

Herbert L. Axelrod,^{a,b} Piotr Kozbial,^{a,c}
 Daniel McMullan,^{a,d} S. Sri Krishna,^{a,c,e}
 Mitchell D. Miller,^{a,d} Polat Abdubek,^{a,d}
 Claire Acosta,^{a,d} Tamara Astakhova,^{a,e}
 Dennis Carlton,^{a,f} Jonathan
 Caruthers,^{a,b} Hsiu-Ju Chiu,^{a,b} Thomas
 Clayton,^{a,f} Marc C. Deller,^{a,f} Lian
 Duan,^{a,e} Ylva Elias,^{a,f} Julie
 Feuerhelm,^{a,d} Slawomir K.
 Grzechnik,^{a,d} Joanna C. Grant,^{a,d}
 Gye Won Han,^{a,f} Lukasz
 Jaroszewski,^{a,c,e} Kevin K. Jin,^{a,b}
 Heath E. Klock,^{a,d} Mark W. Knuth,^{a,d}
 Abhinav Kumar,^{a,b} David Marciano,^{a,f}
 Andrew T. Morse,^{a,e} Kevin D.
 Murphy,^{a,f} Edward Nigoghossian,^{a,d}
 Linda Okach,^{a,d} Silvy Oommachen,^{a,b}
 Jessica Paulsen,^{a,d} Ron Reyes,^{a,b}
 Christopher L. Rife,^{a,b} Henry J. Tien,^{a,f}
 Christina V. Trout,^{a,f} Henry van den
 Bedem,^{a,b} Dana Weekes,^{a,c} Aprilfawn
 White,^{a,d} Qingping Xu,^{a,b} Chloe
 Zubieta,^{a,b} Keith O. Hodgson,^{a,g} John
 Wooley,^{a,e} Marc-André Elsliger,^{a,f}
 Ashley M. Deacon,^{a,b} Adam Godzik,^{a,c}
 Scott A. Lesley,^{a,d,f} and Ian A.
 Wilson^{a,f,*}

^aJoint Center for Structural Genomics, <http://www.jcsg.org>, USA, ^bStanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA, USA, ^cProgram on Bioinformatics and Systems Biology, Burnham Institute for Medical Research, La Jolla, CA, USA, ^dProtein Sciences Department, Genomics Institute of the Novartis Research Foundation, San Diego, CA, USA, ^eCenter for Research in Biological Systems, University of California, San Diego, La Jolla, CA, USA, ^fDepartment of Molecular Biology, The Scripps Research Institute, La Jolla, CA, USA, and ^gPhoton Science, SLAC National Accelerator Laboratory, Menlo Park, CA, USA

Correspondence e-mail: wilson@scripps.edu

Received 16 April 2009

Accepted 9 June 2009

PDB Reference: tetracenomycin polyketide synthesis protein (NP_636471.1), 3h50, r3h50sf.

Conformational changes associated with the binding of zinc acetate at the putative active site of XcTcmJ, a cupin from *Xanthomonas campestris* pv. *campestris*

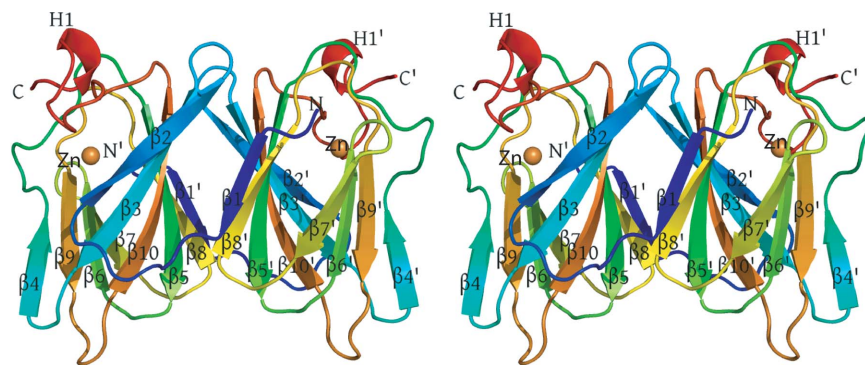
In the plant pathogen *Xanthomonas campestris* pv. *campestris*, the product of the *tcmJ* gene, XcTcmJ, encodes a protein belonging to the RmlC family of cupins. XcTcmJ was crystallized in a monoclinic space group (*C*2) in the presence of zinc acetate and the structure was determined to 1.6 Å resolution. Previously, the apo structure has been reported in the absence of any bound metal ion [Chin *et al.* (2006), *Proteins*, **65**, 1046–1050]. The most significant difference between the apo structure and the structure of XcTcmJ described here is a reorganization of the binding site for zinc acetate, which was most likely acquired from the crystallization solution. This site is located in the conserved metal ion-binding domain at the putative active site of XcTcmJ. In addition, an acetate was also bound within coordination distance of the zinc. In order to accommodate this binding, rearrangement of a conserved histidine ligand is required as well as several nearby residues within and around the putative active site. These observations indicate that binding of zinc serves a functional role in this cupin protein.

