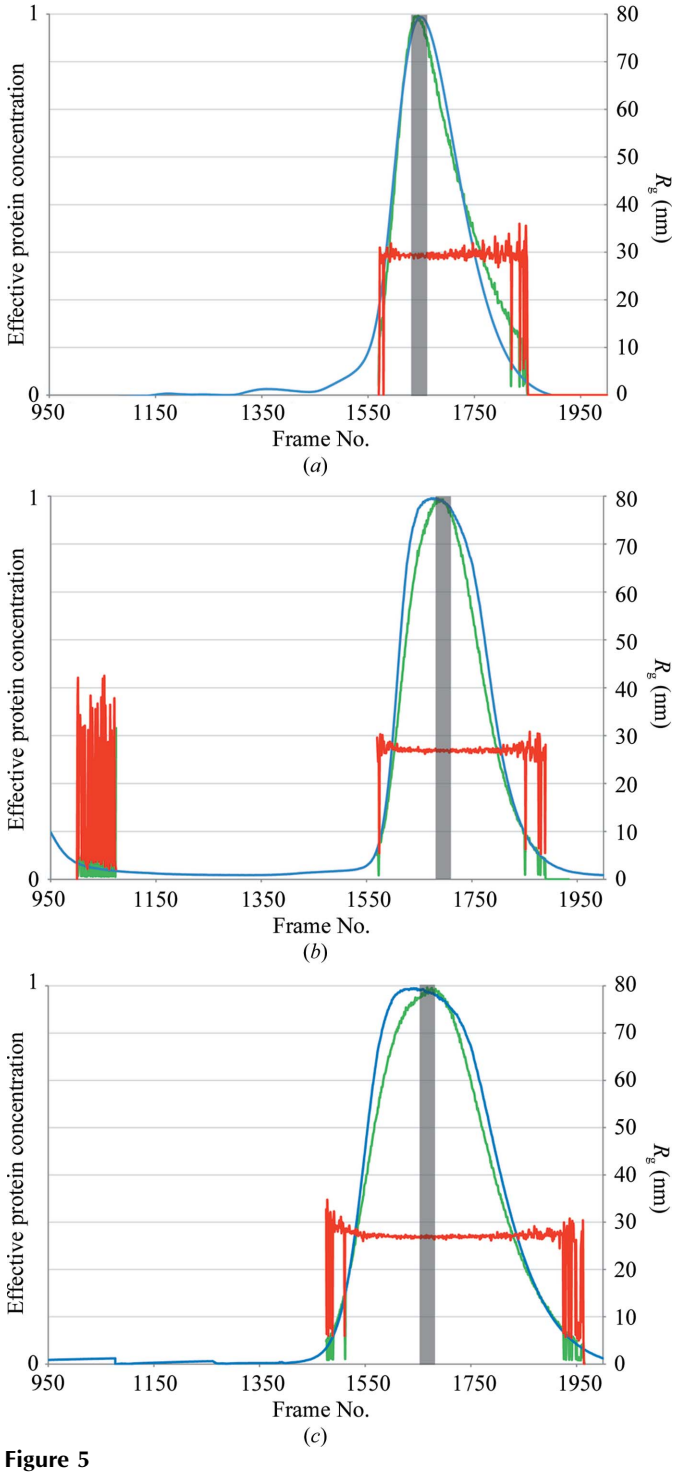

Figure 5
UV absorbance curves, $I(0)$ and R_g values for GH3.12 in ATP-bound (*a*), AMP-bound (*b*) and apo (*c*) forms. UV absorbance traces are coloured blue, $I(0)$ curves are coloured green and R_g values are coloured red. Data frames used to calculate one-dimensional scattering curves are indicated by a grey box.

