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

Structure of 5-formyltetrahydrofolate cyclo-ligase from *Bacillus anthracis* (BA4489)

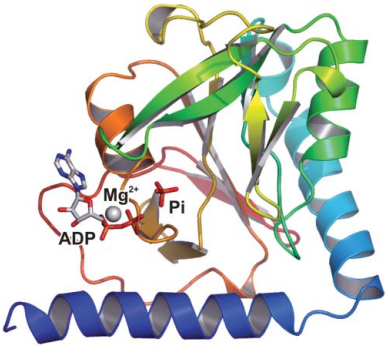
Christoph Meier,[a,b] Lester G. Carter,[a] Graeme Winter,[c] Ray J. Owens,[a] David I. Stuart[a,b] and Robert M. Esnouf[a,b]*

[a]Oxford Protein Production Facility, The Henry Wellcome Building for Genomic Medicine, Oxford University, Roosevelt Drive, Oxford OX3 7BN, England, [b]Division of Structural Biology, The Henry Wellcome Building for Genomic Medicine, Oxford University, Roosevelt Drive, Oxford OX3 7BN, England, and [c]CCLRC Daresbury Laboratory, Warrington, Cheshire WA4 4AD, England

Correspondence e-mail: robert@strubi.ox.ac.uk

Received 21 December 2006
Accepted 12 February 2007

PDB Reference: 5-formyltetrahydrofolate cyclo-ligase, 2jcb, r2jcbsf.



Bacillus anthracis is a spore-forming bacterium and the causative agent of the disease anthrax. The Oxford Protein Production Facility has been targeting proteins from *B. anthracis* in order to develop high-throughput technologies within the Structural Proteomics in Europe project. As part of this work, the structure of 5-formyltetrahydrofolate cyclo-ligase (BA4489) has been determined by X-ray crystallography to 1.6 Å resolution. The structure, solved in complex with magnesium-ion-bound ADP and phosphate, gives a detailed picture of the proposed catalytic mechanism of the enzyme. Chemical differences from other cyclo-ligase structures close to the active site that could be exploited to design specific inhibitors are also highlighted.

1. Introduction

The Oxford Protein Production Facility (OPPF) was established to develop methods for high-throughput protein production and crystallization. As part of these developments, a pilot project was undertaken to study 48 proteins from *Bacillus anthracis*, focusing on protein families that are well conserved across a wide range of bacteria. The causative agent of anthrax, *B. anthracis* is a large Gram-positive spore-bearing bacterium. Sequencing of the genome of the *B. anthracis* Ames strain (Read *et al.*, 2003) revealed two plasmids, pXO1 and pXO2, that carry the major virulence factors, as well as 5.23 megabases of normal chromosomal DNA predicted to code for about 5311 genes. The set of proteins chosen for this study (Au *et al.*, 2006) are all encoded by the chromosomal DNA.

Folate, a water-soluble B vitamin discovered in the 1940s, is a required cofactor for cell growth in all organisms and is synthesized by plants, prokaryotes and yeast. It acts a carrier of one-carbon groups and is necessary for the synthesis of DNA and RNA precursors as well as for methionine biosynthesis. In the cell, folic acid exists as a family of cofactors with different oxidation states [folate, dihydrofolate (DHF) and tetrahydrofolate (THF)]; the one-carbon units carried by these cofactors can also have different oxidation states (methanol to formate) which are enzymatically interconvertable. These different forms of THF donate or accept one-carbon units in a range of metabolic reactions (the so-called one-carbon metabolism), which include the *de novo* synthesis of purines and thymidylate (methylation of dUMP to give dTMP) and the remethylation of homocysteine to methionine [methionine can subsequently be adenylated to form *S*-adenosylmethionine (SAM), itself a cofactor and one-carbon donor for numerous other methylation reactions]. An overview of bacterial folate metabolism is given in Fig. 1(*a*).