1. Introduction

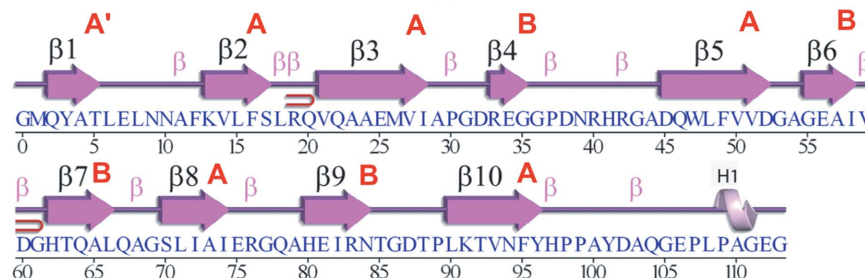
The *tcmJ* gene (Shen & Hutchinson, 1993) of *Xanthomonas campestris* pv. *campestris* strain ATCC 33913 encodes a protein from the cupin superfamily, XcTcmJ, with a molecular weight of 12.2 kDa (residues 1–113) and a calculated isoelectric point of 4.9. XcTcmJ was selected for structure determination in order to extend the structural coverage of proteins of unknown function from pathogenic bacteria, in this case *X. campestris* pv. *campestris* (Leysn *et al.*, 1984). This Gram-negative bacterium causes black rot, which is one of the major worldwide diseases of cruciferous crops (Crossman & Dow, 2004; Williams, 1980). In *X. campestris* pv. *campestris*, a variety of polyketide metabolites are synthesized which are precursors of antibiotics and antitumor therapeutics (Donadio *et al.*, 2007; Tudor-Nelson *et al.*, 2003). XcTcmJ is among several gene products from *X. campestris* that have been annotated as being involved in tetracenomycin polyketide biosynthesis (Overbeek *et al.*, 2005). A sequence-database search found four proteins that are orthologous to XcTcmJ (Fig. 1c), as well as over 1000 more distantly related proteins from several different protein families.

XcTcmJ belongs to the cupin superfamily, the members of which have a conserved β -barrel fold and have been classified into at least 18 functional classes involved in a range of biochemical processes, including plant growth and development (Dunwell *et al.*, 2001, 2004). In enzymatic cupins, the nature of the bound metal cofactor is believed to influence the type of reaction catalyzed. Although some characteristics of the metal ion-binding motif are conserved across all members of the cupin superfamily, several crystal structures of enzymatic cupins have been reported without any bound metal cofactor.

Comparison of apo and metal-bound cupin structures should, therefore, enhance our understanding of how metal-ion binding influences functional specificity in this superfamily of proteins. Since no metal ion was found in the previously solved crystal structure of an identical protein from a different strain (XC5357 from *X. campestris* pv. *campestris* strain 17; PDB code 2gu9; Chin *et al.*, 2006), it was speculated that XC5357 may not require a metal ion for activity.



(a)



(b)

	20	30	40	50	60
XcTcmJ	13	.. KVLFSLRQVQAAEMV IAPGDREGG .. PDNRHRGA .. DQWLFVVDGAGEAIVDG .. HTQA ..			
YP_362950.1	13	.. KVLFSVVRQVQAAEMA IAPQGQEGG .. PDNRHRGA .. DQWLYVVDGTGEAVIEG .. QTHA ..			
YP_411980.1	13	.. RIVLGNPSSQAAEMV LLLPGQAEGG .. PHNRHRGA .. DQWLYVVSGETGIATIND .. KRYR ..			
YP_136688.1	13	.. EVAMEAEAQAAEMTIEPGRSVG .. PENYHADS .. DQWLFVVSGETVTVDG .. DTHH ..			
ZP_00378045.1	29	.. KVLWTGEHAQIVVM TIPPGEIG .. DEVHEHT .. DQILTFVSGTGEADLNC .. HNHP ..			
YP_285289.1	20	.. QVLYTAKNSQLVLM SLKPQEDIG .. AEIHK V .. DQFFRVEEGTGEAVLNG .. ARTP ..			
YP_001159114.1	18	.. RVLVTGKYLQLVVM TIPPGEIG .. DEVHEHT .. DQILTFVSGVGAALVAG .. ERD ..			
YP_138078.1	27	.. PEVTTGGERFSAVV KLEPGKGHE .. RHTHPES .. DEILYVVRGEGEQEVD .. ETRD ..			
NP_107188.1	26	.. RVLVTTRRMQLVLM ALPGEEIG .. AEVHADR .. DQFFRVEKGTGETIWDG .. VASD ..			
2O2A	35	.. RALWTGDHLQVTLMS IQVGEDIG .. LEIHPHL .. DQFLRVEEGRGLVQMGH .. RQDNLF ..			
1V70	20	.. IPVVFQSERMLYDLYA LLLPGQAQ .. KVVHHEGS .. DKVYYALEGEVVVRVGE .. EEAL ..			
2ETE	62	AEWPGTNTLGVSMNRVDFAPGGTN .. PPHIHPRA .. TEIGMVMKGLLGLGSLDSGNKLYSRV ..			
2B8M	19	.. EKLVNTEHVQINHI VLLPRGEQMP .. KHYSN S .. YVHLI I I K G E M T L L G .. QEP ..			
1LRH	32	.. GALNHGMKEVEVWLQ TISPQRT .. IHRHS C .. EVVFTVLKKGKTLMLGS .. SLKYPGPQ ..			
1O4T	37	.. KEAMHNKARLFARMKLPPGSSV .. GLKHHEGE .. FEIYY ILLGEGVFHDNG .. KDVP ..			
2GC3	61	EE .. KEGD LNFATTVL YPGKVG .. KEFFFTKGFHAKLDRAEVYV ALKKGCM L L Q T P .. E .. GDAKWIS ..			

