Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The structure of TTHA0988 from *Thermus thermophilus*, a KipI–KipA homologue incorrectly annotated as an allophanate hydrolase

David A. Jacques,[a] David B. Langley,[a] Seiki Kuramitsu,[b,c] Shigeyuki Yokoyama,[d,e] Jill Trewhella[a] and J. Mitchell Guss[a]*

[a]School of Molecular Bioscience, University of Sydney, Sydney, NSW 2006, Australia, [b]Department of Biological Sciences, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan, [c]RIKEN SPring-8 Center, Harima Institute, Sayo-gun, Hyogo 679-5148, Japan, [d]Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Tokyo, Japan, and [e]RIKEN Systems and Structural Biology Center, Yokohama 230-0045, Japan

Correspondence e-mail: mitchell.guss@sydney.edu.au

Received 27 October 2010
Accepted 6 December 2010

PDB References: TTHA0988, $P6_522$, 3ore; $P2_12_12_1$, 3opf; $P4_32_12$, 3oep.

The *Thermus thermophilus* protein TTHA0988 is a protein of unknown function which represents a fusion of two proteins found almost ubiquitously across the bacterial kingdom. These two proteins perform a role regulating sporulation in *Bacillus subtilis*, where they are known as KipI and KipA. *kipI* and *kipA* genes are usually found immediately adjacent to each other and are often fused to produce a single polypeptide, as is the case with TTHA0988. Here, three crystal forms are reported of TTHA0988, the first structure to be solved from the family of 'KipI–KipA fusion' proteins. Comparison of the three forms reveals structural flexibility which can be described as a hinge motion between the 'KipI' and 'KipA' components. TTHA0988 is annotated in various databases as a putative allophanate hydrolase. However, no such activity could be identified and genetic analysis across species with known allophanate hydrolases indicates that a misannotation has occurred.

1. Introduction

The thermophilic bacterium *Thermus thermophilus* HB8 is the subject of the 'Structural Biological Whole Cell Project' as part of the Riken Structural Genomics/Proteomics Initiative (RSGI; Yokoyama *et al.*, 2000). The relatively small size of the *T. thermophilus* genome (2 Mbp) and the thermostability of its proteins make the organism a promising target for the complete understanding of fundamental biological processes at the molecular level. Despite its small size, analysis of the *T. thermophilus* genome predicts 2226 genes, of which 866 (39%) are hypothetical and as such have no known function. Attributing roles to the predicted proteins is a major obstacle to the complete molecular characterization of the organism.

In *Bacillus subtilis*, the onset of sporulation is governed by a phosphorelay (Burbulys *et al.*, 1991). In response to as yet unidentified signals, one of several histidine kinases, for instance KinA, utilizes ATP to autophosphorylate the side chain of a conserved histidine. The phosphoryl group is subsequently transferred *via* two phosphotransfer proteins (Spo0F and Spo0B) to the response regulator Spo0A. Once phosphorylated, Spo0A is able to alter the expression of more than 120 genes in order to bring about spore formation (Hilbert & Piggot, 2004; Molle *et al.*, 2003). The elaborate nature of the phosphorelay allows the input of multiple signals regulating the decision to commit to sporulation. One mechanism by which negative regulation of the phosphorelay may be

achieved is by KinA inhibition. Two proteins, Sda and KipI, are known to prevent sporulation by binding KinA and subsequently inhibiting the phosphorelay (Rowland *et al.*, 2004; Wang *et al.*, 1997). While Sda (46 amino acids) and KipI (240 amino acids) are structurally unrelated, they both bind the same region on KinA, namely the dimerization and histidine phosphotransfer (DHp) domain (Jacques *et al.*, 2008; Whitten *et al.*, 2007). The inhibition of KipI is itself subject to negative regulation by binding the KipA protein. The regulation of this interaction has yet to be elucidated.

kipI and *kipA* gene homologues are found almost ubiquitously across bacterial species as well as in certain fungi, yet their function remains largely unknown as their occurrence is not restricted to sporulating organisms. The genes are almost always located immediately adjacent to each other and are often fused to produce a single polypeptide. One such 'KipI–KipA fusion' is TTHA0988, which therefore became a priority target for structure solution as it provides a model for the KipI–KipA interaction.

As a collaboration between the RSGI and the University of Sydney, we were able to solve the phase problem using multiple isomorphous replacement (MIR) with the $P6_522$ crystal form and then solve the $P2_12_12_1$ form by molecular replacement. We subsequently identified a third crystal form, $P4_32_12$, which was also solved by molecular replacement. The $P4_32_12$ structure has been previously described in terms of its use as a homology model for the rigid-body modelling of *B. subtilis* KipI–KipA small-angle scattering data (Jacques *et al.*, 2011). In this paper, we report the three crystal forms of TTHA0988 and a comparison between the resulting structural models. We also discuss the possible roles in *T. thermophilus* and other bacterial species for this poorly understood yet near-ubiquitous protein fold.