5-Formyl-THF is a member of the folate cofactor family. However, it is not used directly as a one-carbon donor, but instead acts as a regulator of folate metabolism (Stover & Schirch, 1991; Bertrand & Jolivet, 1989). The enzyme 5-formyl-THF cyclo-ligase (also known as 5,10-methenyltetrahydrofolate synthetase; MTHFS) has a scavenger role as the only enzyme that acts on 5-formyl-THF, converting it to 5,10-methenyl-THF in an ATP- and Mg^{2+}-dependent reaction (Fig. 1*b*). The product is then converted into other reduced folates

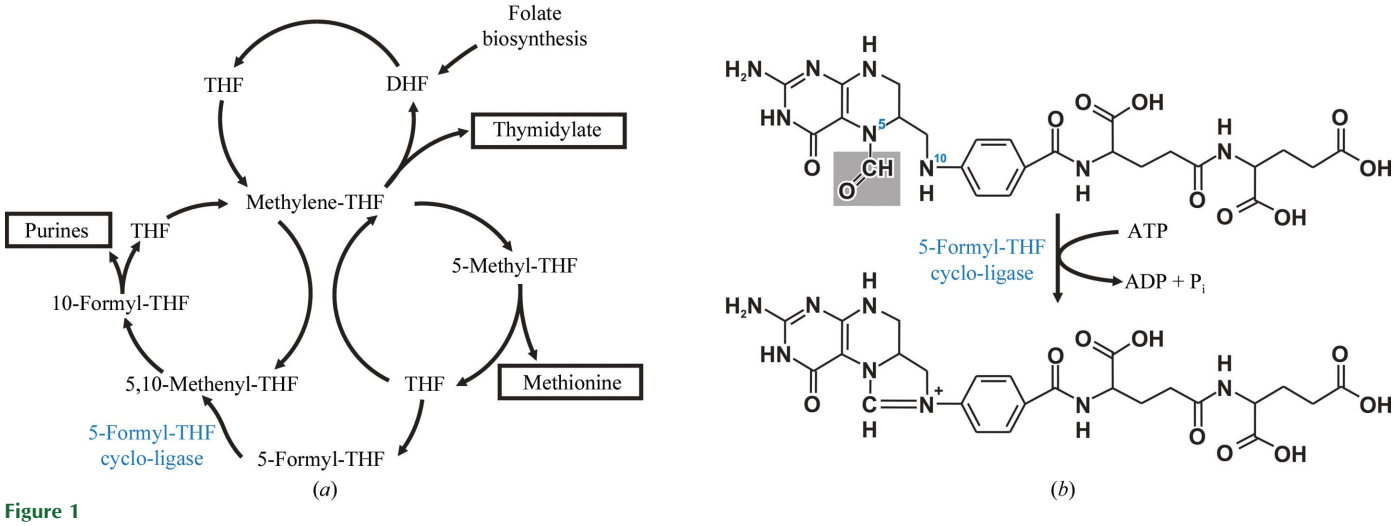


Figure 1
Folate metabolism and the role of 5-formyl-THF cyclo-ligase. (*a*) An overview of bacterial folate metabolism (modified from Zittoun & Zittoun, 1972). Both folate and the one-carbon group it carries can exist in different oxidation states, which are interconvertable. The different forms of folate play a role in several important metabolic pathways, including the biosynthesis of purines, methionine and thymidylate. (*b*) The reaction converting 5-formyl-THF (top) to 5,10-methenyl-THF (bottom) that is catalysed by 5-formyl-THF cyclo-ligase. The metabolically important one-carbon group carried by 5-formyl-THF is highlighted in grey. The catalytically important N5 and N10 positions are labelled.

involved in one-carbon metabolism (Chen *et al.*, 2004). Since 5-formyl-THF acts as an inhibitor of other folate-dependent enzymes, 5-formyl-THF cyclo-ligase may also indirectly regulate several key metabolic processes, such as purine, pyrimidine and amino-acid biosynthesis (Stover & Schirch, 1991; Bertrand & Jolivet, 1989). Since all of these processes are required for cell growth and development, 5-formyl-THF cyclo-ligase has been suggested as a target for drug discovery (Jolivet *et al.*, 1996).