The open and closed conformations of GH3.12 differ in their R_g values and their calculated scattering curves. However, these differences can be difficult to detect in a SAXS experiment owing to small amounts of aggregated protein which can mask the desired data (Fig. 5). In order to overcome this obstacle, an HPLC purification system was integrated into the SAXS sample-handling environment. This allowed the collection of highly redundant data easily and routinely for each sample, dramatically improving the sample homogeneity. During each purification run, all of the material flowing through the HPLC column (both protein and buffer) was exposed to the X-ray beam. The data included the buffer values necessary for subtraction and all possible protein species including aggregates, oligomers and complexes. The number of useful data frames corresponding to a single species was improved and each species was easily identified through the combination of UV absorbance data and automatic processing to obtain R_g values. The signal-to-noise ratio from this type of experiment, in which hundreds of useful frames per experiment are routinely collected, compares very favourably with the tens of frames often collected using a

standard robotic sample changer. Overlays of the HPLC UV trace, the $I(0)$ values and the calculated R_g show good correlation between the selected frames at the maximum peak height based on both UV absorbance and $I(0)$ for the apo and the ligand-bound forms of GH3.12. The selected frames used for the generation of one-dimensional curves have stable R_g values, further indicating the presence of one discrete species in solution. Small peaks corresponding to aggregated protein do not contribute to the scattering (Fig. 5). The scattering curves derived from 31 data frames corresponding to the highest protein concentration, free of aggregates and concentration-dependent interparticle scattering (as confirmed by the uniform R_g over the entire peak), were selected for all further calculations. The number of frames used can be varied empirically depending on the sharpness of the HPLC peak and the most stable R_g and $I(0)$ values in order to achieve the best signal-to-noise ratio. Increasing the number of data frames to greater than 100, for example, led to artificially small error bars. Noise from the lower concentration frames and noise from inherent variance of the beamline led to overall poorer quality scattering curves. In a standard experiment at most ten frames are collected per sample and used to generate scattering curves. Thus, a good compromise between signal to noise was achieved with between ten and 31 frames of data from each peak during an HPLC run. While HPLC purification directly upstream of the X-ray beam dilutes the sample and increases the noise in the measurement, the removal of aggregated species, the independent checks on sample quality [UV absorbance, $I(0)$ and R_g] and the collection of highly redundant data compensate for these drawbacks. By excluding oligomers/aggregates the data quality was dramatically improved, as the R_g values and scattering curves thus obtained were unbiased by the presence of multiple species. Fitting the measured scattering curves to curves calculated from the available crystal structures gave clear differences in fit, with the closed conformation best fitting the experimental data (Fig. 4). The protocol used is generally applicable to any species which are amenable to HPLC separation.

Using combined X-ray crystallographic and scattering data, we were able to study protein conformation in a highly challenging system. GH3.12 putatively undergoes domain rotations during ligand binding and catalysis, resulting in small changes in the calculated R_g and scattering curves. Comparison of the calculated scattering curves based on the crystallographic data with the measured scattering curves has allowed us to predict the most likely protein conformation in solution in the apo form and in the presence of substrate (ATP) and product (AMP). This was not possible using X-ray crystallographic data alone as two unique protein conformations were observed crystallographically in the presence of the same substrate analogue, AMPCPP. The use of these complementary structural techniques, high-resolution X-ray studies which provide a static snapshot of the protein and low-resolution SAXS studies which take advantage of the average conformation of the protein in solution under conditions which better mimic the physiological environment, provides critical insight into mechanism. Small differences in R_g, D_{max} and the Porod–Debye plot (Fig. 4 and Supplementary Fig. 1) corresponded to the predicted flexibility of the apo form and the more ordered 'locked-in' structure of the substrate-bound and product-bound forms. Simulations of scattering curves based on all available crystal structures of GH3.12 revealed the most dominant solution structure of the ATP-bound protein, which corresponded to the active-site closed AMPCPP/salicylate structure determined here as well as the previously determined AMP complexes.

