research papers

Structures of the hydrolase domain of human 10-formyltetrahydrofolate dehydrogenase and its complex with a substrate analogue

Petri Kursula,[a,b]* **Herwig Schüler,**[a]‡ **Susanne Flodin,**[a] **Petra Nilsson-Ehle,**[a] **Derek J. Ogg,**[a] **Pavel Savitsky,**[c] **Pär Nordlund**[a]* **and Pål Stenmark**[a]

[a]Structural Genomics Consortium, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-17177 Stockholm, Sweden, [b]Department of Biochemistry, FIN-90014 University of Oulu, Oulu, Finland, and [c]Structural Genomics Consortium, University of Oxford, Botnar Research Centre, Oxford OX3 7LD, England

‡ Current address: Max Delbrück Center for Molecular Medicine, D-13125 Berlin-Buch, Germany.

Correspondence e-mail: petri.kursula@oulu.fi, par.nordlund@ki.se

10-Formyltetrahydrofolate dehydrogenase is a ubiquitously expressed enzyme in the human body. It catalyses the formation of tetrahydrofolate and carbon dioxide from 10-formyltetrahydrofolate, thereby playing an important role in the human metabolism of one-carbon units. It is a two-domain protein in which the N-terminal domain hydrolyses 10-formyltetrahydrofolate into formate and tetrahydrofolate. The high-resolution crystal structure of the hydrolase domain from human 10-formyltetrahydrofolate dehydrogenase has been determined in the presence and absence of a substrate analogue. The structures reveal conformational changes of two loops upon ligand binding, while key active-site residues appear to be pre-organized for catalysis prior to substrate binding. Two water molecules in the structures mark the positions of key oxygen moieties in the catalytic reaction and reaction geometries are proposed based on the structural data.

1. Introduction

The enzyme 10-formyltetrahydrofolate dehydrogenase (10-FTHFD) is a highly expressed protein; in several human tissues it constitutes up to 1% of the total cytosolic protein. 10-FTHFD catalyses the overall $NADP^+$-dependent formation of carbon dioxide and tetrahydrofolate (THF) from 10-formyltetrahydrofolate (Min *et al.*, 1988; Schirch *et al.*, 1994). The N-terminal domain of 10-FTHFD catalyses a hydrolase reaction resulting in the cleavage of the 10-formyl group to generate formate and THF (Krupenko *et al.*, 1997). The C-terminal aldehyde dehydrogenase domain, on the other hand, is required for the $NADP^+$-dependent dehydrogenase reaction (Krupenko *et al.*, 1997). The N-terminal hydrolase domain has homology to folate-binding domains from other enzymes, such as glycinamide ribonucleotide formyltransferase (GART) and methionyl-tRNA formyltransferase (FMT). His106 is a fully conserved residue in the enzyme family and has been shown to be important both structurally and catalytically (Krupenko *et al.*, 2001). Asp142 is a second conserved residue that has been shown to be directly involved in catalysis (Krupenko & Wagner, 1999).

The role of folate and folate-dependent enzymes has been specifically studied with regard to nervous system development. Based on its expression pattern, 10-FTHFD has been suggested to play a role in CNS development, especially in glial cell function (Neymeyer & Tephly, 1994; Neymeyer *et al.*, 1997). 10-FTHFD also plays a role in the detoxification of methanol *via* the formate pathway (Johlin *et al.*, 1987, 1989; Cook *et al.*, 2001) and 10-FTHFD deficiency in mice causes problems with pathways involving one-carbon units, such as histidine catabolism (Cook, 2001).

Rapidly dividing cells, such as cancer cells, are dependent on a high availability of nucleotides and are hence sensitive to fluctuations in the 10-formyltetrahydrofolate pool, an important precursor in nucleotide biosynthesis. Two steps in purine *de novo* biosynthesis require 10-formyl-THF and C atoms C2 and C8 of the purine ring originate from 10-formyl-THF. In cancer cells, 10-FTHFD is downregulated (Krupenko & Oleinik, 2002), while overexpression induces cell-cycle arrest and apoptosis (Oleinik & Krupenko, 2003), as well as p53 activation (Oleinik *et al.*, 2005).