2. Materials and methods

2.1. Protein expression and purification

The recombinant TTHA0988 which produced the $P4_32_12$ and $P6_522$ crystal forms was expressed and purified as described previously (Jacques *et al.*, 2010). The $P2_12_12_1$ sample was produced as part of the 'Structural Biological Whole Cell Project' (Yokoyama *et al.*, 2000) by overexpression in *Escherichia coli* Rosetta 2 cells from the same plasmid (PC010988-42). The cells were lysed by ultrasonication in 20 m*M* Tris pH 8.0, 50 m*M* NaCl and the lysate was heat-treated (343 K) followed by ultracentrifugation at 200 000*g* for 1 h. The soluble fraction was bound to a Resource PHE column in 50 m*M* sodium phosphate pH 7.0, 1.5 *M* ammonium sulfate. The fraction containing TTHA0988 eluted at 0.29 *M* ammonium sulfate and was subsequently passed through a CHT10-I hydroxyapatite column (10 m*M* sodium phosphate pH 7.0, 0.5 *M* NaCl) followed by precipitation by the addition of 1.8 *M* ammonium sulfate. The sample was finally purified by passage through a HiLoad 16/60 Superdex 200 size-exclusion column equilibrated in 20 m*M* Tris pH 8.0, 0.5 *M* NaCl.

2.2. Protein crystallization and X-ray data collection

Crystals belonging to space group $P6_522$ were prepared by hanging-drop vapour diffusion at 293 K in which 2 µl protein solution (3.1 mg ml^{-1}) was mixed with 2 µl precipitant [3.3 *M* NaCl, 0.04 *M* HEPES, 4.6%(*v*/*v*) 1- propanol pH 6.9]. Crystals (400–500 µm) appeared within 3–5 d. Heavy-atom derivatives were prepared by the addition of 0.4 µl heavy-atom solution (100 m*M*, prepared in precipitant) directly to the drop. The soaking times varied from 2 h to 5 d. The crystals were cryoprotected by dehydrating the drop to reach NaCl saturation (as identified by the appearance of NaCl crystals), at which point the protein crystal was removed and flash-cooled in a cold nitrogen stream (100 K). X-ray diffraction data were collected in-house using Cu $K\alpha$ X-rays produced by a Rigaku RU200H rotating-anode generator with Osmic optics and were recorded on a MAR345 image plate (MAR Research). Data were collected from a native crystal and six heavy-atom derivatives to a maximum resolution of 2.9 Å.

The heavy-atom compounds used were potassium tetrachloroplatinate, uranyl acetate (2 h and 3 d soaks), trimethyllead acetate, ethylmercury thiosalicylic acid and 5-amino-2,4,6-triiodoisophthalic acid (I3C).

A crystal belonging to space group $P2_12_12_1$ was obtained by sitting-drop vapour diffusion in which 70 nl protein solution (2.8 mg ml^{-1}) was mixed with an equal volume of crystallant [0.17 *M* ammonium sulfate, 0.085 *M* sodium cacodylate, 25.5% PEG 8000, 15%(*v*/*v*) glycerol pH 6.5]. X-ray diffraction data were collected to 1.95 Å resolution at a wavelength of 1.000 Å using an ADSC Quantum 210 detector on beamline BL44B2 at the SPring-8 synchrotron.

A crystal belonging to space group $P4_32_12$ was obtained by hanging-drop vapour diffusion in which 2 µl protein solution (3.1 mg ml^{-1}) was mixed with 2 µl precipitant solution [120 m*M* ammonium sulfate, 17.9% polyethylene glycol 8000, 10.5%(*v*/*v*) glycerol, 60 m*M* sodium cacodylate pH 6.5]. A single crystal appeared after seven months of incubation at 293 K. No additional cryoprotection was employed and the crystal was flash-cooled in a cold nitrogen stream (100 K). X-ray diffraction data were collected to 1.75 Å resolution at a wavelength of 0.95369 Å using an ADSC Quantum 315r detector on beamline MX2 at the Australian Synchrotron using the *Blu-Ice* control system (McPhillips *et al.*, 2002).

2.3. Data processing, structure solution and refinement

All data were indexed and scaled using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Structure solution of the $P6_522$ crystal form was achieved by MIR using the program *SOLVE* (Terwilliger & Berendzen, 1999; mean figure of merit = 0.57, score = 39.25) with density modification performed by *RESOLVE* (Terwilliger, 2000). The structure was initially built as separate chains by *Buccaneer* (Cowtan, 2006) and assembled manually in *Coot* (Emsley & Cowtan, 2004). A near-complete molecule of TTHA0988 was built and was used without further refinement as a search model for molecular replacement for the $P2_12_12_1$ data. *Phaser* (McCoy *et al.*, 2007) identified three molecules in the asymmetric unit

Table 1
Diffraction data and refinement statistics for the three crystal forms of TTHA0988 and the heavy-atom derivatives used for MIR.