	70	80	90	100
XcTcmJ	66	... LQAGSLIAIERGQAHEIRNTGDT .. PLKTVNFY .. HPPAY .. DAQ .. GEP		
YP_362950.1	66	... LQAGSLIAIERGETHEIRNTGST .. PLKTVNFY .. LPPAY .. DAQ .. GEP		
YP_411980.1	66	... LDEGSLLLIERGDEHEIRNTGED .. PLQTLNFY .. VPPAY .. TAE .. GEE		
YP_136688.1	66	... VDAGDLIRIEAGERHGIENDGPE .. PLETVNFY .. TPP ..		
ZP_00378045.1	80	... IEAGDQCAVPAGTQHNFRTNGEE .. SLVLYTVY .. SPP .. AVD .. AAF		
YP_285289.1	70	... IQAGFAVLVPAGVNHNIINTGT .. PLKLYTLY .. APPNH .. RDG .. VIH		
YP_001159114.1	69	... IARSSQVVVPAGTRHNIIVNTGPN .. PLVLYTIY .. APP .. ADQ .. AVH		
YP_138078.1	78	... IATGDMVFIPEGVEHGVNTGWE .. PLILFAVY .. APPGP .. EDV .. LRN		
NP_107188.1	77	... IKSDMAMIGPAGARHNVKNTGKK .. PLKLYTLY .. APP .. MDG .. TVH		
2O2A	89	... QEEVFDDYAILIPAGTWHNVRTGNR .. PLKLYSIY .. APP .. PHG .. TVH		
1V70	71	... LAPGMAAFAPAGAPHGVRNESAS .. PALLLVVT .. AP ..		
2ETE	123	... VRAGETFFVIPRGLMHFQFNVGKT .. EAYMVVSFNS .. DPP .. TPV .. LTK		
2B8M	68	... HNYKGENIVVVPFNKMLITQINNSD .. ILEFFVVK .. APHPK .. KLN ..		
1LRH	88	QEIPFFONTTFSIPVNDPHQVWNSDEHEDLQVLVVIS .. RPPAKIFLYDDWSMPHTAAVLKF ..		
1O4T	88	... IKAGDVCFDTSGESHSIENTGNT .. DLEFLAVI .. I ..		
2GC3	123	... MEPGTVVVYVPPYWAHRTVNIIGDE .. PFIFLAIY .. PADAG .. HDY .. GTI		

(c)

Figure 1 Crystal structure of *XcTcmJ* from *X. campestris* pv. *campestris*. (a) Stereo ribbon diagram of the *XcTcmJ* dimer color coded from the N-terminus (blue) to the C-terminus (red). Helices H1 and H1' and β -strands (β 1– β 10, β 1'– β 10') are indicated. A Zn atom in the vicinity of the putative active site of each monomer is shown as a sphere (orange). (b) Diagram showing the secondary-structure elements of *XcTcmJ* superimposed on its primary sequence. The 3_{10} -helices, β -strands and β -turns are indicated. The β -sheets are indicated in red letters. The β -hairpins are depicted as red loops. (c) Multiple sequence alignments for *XcTcmJ* and related proteins. The first four protein sequences belong to *XcTcmJ* (GenBank NP_636471) and its putative orthologs (GenBank YP_362950, YP_411980 and YP_136688). Sequences 5–16 belong to proteins that are similar but have a different pattern of conserved residues (GenBank or PDB codes ZP_00378045, YP_285289, YP_001159114, YP_138078, NP_107188, 2o2a, 1v70, 1o4t, 1lrh and 2gc3). The positions of the three residues (His41, Gln46 and His80) within coordination distance of the Zn atom in the putative active site in the *XcTcmJ* structure are labeled with arrows. The two conserved cupin sequence motifs (Dunwell *et al.*, 2004) are shaded in green and blue.

Table 1

Summary of crystal parameters, data collection and refinement statistics for XcTcmJ (PDB code 3h50).

Values in parentheses are for the highest resolution shell.

	λ_1 MADSe	λ_2 MADSe	λ_3 MADSe
Space group	C2		
Unit-cell parameters (\AA , $^\circ$)	$a = 48.92$, $b = 72.03$, $c = 32.75$, $\beta = 110.76$		
Data collection			
Wavelength (\AA)	0.9184	0.9796	0.9793
Resolution range (\AA)	28.5–1.60	28.5–1.60	28.5–1.60
	(1.64–1.60)	(1.64–1.60)	(1.64–1.60)
No. of observations	100153	100180	66773
No. of unique reflections	13913	13898	13763
Completeness (%)	98.6 (91.4)	98.5 (90.3)	98.0 (95.2)
Mean $I/\sigma(I)$	14.0 (2.6)	12.9 (2.4)	10.5 (1.9)
R_{merge} on I^\dagger (%)	11.9 (67.9)	13.1 (73.6)	10.6 (58.8)
Model and refinement statistics			
Resolution range (\AA)	28.5–1.60		
No. of reflections (total)	13912		
No. of reflections (test)	696		
Completeness (%)	98.6		
Data set used in refinement	λ_1		
Cutoff criterion	$ F > 0$		
$R_{\text{cryst}}^\ddagger$	0.182		
R_{free}^\S	0.220		
Stereochemical parameters			
Restraints (r.m.s.d. observed)			
Bond angles ($^\circ$)	1.42		
Bond lengths (\AA)	0.014		
Average isotropic B value (\AA^2)	13.1		
ESU $^\parallel$ based on R_{free} (\AA)	0.10		
Protein residues/atoms	114/865		
Waters/ions	154/6		

$^\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$. $^\ddagger R_{\text{cryst}} = \frac{\sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum_{hkl} |F_{\text{obs}}|}$, where F_{calc} and F_{obs} are the calculated and observed structure-factor amplitudes, respectively. $^\S R_{\text{free}}$ is the same as R_{cryst} but for 5.0% of the total reflections chosen at random and omitted from refinement. $^\parallel$ Estimated overall coordinate error (Collaborative Computational Project, Number 4, 1994; Cruickshank, 1999).