The solution-state conformation of the protein with ATP and the crystal structure with AMPCPP and salicylate bound suggests that nucleotide binding does not lock the C-terminal domain into an open conformation. This conformation is likely to be important for initial nucleotide and hormone binding; however, upon formation of the ATP–hormone substrate–Mg^{2+} complex GH3.12 is likely to adopt a closed conformation and subsequently performs the adenylation reaction. Complex formation with ATP and magnesium orients the β- and γ-phosphate tail of the ATP molecule, forming electrostatic interactions between the γ-phosphate and Lys104 in the closed conformation. These interactions may be the triggers necessary for domain rotation about the hinge loop. Once the nucleotide is properly positioned in a putatively productive binding mode for catalysis, the protein is able to adopt a closed conformation. Based on the available structural data, hormone binding occurs in the open conformation with subsequent binding of the nucleotide and the adoption of the closed conformation. The hormone-binding pocket is capped by the nucleotide and the hinge loop in the closed conformation, thus precluding diffusion of hormone substrate into or out of the binding pocket. The hormone-binding pocket is distal to the C-terminal domain, with no contributions from residues located in the hinge loop to hormone binding. Indeed, based on the crystal structures available, hormone binding does not lead to any conformational changes in the active site. The conformational change from open to closed may be physiologically relevant and act as a mechanism to tune the reaction, as low hormone concentrations, but high nucleotide concentrations, will lead to catalytically unproductive adoption of the closed conformation. In addition, as the adenylated intermediate is highly reactive, a closed active-site conformation is likely to be important to avoid unwanted hydrolysis of the intermediate. It remains to be determined whether additional C-terminal domain rotations are necessary for the subsequent amino-acid conjugation reaction or for product release.

Great strides have recently been made in integrating SAXS studies and high-resolution structural techniques to simulate conformational changes in solution (see Putnam *et al.*, 2007; Hammel, 2012; Bernadó, 2010; Rambo & Tainer, 2013; Koch *et al.*, 2003; Jacques *et al.*, 2012; Hura *et al.*, 2009; Doniach & Lipfert, 2009; Lipfert *et al.*, 2009; Pons *et al.*, 2010; Pelikan *et al.*, 2009; Mertens & Svergun, 2010; David & Pérez, 2009, and references therein). In order to fully exploit SAXS experiments to probe solution-state conformations, particularly subtle protein conformational changes resulting from the presence or absence of substrate or product ligands, the

sample requires a high level of homogeneity that is often difficult to achieve. These data demonstrate that the integration of on-line HPLC systems within the beamline sample environment provides a critical tool to explore catalysis, bridging the gap between high-resolution structural information and low-resolution solution-state ensembles measured by SAXS experiments.