Values in parentheses are for the highest resolution shell.

Derivative	Native	Native	Native	K_2PtCl_4	Uranyl acetate (2 h soak)	Trimethyllead acetate	Ethylmercury thiosalicylic acid	Uranyl acetate (3 d soak)	I3C
Space group	$P4_32_12$	$P2_12_12_1$	$P6_522$	$P6_522$	$P6_522$	$P6_522$	$P6_522$	$P6_522$	$P6_522$
Unit-cell parameters (Å)	$a = b = 75.3$, $c = 183.0$	$a = 74.505$, $b = 132.107$, $c = 177.833$	$a = b = 142.1$, $c = 259.1$	$a = b = 142.1$, $c = 259.1$	$a = b = 142.1$, $c = 259.1$	$a = b = 142.1$, $c = 259.1$	$a = b = 142.1$, $c = 259.1$	$a = b = 142.1$, $c = 259.1$	$a = b = 142.1$, $c = 259.1$
X-ray source	Australian Synchrotron MX2	SPring-8 BL44B2	Rotating anode	Rotating anode	Rotating anode	Rotating anode	Rotating anode	Rotating anode	Rotating anode
Wavelength (Å)	0.95369	1.000	1.5418	1.5418	1.5418	1.5418	1.5418	1.5418	1.5418
Detector	ADSC Quantum 315r	ADSC Quantum 210	MAR 345	MAR 345	MAR 345	MAR 345	MAR 345	MAR 345	MAR 345
Resolution range (Å)	47.4–1.75 (1.78–1.75)	50.0–1.95 (1.98–1.95)	50.0–2.90 (2.95–2.90)	50.0–3.30 (3.36–3.30)	50.0–3.60 (3.66–3.60)	50.0–3.0 (3.05–3.00)	50.0–3.0 (3.05–3.00)	50.0–3.0 (3.05–3.00)	50.0–7.50 (7.63–7.50)
Observed reflections	763929	498554	228916	162025	151450	202743	271707	236608	12999
Unique reflections	53995	111600	32167	24022	17452	29490	28983	26341	2140
Completeness (%)	100.0 (100.0)	87.3 (51.6)	91.8 (41.4)	99.1 (90.2)	92.8 (40.3)	92.8 (24.9)	90.6 (20.9)	82.6 (9.9)	95.2 (69.5)
Multiplicity	14.1 (14.2)	4.5 (4.2)	7.1 (3.3)	6.7 (4.5)	8.7 (2.2)	6.9 (1.9)	9.2 (1.9)	9.0 (1.7)	6.1 (2.5)
$\langle I/\sigma(I)\rangle$	11.3 (4.10)	11.3 (2.30)	21.1 (2.82)	12.8 (5.0)	20.7 (3.0)	17.6 (3.3)	33.2 (1.7)	22.1 (1.9)	7.4 (3.5)
$R_{\rm merge}$†	0.118 (0.817)	0.131 (0.60)	0.078 (0.296)	0.126 (0.241)	0.083 (0.251)	0.082 (0.178)	0.099 (0.284)	0.079 (0.256)	0.213 (0.261)
Heavy atoms found by *SOLVE*	—	—	—	9	5	9	1	5	4
Phasing power‡	—	—	—	0.95	0.66	1.04	0.21	0.77	1.06
Reflections, working set	51137	105342	28837						
Reflections, test set	2745	5548	1625						
Protomers per ASU	1	3	2						
Total atoms (non-H)	4163	12224	5454						
Protein atoms	3640	11026	5454						
Water atoms	500	1159	0						
Other atoms	23	39	0						
$R_{\rm cryst}$	0.163 (0.191)	0.197 (0.257)	0.285 (0.341)						
$R_{\rm free}$	0.192 (0.223)	0.233 (0.285)	0.316 (0.371)						
R.m.s.d. bond lengths (Å)	0.006	0.006	0.005						
R.m.s.d. bond angles (°)	1.16	1.11	0.94						
All-atom $\langle B\rangle$ (Å^2)	22	39	71						
Cruickshank's DPI§	0.1022	0.1914	0.6868						

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ‡ Phasing power $= \sum |F_{\rm H}| / \sum \big||F_{\rm PHobs}| - |F_{\rm PHcalc}|\big|$. § Diffraction precision indicator (Cruickshank, 1999) as output from *REFMAC*5.