2. Materials and methods

Cloning, expression and protein purification followed standard OPPF pipeline protocols, as described previously (Ren *et al.*, 2005; Alzari *et al.*, 2006). Briefly, the 5-formyl-THF cyclo-ligase gene (BA4489) was amplified from genomic DNA by PCR with the forward primer ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacGTGAGAGAAGAGAAGCTACGTTTACGTAAAC and the reverse primer ggggaccactttgtacaagaaagctgggtctcaCTATACAAGCCCATTTTTAACCATTGTTCC incorporating a ribosome-binding site and an N-terminal hexahistidine tag. The gene was then inserted into the expression vector pDEST14 using Gateway recombinatorial cloning (Invitrogen), resulting in an expressed protein in which the N-terminal methionine residue is replaced by an uncleavable purification tag MAHHHHHHV–. The expression vector was transformed into *Escherichia coli* (strain Rosetta pLysS) and expression was induced by the addition of 0.5 m*M* isopropyl *β*-D-thiogalactopyranoside (IPTG). The protein was purified by a combination of Ni–NTA affinity chromatography and gel filtration.

Crystallization trials and optimizations were carried out using the nanodrop crystallization procedure with standard OPPF protocols (Walter *et al.*, 2003, 2005). In the absence of cofactors, 5-formyl-THF cyclo-ligase gave crystals of up to 160 × 160 × 80 µm in size which diffracted to medium resolution (anisotropically to 3.0 Å at best; space group $P4_22_12$; data not shown). However, in the presence of ATP, Mg^{2+} and folinate cofactors, one larger (380 × 220 × 60 µm) and better ordered crystal was obtained. This grew within 5 d in 200 nl sitting drops containing 5.5 mg ml^{-1} purified protein, 3.8 m*M* folinate, 3.8 m*M* ATP, 30 m*M* $MgCl_2$, 12.5% polyethylene glycol 3350, 50 m*M* bis-Tris pH 5.5 equilibrated against a reservoir solution of 25% polyethylene glycol 3350, 100 m*M* bis-Tris pH 5.5. Crystals were cryoprotected in a mixture containing 20% glycerol and 80% reservoir solution and flash-cooled in liquid nitrogen (100 K).

The crystal of 5-formyl-THF cyclo-ligase grown in the presence of ATP and folinate diffracted well, with measurements extending to Bragg spacings of 1.5 Å. Diffraction data were recorded (at the ESRF synchrotron, beamline ID14-EH1) in two passes: a high-resolution pass of 362 images was initially recorded using 0.5° oscillations, 6 s exposures and a crystal-to-detector distance of 114 mm, followed by a low-resolution pass of 113 images using 2° oscillations, 2 s (the first 21 images) or 1 s exposures and a crystal-to-detector distance of 286 mm. However, processing these images using either *DENZO* or *MOSFLM* was initially problematic owing to scaling problems that we attribute to a high mosaic spread (Bahar *et al.*, 2006. In *DENZO*, refining the mosaicity proved unstable and the most plausible processing was obtained using a fixed mosaicity of 1.5°, while with *MOSFLM* the refined mosaicity varied between 1.2 and 4°. As part of a crystallographic workshop, the diffraction data were reprocessed using the program *XDS* (Kabsch, 1993), which yielded an improved data set with a mosaicity of 0.6° (using the *XDS* definition; see Table 1 and Bahar *et al.*, 2006 for more details). Eventually, the structure was solved using the program *MOLREP* (Vagin & Teplyakov, 1997) with the crystal structure of a homologue from *B. subtilis* (Yqgn) as a search model (PDB code 1ydm; ∼40% sequence identity). Based on the molecular-replacement solution, the program *ARP*/*wARP* (Perrakis *et al.*, 1999) was used to automatically build the structure and carry out ligand fitting. Manual rebuilding used *Coot* (Emsley & Cowtan, 2004) and *REFMAC* (Murshudov *et al.*, 1999) was employed to perform TLS refinement (using each chain as a separate body) using isotropic *B*-factor refinement and weak restraints between main-chain atoms of the two chains. Crystallographic statistics are shown in Table 1. The structure was analysed using *PROCHECK* (Laskowski *et al.*, 1993) and *MolProbity* (Lovell *et al.*, 2003). Structure superpositions were calculated using *SHP* (Stuart *et al.*, 1979). Figures were prepared using *BobScript* (Esnouf, 1997) and *PyMOL* (DeLano, 2002).