The authors would like to acknowledge Jerome Kieffer and Staffan Ohlsson for their work on the *EDNA* BioSAXS pipeline. Work in the Jez laboratory was supported by the National Science Foundation (MCB-0904215 to JMJ). CSW was supported by a US Department of Agriculture National Institute of Food and Agriculture pre-doctoral fellowship (MOW-2010-05240).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Bernadó, P. (2010). *Eur. Biophys. J.* **39**, 769–780.
Chen, Q., Westfall, C. S., Hicks, L. M., Wang, S. & Jez, J. M. (2010). *J. Biol. Chem.* **285**, 29780–29786.
David, G. & Pérez, J. (2009). *J. Appl. Cryst.* **42**, 892–900.
Doniach, S. & Lipfert, J. (2009). *Methods Enzymol.* **469**, 237–251.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Fink, A. L. (1998). *Fold. Des.* **3**, R9–R23.
Grant, T. D., Luft, J. R., Wolfley, J. R., Tsuruta, H., Martel, A., Montelione, G. T. & Snell, E. H. (2011). *Biopolymers*, **95**, 517–530.
Gulick, A. M. (2009). *ACS Chem. Biol.* **4**, 811–827.
Hammel, M. (2012). *Eur. Biophys. J.* **41**, 789–799.
Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L. II, Tsutakawa, S. E., Jenney, F. E. Jr, Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S. J., Scott, J. W., Dillard, B. D., Adams, M. W. & Tainer, J. A. (2009). *Nature Methods*, **6**, 606–612.
Incardona, M.-F., Bourenkov, G. P., Levik, K., Pieritz, R. A., Popov, A. N. & Svensson, O. (2009). *J. Synchrotron Rad.* **16**, 872–879.
Jacques, D. A., Guss, J. M. & Trewhella, J. (2012). *BMC Struct. Biol.* **12**, 9.
Jacques, D. A. & Trewhella, J. (2010). *Protein Sci.* **19**, 642–657.
Jagadeeswaran, G., Raina, S., Acharya, B. R., Maqbool, S. B., Mosher, S. L., Appel, H. M., Schultz, J. C., Klessig, D. F. & Raina, R. (2007). *Plant J.* **51**, 234–246.
Jensen, M. H., Toft, K. N., David, G., Havelund, S., Pérez, J. & Vestergaard, B. (2010). *J. Synchrotron Rad.* **17**, 769–773.
Kabsch, W. (2010*a*). *Acta Cryst.* D**66**, 125–132.
Kabsch, W. (2010*b*). *Acta Cryst.* D**66**, 133–144.
Koch, M. H. J., Vachette, P. & Svergun, D. I. (2003). *Q. Rev. Biophys.* **36**, 147–227.
Kumar, R., Agarwal, P., Tyagi, A. K. & Sharma, A. K. (2012). *Mol. Genet. Genomics*, **287**, 221–235.
Lipfert, J., Columbus, L., Chu, V. B., Lesley, S. A. & Doniach, S. (2007). *J. Phys. Chem. B*, **111**, 12427–12438.
Lipfert, J., Herschlag, D. & Doniach, S. (2009). *Methods Mol. Biol.* **540**, 141–159.
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.
Mertens, H. D. & Svergun, D. I. (2010). *J. Struct. Biol.* **172**, 128–141.
Nobuta, K., Okrent, R. A., Stoutemyer, M., Rodibaugh, N., Kempema, L., Wildermuth, M. C. & Innes, R. W. (2007). *Plant Physiol.* **144**, 1144–1156.
Okrent, R. A., Brooks, M. D. & Wildermuth, M. C. (2009). *J. Biol. Chem.* **284**, 9742–9754.
Okrent, R. A. & Wildermuth, M. C. (2011). *Plant Mol. Biol.* **76**, 489–505.
Peat, T. S., Böttcher, C., Newman, J., Lucent, D., Cowieson, N. & Davies, C. (2012). *Plant Cell*, **24**, 4525–4538.
Pelikan, M., Hura, G. & Hammel, M. (2009). *Gen. Physiol. Biophys.* **28**, 174–189.
Pons, C., D'Abramo, M., Svergun, D. I., Orozco, M., Bernadó, P. & Fernández-Recio, J. (2010). *J. Mol. Biol.* **403**, 217–230.
Pretto, D. I., Tsutakawa, S., Brosey, C. A., Castillo, A., Chagot, M. E., Smith, J. A., Tainer, J. A. & Chazin, W. J. (2010). *Biochemistry*, **49**, 2880–2889.
Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. (2007). *Q. Rev. Biophys.* **40**, 191–285.
Rambo, R. P. & Tainer, J. A. (2010). *Curr. Opin. Struct. Biol.* **20**, 128–137.
Rambo, R. P. & Tainer, J. A. (2011). *Biopolymers*, **95**, 559–571.
Rambo, R. P. & Tainer, J. A. (2013). *Nature (London)*, **496**, 477–481.
Schneidman-Duhovny, D., Kim, S. J. & Sali, A. (2012). *BMC Struct. Biol.* **12**, 17.
Sibille, N. & Bernadó, P. (2012). *Biochem. Soc. Trans.* **40**, 955–962.
Staswick, P. E., Serban, B., Rowe, M., Tiryaki, I., Maldonado, M. T., Maldonado, M. C. & Suza, W. (2005). *Plant Cell*, **17**, 616–627.
Svergun, D., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.
Westfall, C. S., Herrmann, J., Chen, Q., Wang, S. & Jez, J. M. (2010). *Plant Signal. Behav.* **5**, 1607–1612.
Westfall, C. S., Zubieta, C., Herrmann, J., Kapp, U., Nanao, M. H. & Jez, J. M. (2012). *Science*, **336**, 1708–1711.
Yang, S., Blachowicz, L., Makowski, L. & Roux, B. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 15757–15762.