(log-likelihood gain = 2529). The resulting model was first rigid-body refined, which was followed by multiple rounds of restrained refinement using *REFMAC*5 (Murshudov *et al.*, 1997). Between rounds of refinement, the model was manually checked and corrected against the corresponding σ_A-weighted electron-density maps in *Coot*. Solvent molecules were added as the refinement progressed either manually or automatically within *Coot* and were routinely checked for correct stereochemistry, for sufficient supporting density above a $2F_o - F_c\,\sigma$ threshold of 1.0σ and for a reasonable thermal factor. The quality of the model was regularly checked for steric clashes, incorrect stereochemistry and rotamer outliers using *MolProbity* (Chen *et al.*, 2010). Chain *A* of the fully refined structure was used as a search model for molecular replacement for the $P4_32_12$ data. *Phaser* identified one molecule in the asymmetric unit (log-likelihood gain = 3500) and the structure was refined as described for the $P2_12_12_1$ structure. Chain *A* of the $P2_12_12_1$ structure was also used as a search model to re-solve the $P6_522$ data (log-likelihood gain = 2876). *Phaser* identified two molecules in the asymmetric unit; however, the electron density associated with residues 1–220 (representing domains A and B) of chain *B* was too poor to build accurately. Refinement proceeded as described for the other two crystal forms, with no attempt to model the missing residues. After refinement, the density still had not improved sufficiently to be able to model these two domains, which have been omitted from the final structure.

2.4. DNA-binding assay

The vector pGEM3ZF(+) (Promega) was digested with *Hae*III to give fragments of 11, 18, 80, 102, 142, 174, 267, 289, 314, 323, 434, 458 and 587 bp. The digest was incubated with increasing molar ratios of TTHA0988:DNA in 50 or 500 m*M* NaCl, 10 m*M* $MgCl_2$, 10 m*M* Tris pH 7.9, 1 m*M* dithiothreitol and 10%(*v*/*v*) glycerol for 30 min at room temperature. The procedure was also applied to the individual 80, 102 and 142 bp fragments, which were first gel purified. The incubated samples were electrophoretically separated on 3.5% Metaphor agarose gels (Cambrex Bio Science), set and run in 0.5×

Tris–borate–EDTA buffer, after which the gel was stained with ethidium bromide and photographed atop a UV transilluminator.

3. Results and discussion

3.1. The structure of TTHA0988 and comparison of the three crystal forms

The data-processing and refinement statistics are listed in Table 1. TTHA0988 is a monomeric four-domain protein (Fig. 1). The facts that no structural homologues were available and that there is only a single ordered methionine in 494 residues explain why, despite the collection of diffraction data to 1.95 Å resolution, the RSGI structure resisted solution by molecular replacement and selenomethionine MAD phasing.

The structure reveals that both KipI (equivalent to domains A and B) and KipA (equivalent to domains C and D) of *B. subtilis* are two-domain proteins (Jacques *et al.*, 2010). Domain A (residues 1–75) belongs to the arginine-repressor family and consists of a three-stranded β-sheet and two α-helices. Domain C (211–386), the largest domain, belongs to the cupin family and comprises a 12-stranded β-sandwich and a single α-helix. Domains B (76–210) and D (387–494) both belong to the cyclophilin family and each consist of a seven-stranded β-barrel core and two α-helices. Cyclophilins are proline *cis*/*trans* isomerases and their active sites are located on a concave surface formed from four β-strands (Ikura & Ito, 2007). The structurally homologous surfaces on domains B and D are positioned facing each other so that they form a large cleft that runs along the length of the protein. This surface on domain B carries a cluster of highly conserved aromatic residues, while the corresponding surface on domain D is largely aliphatic. The two cyclophilin domains are linked by a salt bridge between Arg137 and Glu434. Domains A and B are connected by a flexible linker (residues 69–76) as is evident from the absence of electron density in all but chain *B* of the $P2_12_12_1$ crystal form. Domains B and C are connected by a proline-rich linker (residues 202–210) which forms a polyproline II helix. The linker between domains C and D (residues 372–389) also has a high proline content and residues 382–387 form another polyproline II helix. Domain B also possesses a rare nonproline *cis*-peptide between Ala127 and Glu128.

Table 2
Comparison between TTHA0988 structures.

Pairwise r.m.s.d. values (Å) are reported based on full C^{α} alignments. Note: chain *B* of the $P6_522$ crystal form is missing residues 1–219.

Crystal form (chain)	$P4_32_12$ (*A*)	$P6_522$ (*A*)	$P6_522$ (*B*)	$P2_12_12_1$ (*A*)	$P2_12_12_1$ (*B*)
$P6_522$ (*A*)	0.79				
$P6_522$ (*B*)	0.59	0.62			
$P2_12_12_1$ (*A*)	0.64	0.74	0.48		
$P2_12_12_1$ (*B*)	0.88	0.88	0.52	0.40	
$P2_12_12_1$ (*C*)	0.50	0.74	0.48	0.59	0.79

In the $P6_522$ crystal form, chain *A* is well resolved, while the electron density for domains A and B of chain *B* is too poor to be able to accurately model these domains. Analysis of the symmetry-related molecules reveals a very large solvent channel formed along the length of the sixfold axis (Fig. 2). A superposition of the complete chain *A* on chain *B* (Fig. 2, red) predicts that domains A and B project into this large solvent channel and are unavailable for participation in crystal contacts. Conversely, chain *A* and domains C and D of chain *B* participate in several crystal contacts. The disordering of two whole domains owing to the absence of crystal contacts suggests that domains A and B (the KipI-equivalent component) are flexible with respect to domains C and D (the KipA-equivalent component).