Table 1
Crystallographic data-collection and processing statistics.

Values in parentheses are for the outer resolution shell.

Data collection	
Experiment type	Single wavelength
Beamline	ESRF ID14-EH1
Wavelength (Å)	0.934
Data processing	
Resolution limits (Å)	20–1.6 (1.69–1.60)
Unique reflections	49797 (7241)
Space group	*P*1
Unit-cell parameters (Å, °)	a = 37.15, b = 45.50, c = 67.00, α = 74.9, β = 78.0, γ = 70.1
Completeness (%)	95.5 (94.3)
Multiplicity	2.5 (2.0)
$I/\sigma(I)$	13.1 (2.9)
$R_{\rm merge}$†	0.042 (0.341)
Refinement statistics	
Resolution limits (Å)	20–1.6 (1.64–1.60)
R factor	0.189 (0.280)
Free *R* factor (random 5% of reflections)	0.238 (0.346)
No. of non-H atoms (protein/cofactor/water)	3168/66/325
R.m.s.d. bond lengths (Å)	0.016
R.m.s.d. bond angles (°)	1.6
Mean *B* value (Å^2)	33.7
Ramachandran plot‡	
Residues in allowed regions (%)	98.6
Residues in disallowed regions (%)	0.6

† $R_{\rm merge} = \sum|I - \langle I\rangle|/\sum\langle I\rangle$. ‡ Calculated using the program *PROCHECK* (Laskowski *et al.*, 1993).

3. Results and discussion

The structure of 5-formyl-THF cyclo-ligase (Fig. 2) forms a single domain with an $\alpha+\beta$ fold: a central layer of mixed (parallel and antiparallel) β-sheet flanked by helices on either side. This fold belongs to the NagB/RpiA/CoA transferase-like superfamily of enzymes (Murzin *et al.*, 1995), the members of which are involved in a diverse range of metabolic processes but share the common property of phosphate binding.

The crystal structure comprises two protein chains in the crystallographic asymmetric unit. Both chain traces are continuous and include parts of the hexahistidine purification tag (the *A* chain extends from residue −4 to 189 and the *B* chain from −1 to 188). The conformations of the two chains are nearly identical [a root-mean-square deviation (r.m.s.d.) between Cα-atom positions of 0.39 Å for residues 1–188 in each chain], but the electron density for the *A* chain is of somewhat better quality and the discussion focuses on this chain.

It has been suggested that 5-formyl-THF cyclo-ligase forms dimers in solution and that this oligomeric state is related to the cooperative binding of its substrate 5-formyl-THF (Chen *et al.*, 2005). While the structure presented here is, in principle, consistent with a dimeric protein, an analysis (using PISA; Krissinel & Henrick, 2005) suggests that no biologically significant interfaces are present in our structure. All the potential interface regions in the crystal consist exclusively of hydrogen bonds and salt bridges and the largest possible interface area comprises only 500 Å^2 (∼5% of the total surface of the protein monomer), which is significantly less than for most biological dimers (Lesk, 2004). However, it is conceivable that dimerization occurs transiently in solution, especially at higher concentrations (Chen *et al.*, 2005).

The structure of *B. anthracis* 5-formyl-THF cyclo-ligase was obtained by cocrystallization with folinate, ATP and Mg^{2+}. The structure shows a pocket in the protein containing a bound nucleotide cofactor (Fig. 3*a*). Interestingly, the ATP molecule which was present in the crystallization experiment has been hydrolysed to ADP and inorganic phosphate. The high-resolution map shows clear and separated electron density for these moieties separated by a water molecule. The conformation of the ADP α- and β-phosphates is stabilized as part of an octahedral Mg^{2+} coordination shell (Fig. 3*b*). There is some weak ill-defined electron density in the region expected to be occupied by the substrate and/or product, but it is not of sufficient quality to allow it to be built. However, ATP/ADP only occupies as small part of the active-site pocket, adjacent to a wider cavity which could accommodate the pteridine ring of folinate. Furthermore, this cavity is lined by a number of hydrogen-bond donors and acceptors, as well as hydrophobic residues, a suitable environment for binding a heterocyclic aromatic ring (Fig. 3*a*).