In order to obtain detailed structural information on this abundant enzyme central to human one-carbon unit metabolism, we have determined high-resolution crystal structures of the N-terminal hydrolase domain of human 10-FTHFD both in the presence and absence of the substrate analogue 6-formyltetrahydropterin.

2. Materials and methods

2.1. Cloning, expression and purification

The cDNA for the hydrolase domain of human 10-FTHFD was cloned into the pNIC-Bsa4 vector and the predicted protein contained an N-terminal hexahistidine tag with an integrated TEV protease cleavage site (MHHHHHHSSGV-DLGTENLYFQ). Protein expression was carried out in *Escherichia coli* BL21(DE3) using IPTG induction at 291 K overnight. Cells were harvested by centrifugation and pellets were resuspended in 50 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol. 50 μg ml^{-1} lysozyme was added as well as one tablet of Complete EDTA-free protease inhibitors (Roche) per cell pellet. Cells were disrupted by sonication and DNA was precipitated by adding polyethyleneimine to 0.15%. The sample was incubated on ice for 30 min and centrifuged for 1 h at 40 000g. The soluble fraction was filtered and subjected to further purification.

Purification was conducted on an ÄKTA Xpress system. HisTrap HP and Superdex 75 columns were equilibrated with IMAC buffer 1 (50 mM HEPES pH 7.5, 10 mM imidazole, 500 mM NaCl, 10% glycerol, 0.5 mM TCEP) and gel-filtration buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 10% glycerol, 0.5 mM TCEP), respectively. The protein sample was loaded on the HisTrap HP column, which was washed with IMAC buffer 1 followed by IMAC buffer 2 (50 mM HEPES pH 7.5, 50 mM imidazole, 500 mM NaCl, 10% glycerol, 0.5 mM TCEP). Bound protein was eluted from the IMAC columns with 7.5 ml IMAC elution buffer (50 mM HEPES pH 7.5, 400 mM imidazole, 500 mM NaCl, 10% glycerol, 0.5 mM TCEP) and loaded onto the gel-filtration column. The chromatogram from gel filtration showed one major protein peak of 10-FTHFD of high purity as shown by SDS–PAGE analysis. 2 mM TCEP was added to the pooled protein. The protein was concentrated to 33 mg ml^{-1} and stored at 193 K.

2.2. Crystallization, data collection and structure solution

Crystals were obtained using the sitting-drop method at 293 K. Drops were prepared using 900 nl protein solution

Table 1
Data processing and refinement.

	Apo	Complex
Space group	$C2$	$C2$
Unit-cell parameters (Å, °)	a = 108.2, b = 64.5, c = 58.9, α = 90, β = 98.2, γ = 90	a = 108.1, b = 64.4, c = 58.9, α = 90, β = 98.8, γ = 90
Resolution range (Å)	41.7–1.70 (1.79–1.70)	20–1.85 (1.90–1.85)
R_{sym} (%)	9.0 (36.8)	6.7 (40.8)
$\langle I/\sigma(I)\rangle$	12.3 (3.2)	14.3 (4.0)
Completeness (%)	99.6 (99.7)	99.6 (99.5)
Redundancy	3.7 (3.7)	4.8 (4.5)
R_{cryst} (%)	17.1	17.8
R_{free} (%)	20.0	22.5
R.m.s.d. bond lengths (Å)	0.016	0.014
R.m.s.d. angles (°)	1.6	1.6
Ramachandran analysis		
Most favoured (%)	91.1	89.5
Additionally allowed (%)	8.9	9.7
Disallowed (%)	0	0.8

(16.5 mg ml^{-1} concentration) and 900 nl well solution [1.4 M (NH$_4$)$_2$SO$_4$, 50 mM HEPES pH 7.8]. A liganded complex was prepared by soaking a crystal in a 5 mM solution of THF and 20 mM formate for 1 h. Prior to cryocooling in liquid nitrogen, the crystals were transferred into a cryoprotectant solution [2.3 M (NH$_4$)$_2$SO$_4$, 20% glycerol, 0.2 M NaCl, 2 mM TCEP, 20 mM HEPES pH 7.5, 50 mM HEPES pH 7.8].