In an attempt to investigate the range of domain motions in the molecule, r.m.s.d. values were calculated for pairwise alignments of each of the six structures available from the three crystal forms (Table 2). The largest r.m.s.d. value (0.88 Å) was recorded from the alignment of the $P4_32_12$ structure with chain *B* of the $P2_12_12_1$ crystal form. A second alignment was performed between these two structures based solely on residues 220–493 (domains C and D; r.m.s.d. of 0.43 Å). This alignment (Fig. 3*a*) reveals that domains A and B can pivot relative to domains C and D by 5.3°, corresponding to a shift in C^{α} positions of up to 4.0 Å. An alignment of the five complete molecules by domains C and D revealed that the range of positions sampled by domains A and B lie on the same trajectory (Fig. 3*b*), consistent

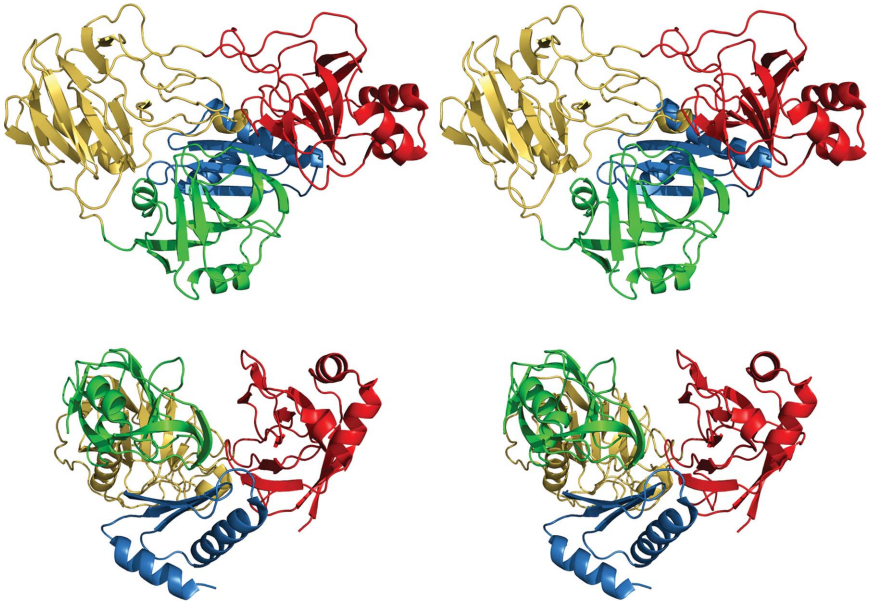

Figure 1
Stereoview of the structure of TTHA0988 coloured by domains. Domain A is shown in blue, domain B in red, domain C in yellow and domain D in green. A second orientation is shown to illustrate the cleft formed between domains B and D.

with a discreet 'hinge' motion. This range of domain movement can be used to rationalize the disordering of domains A and B observed in the $P6_522$ crystal form when the protein is free from conformationally restrictive crystal contacts.

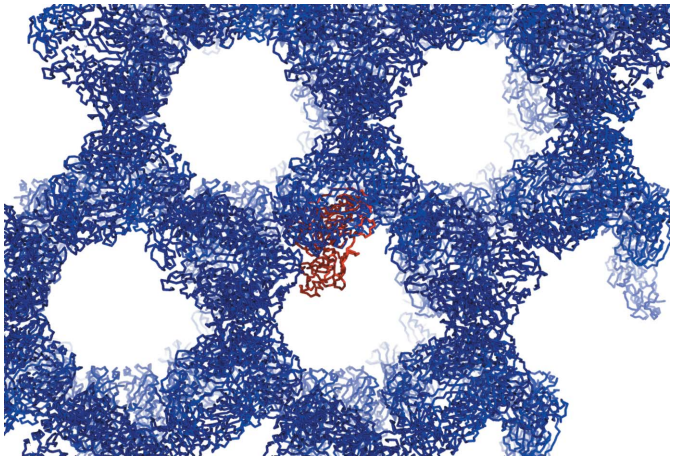

Figure 2
Crystal packing of TTHA0988 in space group $P6_522$ viewed along the sixfold axis. A molecule of chain *A* (red) is superposed on the two modelled domains of chain *B*. The superposition predicts that the disordered domains project into the large solvent channels, where they are unable to form stabilizing crystal contacts.