Here, we report the crystal structure of XcTcmJ from *X. campestris* pv. *campestris* with zinc acetate bound at the putative active site at 1.6 \AA resolution which offers additional insights into the role of metal ions in the structure and function of cupins. The structure was determined using the semi-automated high-throughput pipeline of the Joint Center for Structural Genomics (JCSG; Lesley *et al.*, 2002) as part of the National Institute of General Medical Sciences Protein Structure Initiative (PSI; <http://www.nigms.nih.gov/Initiatives/PSI/>).

2. Materials and methods

2.1. Protein production and crystallization

The gene encoding XcTcmJ (GenBank NP_636471.1; gi:21230554) was amplified by polymerase chain reaction (PCR) from genomic DNA using *PfuTurbo* DNA polymerase (Stratagene) and primers corresponding to the predicted 5' (aacctgtacttcaggcATGCAGTACGCAACGTTGGA) and 3' (gagttaattaagtgcgtaGCCTTCGCCTGCCGCAA) ends. The PCR product was cloned into plasmid pSpeedET, which encodes an expression and purification tag followed by a tobacco etch virus (TEV) protease cleavage site (MGSDKIHSHHHHHENLYFQ/G) at the amino-terminus of the full-length protein. The cloning junctions were confirmed by DNA sequencing. Protein expression was performed in selenomethionine-containing medium with suppression of normal methionine synthesis using *Escherichia coli* strain GeneHogs (Invitrogen). At the end of fermentation, lysozyme was added to the culture to a final concentration of 250 $\mu\text{g ml}^{-1}$ and the cells were harvested. After one freeze-thaw cycle, the cells were sonicated in lysis buffer [50 mM HEPES pH 8.0, 50 mM NaCl, 10 mM imidazole, 1 mM tris(2-carboxyethyl)-

phosphine hydrochloride (TCEP)]. The lysate was clarified by centrifugation at 32 500g for 30 min and loaded onto nickel-chelating resin (GE Healthcare) pre-equilibrated with lysis buffer. The resin was washed with wash buffer [50 mM HEPES pH 8.0, 300 mM NaCl, 40 mM imidazole, 10% (v/v) glycerol, 1 mM TCEP] and the protein was eluted with elution buffer [20 mM HEPES pH 8.0, 300 mM imidazole, 10% (v/v) glycerol, 1 mM TCEP]. The eluate was buffer-exchanged with HEPES crystallization buffer (20 mM HEPES pH 8.0, 200 mM NaCl, 40 mM imidazole, 1 mM TCEP) and treated with 1 mg of TEV protease per 15 mg of eluted protein. The digested protein was passed over nickel-chelating resin (GE Healthcare) pre-equilibrated with HEPES crystallization buffer and the resin was washed with the same buffer. The flowthrough and wash fractions were combined and concentrated to 18 mg ml^{-1} by centrifugal ultrafiltration (Millipore) for crystallization assays. XcTcmJ was crystallized using the nanodroplet vapor-diffusion method (Santarsiero *et al.*, 2002) with standard JCSG crystallization protocols (Lesley *et al.*, 2002). Sitting drops composed of 200 nl protein mixed with 200 nl crystallization solution were equilibrated against a 50 μl reservoir at 277 K for 34 d prior to harvesting. Initial screening for diffraction was carried out using the Stanford Automated Mounting system (SAM; <http://smb.slac.stanford.edu/facilities/hardware/SAM/UserInfo>; Cohen *et al.*, 2002) at the Stanford Synchrotron Radiation Lightsource (SSRL). The crystallization reagent contained 35% (v/v) 2-propanol, 0.2 M zinc acetate and 0.1 M imidazole pH 8.0. No additional cryoprotectant was added to the crystal. Diffraction data from a cube-shaped crystal of approximate dimensions 100 \times 100 \times 100 μm mounted in a nylon loop were indexed in monoclinic space group C2 (Table 1). The molecular weight and oligomeric state of XcTcmJ were determined using a 1 \times 30 cm Superdex 200 column (GE Healthcare) in combination with static light scattering (Wyatt Technology). The mobile phase consisted of 20 mM Tris pH 8.0, 150 mM NaCl and 0.02% (w/v) sodium azide.

2.2. Data collection, structure solution and refinement

Multi-wavelength anomalous diffraction (MAD) data were collected at SSRL on beamline BL11-1 at wavelengths corresponding to the high-energy remote (λ_1), inflection (λ_2) and peak (λ_3) of a selenium MAD experiment. The data sets were collected at 100 K using an ADSC Quantum 315 CCD detector. The MAD data were integrated and scaled using the XDS package (Kabsch, 1993). Data statistics are summarized in Table 1. The selenium substructure was solved with *SHELXD* (Sheldrick, 2008) and the MAD phases were refined with *autoSHARP* (Vonnrhein *et al.*, 2007), which resulted in a mean figure of merit of 0.42 with two selenium sites. Automated model building was performed by iterative cycles of *ARP/wARP* (Cohen *et al.*, 2004). Model completion and refinement were performed using *Coot* (Emsley & Cowtan, 2004) and *REFMAC* (Winn *et al.*, 2003). Refinement statistics are summarized in Table 1.