The crystals diffracted to 1.7 Å using synchrotron radiation. Data were collected at ESRF ID14-EH4 and BESSY 14.1 and processed with *MOSFLM* (Leslie, 1992) and *SCALA* (Evans, 2006) for the apo form and *XDS/XDSi* (Kabsch, 1993; Kursula, 2004) for the liganded form. The data-processing statistics are presented in Table 1.

The structure was solved by molecular replacement, using the structure of the corresponding enzyme from rat (PDB code 1s3i) as a model (Chumanevich *et al.*, 2004). Model building was performed in *Coot* (Emsley & Cowtan, 2004) and refinement in *REFMAC*5 (Murshudov *et al.*, 1997) using TLS parameters (Winn *et al.*, 2001). Water molecules were added using *ARP/wARP* (Perrakis *et al.*, 1999). The structure of the liganded form was solved using the apo form as a model and refined as above. The refinement statistics are presented in Table 1.

The coordinates and structure factors for the structures presented in this paper were deposited in the PDB under codes 2bw0 (apo) and 2cfi (complex). The figures were produced using *MOLSCRIPT/POVSCRIPT+* (Fenn *et al.*, 2003), *POV-Ray* (http://www.pov-ray.org), *UCSF Chimera* (Pettersen *et al.*, 2004), *DINO* (http://www.dino3d.org) and *ESPript* (Gouet *et al.*, 1999).

3. Results and discussion

3.1. The overall structure of human 10-FTHFD hydrolase domain

The structure of the hydrolase domain of human 10-FTHFD was determined at 1.7 Å resolution with good geometrical quality and crystallographic statistics (Fig. 1*a*;

Table 1). The final refined model consists of residues 1–307 of human 10-FTHFD and two N-terminal residues from the affinity tag. Three regions of high temperature factors were observed, corresponding to loops around residues 35, 60 and 85.

The hydrolase domain can further be divided into the N- and C-terminal subdomains which are also found in FMT