Our findings support a catalytic mechanism in which 5-formyl-THF interacts directly with the γ-phosphate of the ATP cofactor and thereby allows the formation of the azoline ring in the substrate. Based on NMR-labelling studies, a catalytic mechanism has been proposed in which 5-formyl-THF becomes transiently phosphorylated by ATP to form a phospho-enol intermediate (Chen *et al.*, 2005; Fig. 3*c*), which is entirely consistent with our observations. Based on this scheme, our crystal structure would represent the final stage of the catalytic mechanism, after the enzyme has turned over and released the product.

At present, structures are available for two homologues of *B. anthracis* 5-formyl-THF cyclo-ligase, those from *Mycoplasma*

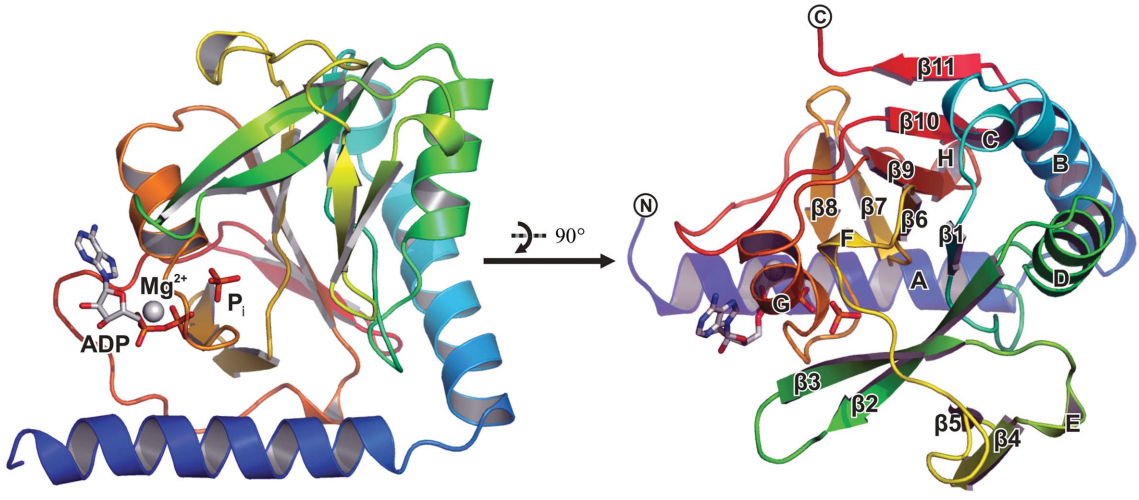


Figure 2
Two orthogonal views of the secondary structure of *B. anthracis* 5-formyl-THF cyclo-ligase (coloured from blue at the N-terminus to red at the C-terminus). Secondary-structural elements are labelled on the right-hand view. The ADP and phosphate cofactors are shown in a stick representation, with the Mg^{2+} ion shown as a grey sphere.

pneumoniae (PDB codes 1sbq, 1u3f and 1u3g; Chen *et al.*, 2004, 2005) and from *Bacillus subtilis* (PDB code 1ydm; unpublished work). Sequence and structural alignments (Fig. 4) reveal only moderate sequence similarities to the *B. anthracis* enzyme (∼25% and ∼40% identity, respectively) but a high degree of structure conservation (r.m.s.d. for Cα-atom superpositions of ∼1.1 Å for 150 and 160 residues, respectively). These findings underscore the critical metabolic function of the enzyme, for which there appears to be strongly conservative evolutionary pressure. The *B. anthracis* and *B. subtilis* sequences both contain a 20 amino-acid insertion compared with the *M. pneumoniae* homologue (residues 93–112 in *B. anthracis* numbering). Structurally, this insertion lines part of the substrate-binding site and, interestingly, it is positioned such that it precludes the substrate-binding mode postulated for the *M. pneumoniae* structure (PDB code 1u3g; Chen *et al.*, 2005). However, we note that in the model the atoms forming the portion of the substrate that impinges on this insertion have been assigned occupancies of zero, suggesting uncertainty in the interpretation. This difference, along with other changes to the amino acids lining the site (particularly Thr50, Met53 and Tyr138; Fig. 4*a*) suggest that despite the structural similarity,