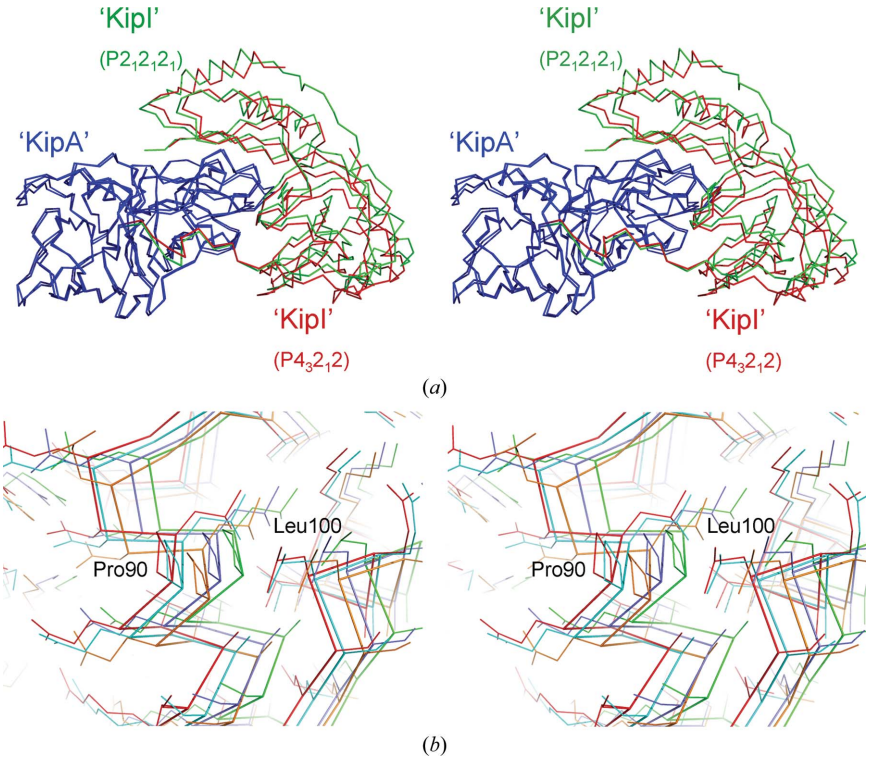

Figure 3
(*a*) Stereoview of the alignment between the $P4_32_12$ structure and chain *B* of the $P2_12_12_1$ structure based on the KipA-equivalent domains (residues 220–494, blue). The KipI-equivalent domains (residues 1–219) are coloured red and green for the $P4_32_12$ and $P2_12_12_1$ structures, respectively. The KipI-equivalent domains pivot by as much as 5.3°, with C^{α} displacements of up to 4.0 Å. An identical alignment of all five structures reveals a range of 'KipI' positions that lie on the same trajectory. (*b*) A stereoview of these alignments focused on the region around Pro90 and Leu100 illustrates that the observed domain positions are consistent with a hinge motion (red, $P4_32_12$; cyan, $P2_12_12_1$ chain *C*; orange, $P6_522$ chain *A*; purple, $P2_12_12_1$ chain *A*; green, $P2_12_12_1$ chain *B*).

3.2. TTHA0988 function: the allophanate hydrolase question

The folds represented by TTHA0988 (and KipI–KipA) are annotated as comprising subunits 1 and 2 of putative allophanate hydrolases; however, attempts to demonstrate such activity in TTHA0988 were unsuccessful (data not shown). Allophanate hydrolase is an enzyme that catalyzes the degradation of allophanate (carboxylated urea) to ammonia and carbon dioxide. This function was originally attributed to these domains based on the characterization of a multifunctional enzyme, urea amidolyase, from *Saccharomyces cerevisiae* (Whitney & Cooper, 1972). Sequence analysis reveals that this large enzyme has five domains: an N-terminal amidase domain, followed by a carboxylase domain, a KipA-like domain, a KipI-like domain and a small C-terminal biotin ligase domain (Fig. 4). Biochemical studies identified this protein as possessing two enzymatic activities: an ATP- and biotin-dependent urea carboxylase (which produces allophanate from urea and bicarbonate) and allophanate hydrolase. Recombinatorial mapping attributed loss of allophanate hydrolase activity to the N-terminal half of the protein and loss of urea carboxylase activity to the C-terminal half (Cooper *et al.*, 1980). The original data, therefore, showed the KipI-like and KipA-like domains to be part of the urea carboxylase, while the allophanate hydrolase activity could be mapped to the amidase domain. While most KipI and KipA homologues in prokaryotes are annotated as having allophanate hydrolase function, to our knowledge the only proven prokaryotic allophanate hydrolases have been found in *Pseudomonas* sp. strain ADP (Shapir *et al.*, 2005) and *Oleomonas sagarensis* (Kanamori *et al.*, 2005). Neither of the enzymes in question is homologous to TTHA0988; however, both belong to the amidase superfamily. This argument is further supported by the adjacent gene in *O. sagarensis*, which is a proven urea carboxylase (Kanamori *et al.*, 2004) and has the same domain architecture as the *S. cerevisiae* urea amidolyase, except for the amidase domain (Fig. 4). With amidase domains being responsible for allophanate hydrolase function in the yeast and prokaryotic examples, it is clear that a misannotation has occurred. There is no evidence that any TTHA0988 homologues are responsible for allophanate hydrolase activity; however, they have been found, performing an unknown role, within urea carboxylases.