2.3. Validation and deposition

The quality of the crystal structure was analyzed using the *JCSG Quality Control* server, which verifies the stereochemical quality of the model using *AutoDepInputTool* (Yang *et al.*, 2004), *MolProbity* (Lovell *et al.*, 2003) and *WHATIF* 5.0 (Vriend, 1990), the agreement between the atomic model and the data using *SFHECK* 4.0 (Laskowski *et al.*, 2005) and *RESOLVE* (Terwilliger, 2002), the protein sequence using *ClustalW* (Chenna *et al.*, 2003), the atom occupancies using *MOLEMAN2* (Kleywegt *et al.*, 2001a,b) and the consistency of NCS pairs. Protein quaternary-structure analysis used the EBI *PISA* server (Krissinel & Henrick, 2007). Fig. 1(b) was

adapted from an analysis using *PDBsum* (Laskowski *et al.*, 2005). Figs. 1(a) and 2(a) were prepared with *PyMOL* (DeLano Scientific) and Fig. 1(c) was prepared with *ESPrnt* (Gouet *et al.*, 2003). Atomic coordinates and experimental structure factors for *XcTcmJ* from *X. campestris* pv. *campestris* strain ATCC 33913 at 1.6 Å resolution have been deposited in the PDB under accession code 3h50.

3. Results and discussion

3.1. Overall structure

The crystal structure of *XcTcmJ* (Fig. 1a) was determined to 1.6 Å resolution using the MAD method. Data collection, model and refinement statistics are summarized in Table 1. The final model is comprised of a protein monomer (residues 1–113 and Gly0 from the TEV protease cleavage site), four zinc ions, two acetate molecules and 151 water molecules in the asymmetric unit. As the electron density was not well resolved for some of the side-chain atoms of Lys13, Gln20 and Arg42, these atoms were not included in the structural model. The Matthews coefficient (V_M ; Matthews, 1968) for *XcTcmJ* is 2.19 Å³ Da⁻¹ and the estimated solvent content is 43.8%. The Ramachandran plot produced by *MolProbity* shows that 98.2% and 100% of the residues are in favored and allowed regions, respectively.

XcTcmJ is composed of ten β-strands (β1–β10) and one 3₁₀-helix (Fig. 1a). The total β-sheet content is 45.1% and the 3₁₀-helical content is 2.6%. The *XcTcmJ* monomer (Fig. 1a) contains two anti-parallel β-sheets that form a jelly-roll β-sandwich and a topology characteristic of the RmlC-like cupin superfamily (SCOP sunid 51182; Andreeva *et al.*, 2004). The larger β-sheet has a six-stranded 23(10)581' topology, while the smaller β-sheet has a four-stranded 4967 topology (Fig. 1a).

The crystallographic packing of *XcTcmJ* indicates that a homodimer is the biologically relevant form. This finding is consistent with previous reports (Chin *et al.*, 2006; Chu *et al.*, 2005) and with the results of our analytical size-exclusion chromatography coupled with

static light scattering. In the homodimer, the N-terminal β-strand β1 of each monomer interacts with β8 in the twofold symmetry-related monomer in a domain-swapped arrangement. The dimer resembles two juxtaposed barrels. One end of each barrel, proximal to the C-terminus and zinc-binding site, is solvent accessible and forms a putative active site. The two symmetry-related, six-stranded β-sheets in the dimer show a pronounced twist, which enhances structural complementarity to the tightly packed, monomer–monomer interface.

A *DALI* search (Holm *et al.*, 2008) for entries in the PDB similar to *XcTcmJ* revealed a multitude of related cupin structures. Alignment of the amino-acid sequence of *XcTcmJ* with those of related structures is shown in Fig. 1(c). In general, *XcTcmJ* shares only 10–20% sequence identity with related cupins, even though the structures are quite similar. Many structural homologues contain a bound metal ion, but others do not. An example from the multitude of similar structures is the metal-deficient BH2720 from *Bacillus halodurans* (gi:10175341; PDB code 2oa2), which was reported by the JCSG and which has 28% sequence identity and an r.m.s.d. of 1.46 Å with *XcTcmJ* for 99 superimposed C^α atoms.

3.2. The zinc acetate-binding site

X-ray fluorescence emission spectroscopy on the same *XcTcmJ* crystal as used for structure determination indicated the presence of selenium and zinc and the absence of other transition metals. To corroborate that zinc is bound at specific sites in the structure and is not just present in the bulk solvent, an anomalous difference map was calculated using the Se remote data (λ_1 MADSe in Table 1). The map revealed that, in addition to two electron-dense peaks $\sim 16\sigma$ above the r.m.s. background level at positions corresponding to the Se atoms, other peaks at $\sim 13\sigma$ were found that were not associated with the selenium sites. These peak-height ratios were consistent with the theoretical ratio of $\Delta f''$ for selenium versus zinc at λ_1 . One of the bound zinc ions is specifically coordinated to the metal-binding site (Figs. 1a and 2), while three other zinc ions are found in crystal-