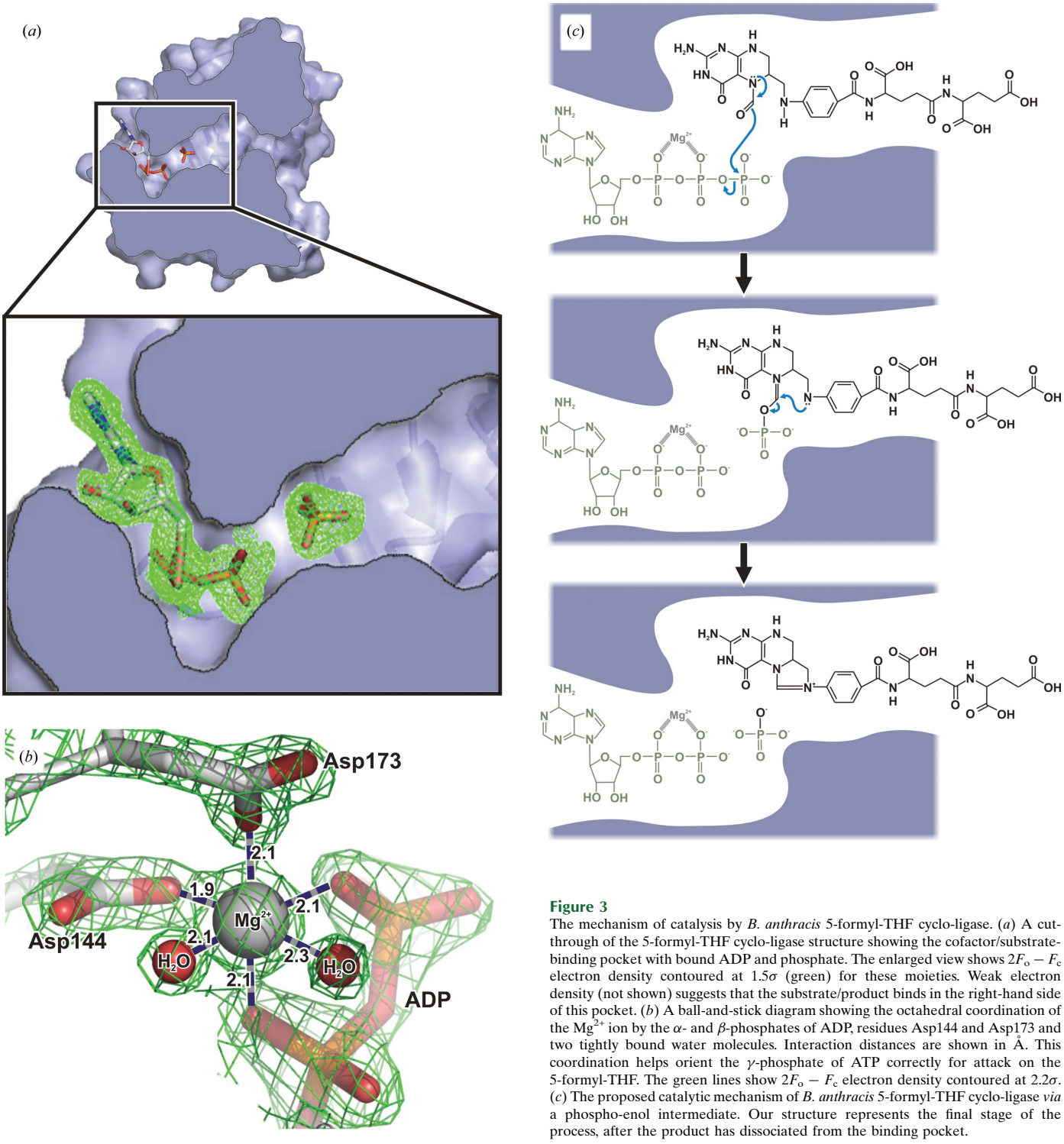


Figure 3
The mechanism of catalysis by *B. anthracis* 5-formyl-THF cyclo-ligase. (*a*) A cut-through of the 5-formyl-THF cyclo-ligase structure showing the cofactor/substrate-binding pocket with bound ADP and phosphate. The enlarged view shows $2F_o - F_c$ electron density contoured at 1.5σ (green) for these moieties. Weak electron density (not shown) suggests that the substrate/product binds in the right-hand side of this pocket. (*b*) A ball-and-stick diagram showing the octahedral coordination of the Mg^{2+} ion by the α- and β-phosphates of ADP, residues Asp144 and Asp173 and two tightly bound water molecules. Interaction distances are shown in Å. This coordination helps orient the γ-phosphate of ATP correctly for attack on the 5-formyl-THF. The green lines show $2F_o - F_c$ electron density contoured at 2.2σ. (*c*) The proposed catalytic mechanism of *B. anthracis* 5-formyl-THF cyclo-ligase *via* a phospho-enol intermediate. Our structure represents the final stage of the process, after the product has dissociated from the binding pocket.