An unusual feature of the TTHA0988 protein is its very high theoretical pI of 9.6,

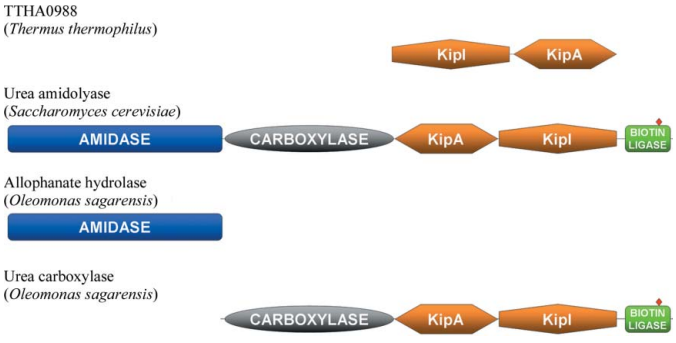

Figure 4
Domain analysis of TTHA0988 and known allophanate hydrolases. TTHA0988 comprises a KipI-like domain and a KipA-like domain (orange). Urea amidolyase from *S. cerevisiae* is a five-domain multifunctional protein that possesses both biotin- and ATP-dependent urea carboxylase activity as well as allophanate hydrolase activity. This second activity was originally attributed to the KipI and KipA domains. The first proven bacterial allophanate hydrolase was found in *O. sagarensis* and only possesses a single amidase domain (blue). The gene immediately adjacent to this enzyme was found to be a urea carboxylase, suggesting that the original attribution of allophanate hydrolase activity to the KipI–KipA domains was incorrect. The presence of a carboxylase domain (grey) and a biotin ligase domain (green) with a biotinylation site (red diamond) would seem to be sufficient for urea carboxylase activity. It is therefore unclear what role the KipI and KipA domains play in this enzyme.

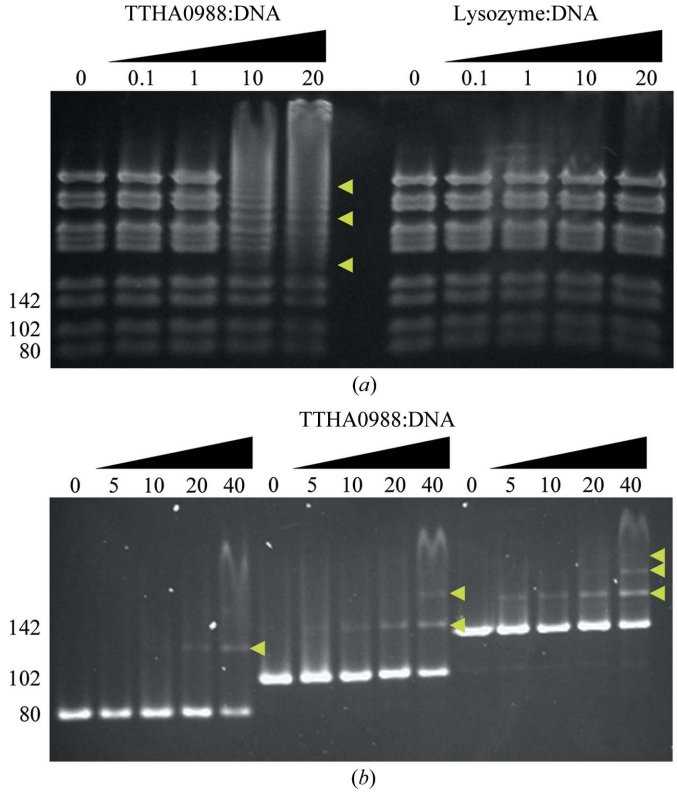

Figure 5
TTHA0988 binds DNA in a band-shift assay. (*a*) The vector pGEM3ZF(+) was digested with *Hae*III and incubated with various molar ratios of TTHA0988 (the ratios relative to the number of vector fragments pre-digestion are indicated above the lanes). Incubations with lysozyme, which has a similar pI but is not a DNA-binding protein, were performed as a negative control. (*b*) The bottom three bands from the digest displayed in (*a*) (sizes 80, 102 and 142 bp) were gel purified and the band-shift experiment was repeated. Yellow arrows indicate DNA species retarded by TTHA0988. Results were unaffected by incubation with NaCl (to 500 m*M*) prior to electrophoresis.