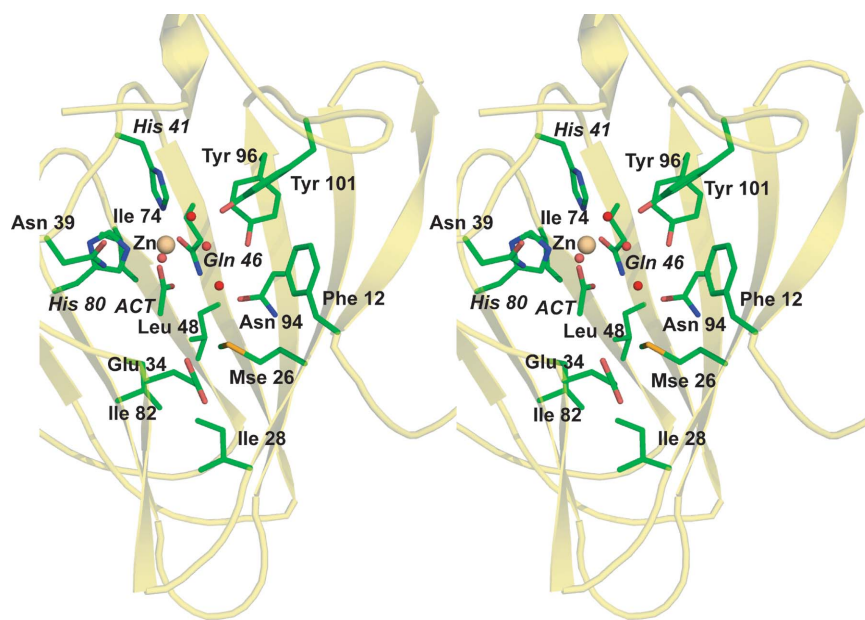


Figure 2 Stereo diagram of residues in the putative active site of *XcTcmJ* from *X. campestris* pv. *campestris*. The C, O, N and Se (from selenomethionine) atoms of the side chains of *XcTcmJ* and the bound acetate ligand (labeled ACT) are shown in green, red, blue and orange, respectively. The bound Zn atom and water molecules are shown as orange and red spheres, respectively. His41, Gln46 and His80, which are within coordination distance of the bound zinc, are labeled in italics.

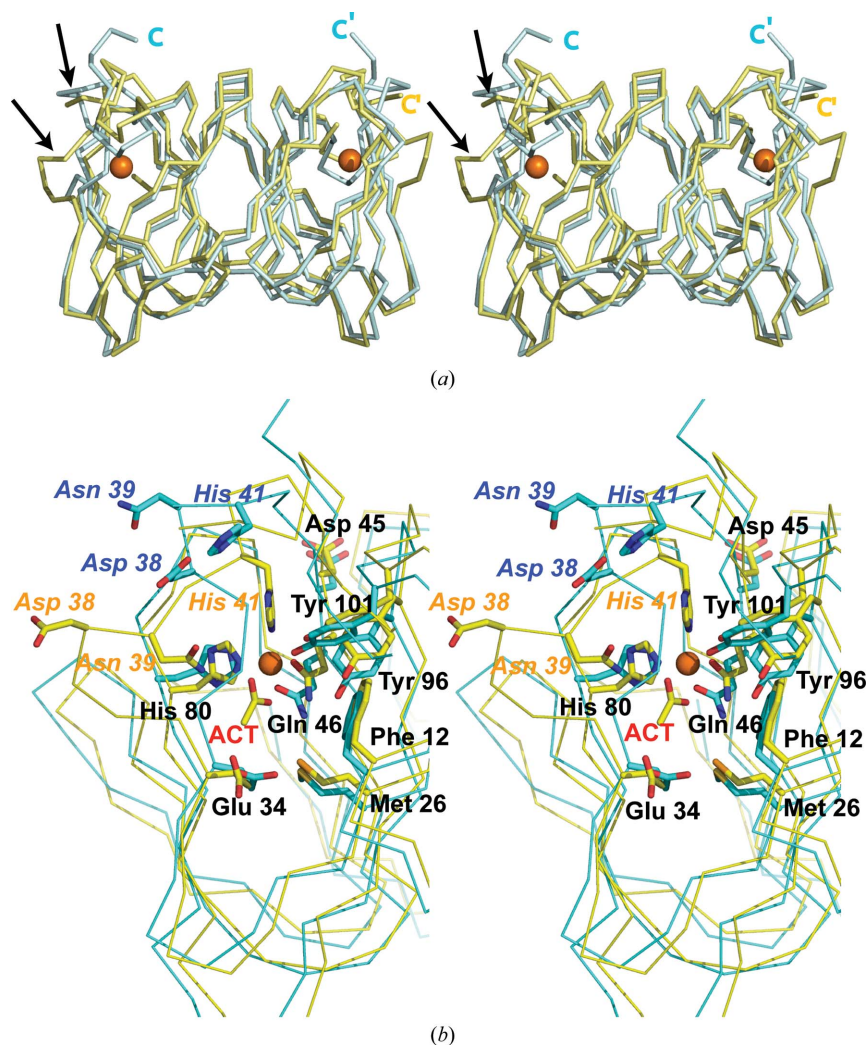
Table 2Short-range contacts to the Zn atom in the putative active site of *XcTcmJ*.

Atom	Interatomic distance (Å)
His41 NE2	2.0
Gln46 OE1	2.1
His80 NE2	2.2
Acetate (ACT) 118 O	2.2
Acetate (ACT) 118 OXT	2.2
Wat181 O	2.3

packing interfaces (not shown). The active-site zinc is within coordination distance of the side-chain NE2 atoms of two histidines (His41 and His80) and the carboxamide O atom of Gln46. A bound acetate was also identified in an ~ 5 Å deep hydrophobic pocket and was most likely acquired from the crystallization solution. The positions of the acetate carboxylate emulates coordination to the zinc by Glu or Asp side chains in other structures. Here, the Glu34 carboxyl is 5 Å from the bound acetate and is too distant for zinc coordination even if it adopted another rotamer (Fig. 2*b*).

A bound water molecule, acetate and three side-chain ligands indicate octahedral coordination to the bound zinc. In zinc metallo-proteins the metal ion is thought to play either a structural or a catalytic role, with tetrahedral coordination being the most prevalent (McCall *et al.*, 2000). Structural zinc sites usually contain four protein side-chain ligands (Cys, Glu or Asp) in a tetrahedral coordination, whereas catalytic zinc sites tend to utilize more than four ligands, with one being a water (Auld, 2009). Thus, the octahedral coordination geometry and coordinating water molecule both suggest a likely functional role for the bound zinc. A list of interactions surrounding the zinc at the putative active site is shown in Table 2.