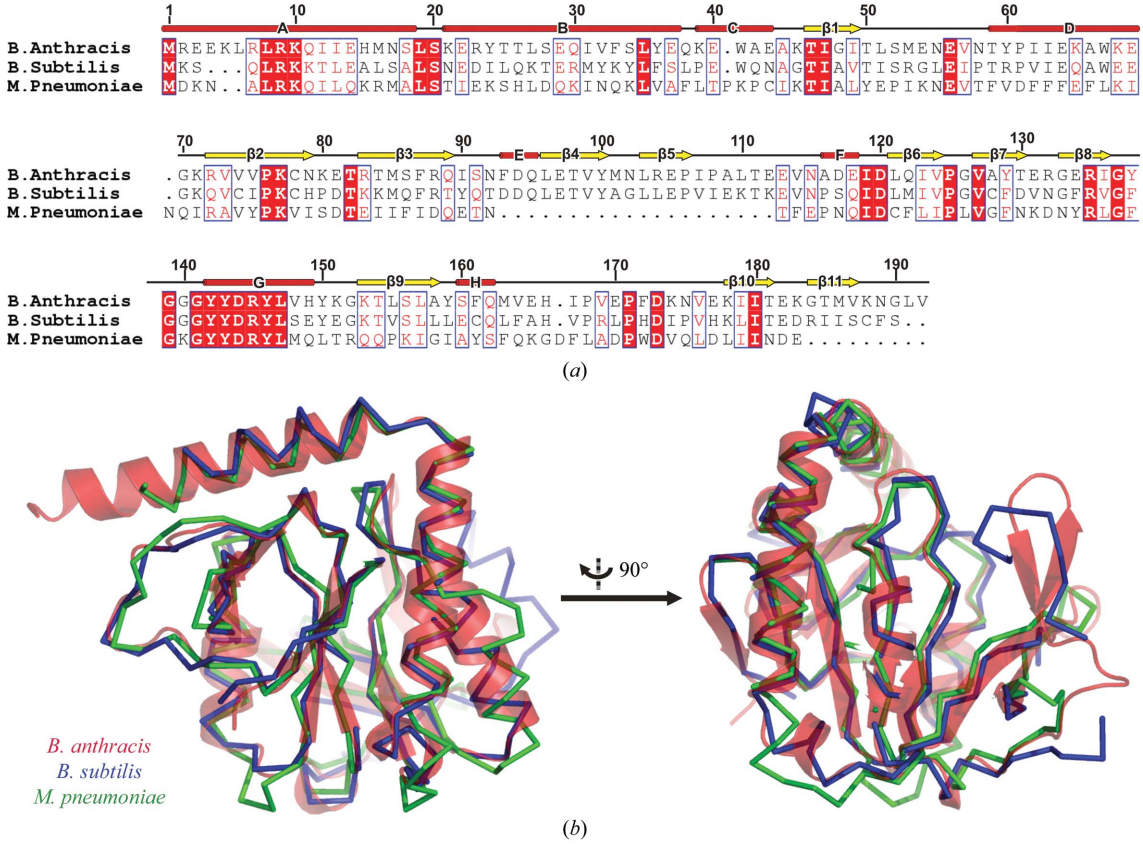


Figure 4
Sequence and structural alignments of prokaryotic 5-formyl-THF cyclo-ligase enzymes. (*a*) Annotated sequence alignment of the *B. anthracis*, *B. subtilis* and *M. pneumoniae* 5-formyl-THF cyclo-ligases. The residue numbering refers to the *B. anthracis* sequence, with secondary-structural assignments shown above the alignment. (*b*) Orthogonal views of the superposition of the corresponding crystal structures (calculated using the program *SHP*; Stuart *et al.*, 1979). The *B. anthracis* structure is shown in a red semitransparent cartoon representation; the crystal structures for the *B. subtilis* (PDB code 1ydm) and *M. pneumoniae* (PDB code 1u3g) enzymes are shown as coloured Cα traces in blue and green, respectively. The overall fold of the protein is tightly conserved, except for the 20 amino-acid insertion at the bottom right of the right-hand figure.

there is potential for designing inhibitors of 5-formyl-THF cyclo-ligase that are specific to particular pathogen species.

The OPPF is funded by the UK Medical Research Council (MRC) and this work was carried out as part of the Structural Proteomics in Europe (SPINE) consortium (European Commission Grant No. QLG2-CT-2002-00988). CM and LGC were supported by Wellcome Trust and MRC studentships, respectively. GW was supported by the BBSRC e-HTPX project, while DIS and RME received additional MRC support. We also thank the staff of the ESRF and EMBL, Grenoble for help with data collection..

References

Alzari, P. M. *et al.* (2006). *Acta Cryst.* D**62**, 1103–1113.
Au, K. *et al.* (2006). *Acta Cryst.* D**62**, 1267–1275.
Bahar, M. *et al.* (2006). *Acta Cryst.* D**62**, 1170–1183.
Bertrand, R. & Jolivet, J. (1989). *J. Biol. Chem.* **264**, 8843–8846.
Chen, S., Shin, D. H., Pufan, R., Kim, R. & Kim, S.-H. (2004). *Proteins*, **56**, 839–843.