raising the possibility that the protein might have a role that involves it interacting with nucleic acids. The protein carries no obvious DNA-binding motifs; however, the cleft formed by the cyclophilin domains has approximately the dimensions of a double-stranded DNA helix. When tested in a band-shift assay, TTHA0988 was indeed capable of retarding the migration of restriction fragments generated from a random vector on an agarose gel (Fig. 5). The interactions appear to be nonspecific, perhaps reflecting a simple ion-exchange phenomenon, although the addition of 500 m*M* NaCl to the assay did not affect binding (data not shown). A similar experiment using lysozyme, which has a similar pI (9.3) but does not bind DNA, showed no shift in the DNA bands. In contrast to TTHA0988, the pI of the *B. subtilis* KipI–KipA complex is calculated to be 5.2, supporting the notion that these proteins, which are involved in histidine kinase regulation, are unlikely to have a supporting role that involves being bound to DNA.

4. Conclusions

TTHA0988 was prioritized for study as it provides a model for the interaction between two near-ubiquitous bacterial proteins, KipI and KipA. Here, we have reported three crystal forms of TTHA0988, comparisons of which reveal that the KipI- and KipA-equivalent components are flexible relative to each other. While genetic and biochemical evidence precludes the annotated allophanate hydrolase activity, the roles of TTHA0988 and its homologues across the bacterial kingdom remain largely unknown, but they may be involved in mediating macromolecular interactions, enzyme catalysis or chaperone activity.

This research was supported in part by Australian Research Council Discovery Project Grant DP0984536 awarded to JT and JMG and the RIKEN Structural Genomics/Proteomics Initiative (RSGI), the National Project on Protein Structural and Functional Analyses, Ministry of Education, Culture, Sports, Science and Technology of Japan. DAJ is supported by an Australian Institute of Nuclear Science and Engineering Postgraduate Research Award. Data collection for this research was undertaken on the MX2 beamline at the Australian Synchrotron, Victoria, Australia and on the BL44B2 beamline at the SPring-8 synchrotron, Hyogo, Japan

References

Burbulys, D., Trach, K. A. & Hoch, J. A. (1991). *Cell*, **64**, 545–552.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Cooper, T. G., Lam, C. & Turoscy, V. (1980). *Genetics*, **94**, 555–580.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.

Cruickshank, D. W. J. (1999). *Acta Cryst.* D**55**, 583–601.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Hilbert, D. W. & Piggot, P. J. (2004). *Microbiol. Mol. Biol. Rev.* **68**, 234–262.
Ikura, T. & Ito, N. (2007). *Protein Sci.* **16**, 2618–2625.
Jacques, D. A., Langley, D. B., Hynson, R. M. G., Whitten, A. E., Kwan, A., Guss, J. M. & Trewhella, J. (2011). *J. Mol. Biol.* **405**, 214–226.
Jacques, D. A., Langley, D. B., Jeffries, C. M., Cunningham, K. A., Burkholder, W. F., Guss, J. M. & Trewhella, J. (2008). *J. Mol. Biol.* **384**, 422–435.
Kanamori, T., Kanou, N., Atomi, H. & Imanaka, T. (2004). *J. Bacteriol.* **186**, 2532–2539.
Kanamori, T., Kanou, N., Kusakabe, S., Atomi, H. & Imanaka, T. (2005). *FEMS Microbiol. Lett.* **245**, 61–65.
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.
McPhillips, T. M., McPhillips, S. E., Chiu, H.-J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M. & Kuhn, P. (2002). *J. Synchrotron Rad.* **9**, 401–406.
Molle, V., Fujita, M., Jensen, S. T., Eichenberger, P., González-Pastor, J. E., Liu, J. S. & Losick, R. (2003). *Mol. Microbiol.* **50**, 1683–1701.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Rowland, S. L., Burkholder, W. F., Cunningham, K. A., Maciejewski, M. W., Grossman, A. D. & King, G. F. (2004). *Mol. Cell*, **13**, 689–701.
Shapir, N., Sadowsky, M. J. & Wackett, L. P. (2005). *J. Bacteriol.* **187**, 3731–3738.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Wang, L., Grau, R., Perego, M. & Hoch, J. A. (1997). *Genes Dev.* **11**, 2569–2579.
Whitney, P. A. & Cooper, T. G. (1972). *Biochem. Biophys. Res. Commun.* **49**, 45–51.
Whitten, A. E., Jacques, D. A., Hammouda, B., Hanley, T., King, G. F., Guss, J. M., Trewhella, J. & Langley, D. B. (2007). *J. Mol. Biol.* **368**, 407–420.
Yokoyama, S., Hirota, H., Kigawa, T., Yabuki, T., Shirouzu, M., Terada, T., Ito, Y., Matsuo, Y., Kuroda, Y., Nishimura, Y., Kyogoku, Y., Miki, K., Masui, R. & Kuramitsu, S. (2000). *Nature Struct. Biol.* **7**, 943–945.