Most members of the cupin superfamily typically share two conserved sequence motifs (Dunwell *et al.*, 2001, 2004; see Fig. 1*c*) that contribute His, Glu and Asp side chains to a conserved metal ion-binding site. An alignment between *XcTcmJ* and related sequences is shown in Fig. 1*c*), with the conserved N-terminal sequence motif highlighted in green and the C-terminal sequence motif in blue. This alignment indicates that, in *XcTcmJ*, either Asp38 or Asp45 is likely to be a ligand for a bound metal ion. However, the

**Figure 3**

Comparison of the zinc acetate-bound and metal-free (PDB code 2gu9) structures of *XcTcmJ* from *X. campestris* pv. *campestris*. (a). Stereo diagram showing a superposition of the C α traces of the zinc-bound (yellow) and the apo (cyan) forms. The two regions that show significant conformational differences are indicated by arrows. The Zn atom is shown as an orange sphere. (b). Stereo diagram of the residues in the putative active site of the zinc acetate-bound structure superimposed on the apo structure (PDB code 2gu9). The C atoms from the side chains and acetate ligand (ACT) of the zinc acetate-bound structure are shown in yellow, while the C atoms of the apo structure are shown in cyan. O, N and Se (SeMet) atoms in both structures are shown in red, blue and orange, respectively. The Zn atom in the metal ion-bound form is shown as an orange sphere. Those side chains that show significant structural differences in the ligand-bound and the apo states are labeled in italics.

crystal structure shows that the ligand of the metal ion is the carboxamide side chain of Gln46, even though Glu and Asp side chains are more prevalent ligands than the carboxamide side chains of Gln and Asn residues (Hsin *et al.*, 2008). This type of coordination is also observed in the crystal structure of a related cupin, phosphomannose isomerase from *Candida albicans* (Cleasby *et al.*, 1996), where an amide side chain is also within coordination distance of the bound metal.

3.3. Comparison with apo XC5357

Apo *XcTcmJ* has previously been reported at 1.4 Å resolution for XC5357 from a different strain of *X. campestris* (PDB code 2gu9; Chin *et al.*, 2006). Although XC5357 and *XcTcmJ* share identical amino-acid sequences, their structures show significant differences. XC5357 was crystallized in a triclinic form without zinc, whereas *XcTcmJ* crystallized in a monoclinic space group in the presence of zinc acetate. Examination of the structures with and without zinc reveals significant rearrangements around the metal-binding site (Figs. 3*a* and 3*b*). Strand β_4 (Gly35–Asp38) in the apo structure is repositioned as a surface loop in the *XcTcmJ* structure. Additional rearrangements (Asn39–Gly43) in *XcTcmJ* contribute to the formation of a longer loop segment connecting β_4 (Asp32–Gly36) and β_5 (Asp45–Asp52) at the open end of the cupin barrel. The conserved His41 is located in one of the loops and the backbone rearrangements in *XcTcmJ* result in His41 relocating by ~ 8 Å to coordinate with the Zn atom near the center of the cupin barrel. In contrast, the other zinc ligands, Gln46 and His80, do not move significantly.

Comparison of the solvent-accessible surfaces (see supplementary material¹) indicates that a deep surface cavity is eliminated upon the conformational rearrangements arising from the binding of zinc acetate. In the apo structure, this cavity is located between anti-parallel β -strands at the C-terminus of each monomer and a loop in the vicinity of His41. In *XcTcmJ*, this histidine and loop are repositioned upon metal ligation, so that the interactions with the C-terminal region are lost. One possible explanation for the loss of these interactions is that the crystallized XC5357 contained an eight-residue sequence tag at the C-terminus (Chu *et al.*, 2005) that was absent from the crystallized *XcTcmJ*. Consequently, this cavity is replaced by a shallow depression in *XcTcmJ* and the two-stranded β -sheet becomes a loop and a small 3_{10} helix. Notwithstanding, both the *XcTcmJ* and apo structures retain the cavity corresponding to the zinc acetate-binding site. Furthermore, a new channel is formed on the surface of *XcTcmJ* that is surrounded by nonpolar residues including Ala11, Phe12, Val27 and Ile28 that may reflect the natural substrate-binding site.

3.4. Functional implications and conclusions

Extensive investigations have demonstrated the importance of metal ions in the structure and function of diverse cupins (Dunwell *et al.*, 2001, 2004). However, it is not understood why some cupins require metal ions for biological activity whereas others do not. Conformational changes induced by metal binding could exert control of reaction specificity or even inhibition of certain biological processes (Babor *et al.*, 2005; Barondeau & Getzoff, 2004). From the structural studies alone, it is difficult to define the exact functional implications of metal ion binding to *XcTcmJ*. However, metal ion binding can have a profound impact on the structure, as illustrated here. Zinc is specifically bound and reorganizes the structure and

acetate (likely from the crystallization solution) then serves as a counter ion in the formation of the zinc–protein complex.

Additional information about the protein described in this study is available from TOPSAN (Krishna *et al.*, 2010) <http://www.topsan.org/explorer?PDBid=3h50>.