Chen, S., Yakunin, A. F., Proudfoot, M., Kim, R. & Kim, S.-H. (2005). *Proteins*, **61**, 433–443.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Esnouf, R. M. (1997). *J. Mol. Graph. Model.* **15**, 132–134.
Jolivet, J., Dayan, A., Beauchemin, M., Chahla, D., Mamo, A. & Bertrand, R. (1996). *Oncologist*, **1**, 248–254.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Krissinel, E. & Henrick, K. (2005). *CompLife 2005*, edited by M. R. Berthold, R. Glen, K. Diederichs, O. Kohlbacher & I. Fischer, pp. 163–174. Berlin/Heidelberg: Springer-Verlag.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lesk, A. (2004). *Introduction to Protein Science: Architecture, Function and Genomics.* Oxford University Press.
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.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *J. Mol. Biol.* **247**, 536–540.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Read, T. D. *et al.* (2003). *Nature (London)*, **423**, 81–86.
Ren, J., Sainsbury, S., Berrow, N. S., Alderton, D., Nettleship, J. E., Stammers, D. K., Saunders, N. J. & Owens, R. J. (2005). *BMC Struct. Biol.* **10**, 13.
Stover, P. & Schirch, V. (1991). *J. Biol. Chem.* **266**, 1543–1550.
Stuart, D. I., Levine, M., Muirhead, H. & Stammers, D. K. (1979). *J. Mol. Biol.* **134**, 109–142.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Walter, T. S., Diprose, J., Brown, J., Pickford, M., Owens, R. J., Stuart, D. I. & Harlos, K. (2003). *J. Appl. Cryst.* **36**, 308–314.
Walter, T. S., Diprose, J. M., Mayo, C. J., Siebold, C., Pickford, M. G., Carter, L., Sutton, G. C., Berrow, N. S., Brown, J., Berry, I. M., Stewart-Jones, G. B., Grimes, J. M., Stammers, D. K., Esnouf, R. M., Jones, E. Y., Owens, R. J., Stuart, D. I. & Harlos, K. (2005). *Acta Cryst.* D**61**, 651–657.
Zittoun, J. & Zittoun, R. (1972). *Rev. Eur. Etud. Clin. Biol.* **17**, 139–146.