This work was supported by the National Institutes of Health, Protein Structure Initiative (<http://www.nigms.nih.gov>), grant Nos. P50 GM62411 and U54 GM074898. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource (SSRL). The SSRL is a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research and by the National Institutes of Health (National Center for Research Resources, Biomedical Technology Program and the National Institute of General Medical Sciences). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

References

- Andreeva, A., Howorth, D., Brenner, S. E., Hubbard, T. J., Chothia, C. & Murzin, A. G. (2004). *Nucleic Acids Res.* **32**, D226–D229.
- Auld, D. (2009). *Biometals*, **22**, 141–148.
- Babor, M., Greenblatt, H. M., Edelman, M. & Sobolev, V. (2005). *Proteins*, **59**, 221–230.
- Barondeau, D. P. & Getzoff, E. D. (2004). *Curr. Opin. Struct. Biol.* **14**, 765–774.
- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D. (2003). *Nucleic Acids Res.* **31**, 3497–3500.
- Chin, K.-H., Chou, C.-C., Wang, A. H. & Chou, S.-H. (2006). *Proteins*, **65**, 1046–1050.
- Chu, C.-L., Chin, K.-H., Lin, F.-Y., Chou, C.-C., Lee, C.-C., Shr, H.-L., Lyu, P.-C., Wang, A. H.-J. & Chou, S.-H. (2005). *Acta Cryst.* **F61**, 697–699.
- Cleasby, A., Wonacott, A., Skarzynski, T., Hubbard, R. E., Davies, G. J., Proudfoot, A. E., Bernard, A. R., Payton, M. A. & Wells, T. N. (1996). *Nature Struct. Biol.* **3**, 470–479.
- Cohen, A. E., Ellis, P. J., Miller, M. D., Deacon, A. M. & Phizackerley, R. P. (2002). *J. Appl. Cryst.* **35**, 720–726.
- Cohen, S. X., Morris, R. J., Fernandez, F. J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V. S., Kleywegt, G. J. & Perrakis, A. (2004). *Acta Cryst.* **D60**, 2222–2229.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Crossman, L. & Dow, J. M. (2004). *Microbes Infect.* **6**, 623–629.
- Cruickshank, D. W. J. (1999). *Acta Cryst.* **D55**, 583–601.
- Donadio, S., Monciardini, P. & Sosio, M. (2007). *Nat. Prod. Rep.* **24**, 1073–1109.
- Dunwell, J. M., Culham, A., Carter, C. E., Sosa-Aguirre, C. R. & Goodenough, P. W. (2001). *Trends Biochem. Sci.* **26**, 740–746.
- Dunwell, J. M., Purvis, A. & Khuri, S. (2004). *Phytochemistry*, **65**, 7–17.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Gouet, P., Robert, X. & Courcelle, E. (2003). *Nucleic Acids Res.* **31**, 3320–3323.
- Holm, L., Kääriäinen, S., Rosenstrom, P. & Schenkel, A. (2008). *Bioinformatics*, **24**, 2780–2781.
- Hsin, K., Sheng, Y., Harding, M. M., Taylor, P. & Walkinshaw, M. D. (2008). *J. Appl. Cryst.* **41**, 963–968.
- Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
- Kleywegt, G. J., Zou, J.-Y., Kjeldgaard, M. & Jones, T. A. (2001*a*). *International Tables for Crystallography*, Vol. F, edited by M. G. Rossmann & E. Arnold, pp. 353–356. Dordrecht: Kluwer Academic Publishers.
- Kleywegt, G. J., Zou, J.-Y., Kjeldgaard, M. & Jones, T. A. (2001*b*). *International Tables for Crystallography*, Vol. F, edited by M. G. Rossmann & E. Arnold, pp. 497–506. Dordrecht: Kluwer Academic Publishers.
- Krishna, S. S., Weekes, D., Bakolitsa, C., Elsliger, M.-A., Wilson, I. A., Godzik, A. & Wooley, J. (2010). *Acta Cryst.* **F66**, 1143–1147.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Laskowski, R. A., Chistyakov, V. V. & Thornton, J. M. (2005). *Nucleic Acids Res.* **33**, D266–D268.
- Lesley, S. A. *et al.* (2002). *Proc. Natl Acad. Sci. USA*, **99**, 11664–11669.

¹ Supplementary material has been deposited in the IUCr electronic archive (Reference: WD5112).

- Leyns, F., De Cleene, M., Swings, J.-G. & De Ley, J. (1984). *Bot. Rev.* **50**, 308–355.
- Lovell, S. C., Davis, I. W., Arendall, W. B. III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCall, K. A., Huang, C. & Fierke, C. A. (2000). *J. Nutr.* **130**, 1437S–1446S.
- Overbeek, R. *et al.* (2005). *Nucleic Acids Res.* **33**, 5691–5702.
- Santarsiero, B. D., Yegian, D. T., Lee, C. C., Spraggon, G., Gu, J., Scheibe, D., Uber, D. C., Cornell, E. W., Nordmeyer, R. A., Kolbe, W. F., Jin, J., Jones, A. L., Jaklevic, J. M., Schultz, P. G. & Stevens, R. C. (2002). *J. Appl. Cryst.* **35**, 278–281.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Shen, B. & Hutchinson, C. R. (1993). *Science*, **262**, 1535–1540.
- Terwilliger, T. C. (2002). *Acta Cryst.* **D58**, 1937–1940.
- Tudor-Nelson, S. M., Minsavage, G. V., Stall, R. E. & Jones, J. B. (2003). *Phytopathology*, **93**, 1415–1421.
- Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.
- Vriend, G. (1990). *J. Mol. Graph.* **8**, 52–56.
- Williams, P. H. (1980). *Plant Dis.* **64**, 736–742.
- Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). *Methods Enzymol.* **374**, 300–321.
- Yang, H., Guranovic, V., Dutta, S., Feng, Z., Berman, H. M. & Westbrook, J. D. (2004). *Acta Cryst.* **D60**, 1833–1839.