(Figs. 1a and 1b). The N-terminal folate-binding domain (1–186), consisting of a seven-stranded β-sheet and four α-helices, can be further divided into two halves consisting of residues before and after the α4–β5 loop. This division can be made based on sequence and structural conservation (see below). The C-terminal domain, also found in FMT (Chumanevich et al., 2004), folds into a slightly open β-barrel
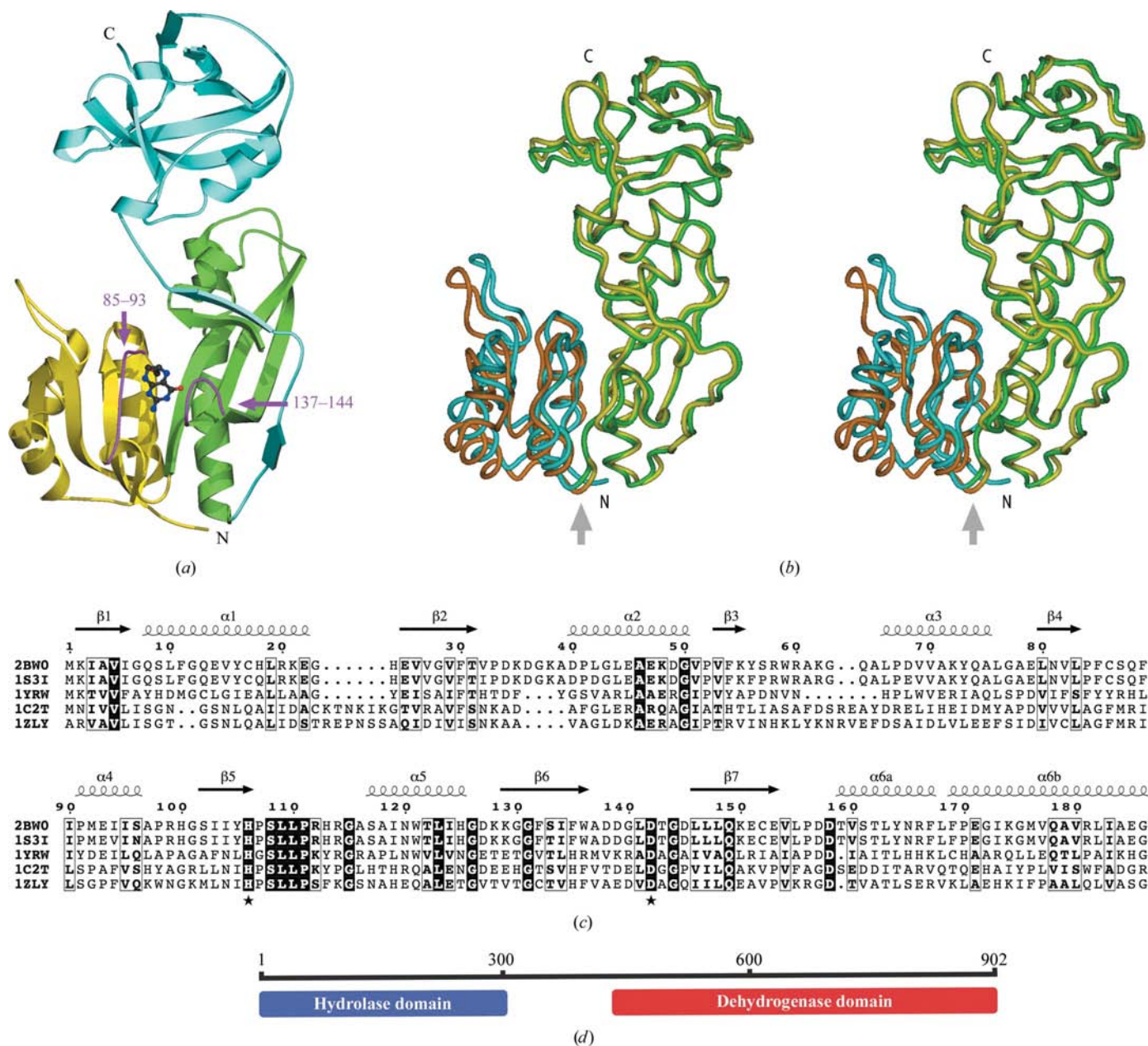

Figure 1
(a) Overall structure of human 10-FTHFD. The ligand 6-formyltetrahydropterin is shown as a ball-and-stick model bound to the THF-binding site in the N-terminal domain. The two halves of the folate-binding domain are shown in yellow and green and the two loops covering the binding site in magenta. The C-terminal domain is shown in light blue. (b) Superposition of the human 10-FTHFD with the rat enzyme. The human enzyme is coloured green/blue and the rat enzyme yellow/orange. N-terminal residues up to residue 99 are coloured blue (human) and orange (rat) to illustrate the apparent hinge movement between this region and the rest of the protein. The loop segment acting as a hinge is highlighted by the grey arrow. (c) Structural alignment of human (this study) and rat (Chumanevich et al., 2004) 10-FTHFD, E. coli FMT (Gatzeva-Topalova et al., 2005) and human (Dahms et al., 2005) and E. coli (Greasley et al., 1999) GART. The alignment was produced with the SSM server based on structural superposition of the coordinates and includes residues of the folate-binding domain (residues 1–186 in 10-FTHFD). Fully conserved residues are shown on a black background and the secondary-structure elements of human 10-FTHFD are shown above the alignment. Asterisks mark the catalytic residues His106 and Asp142. (d) The domain structure of human 10-FTHFD.
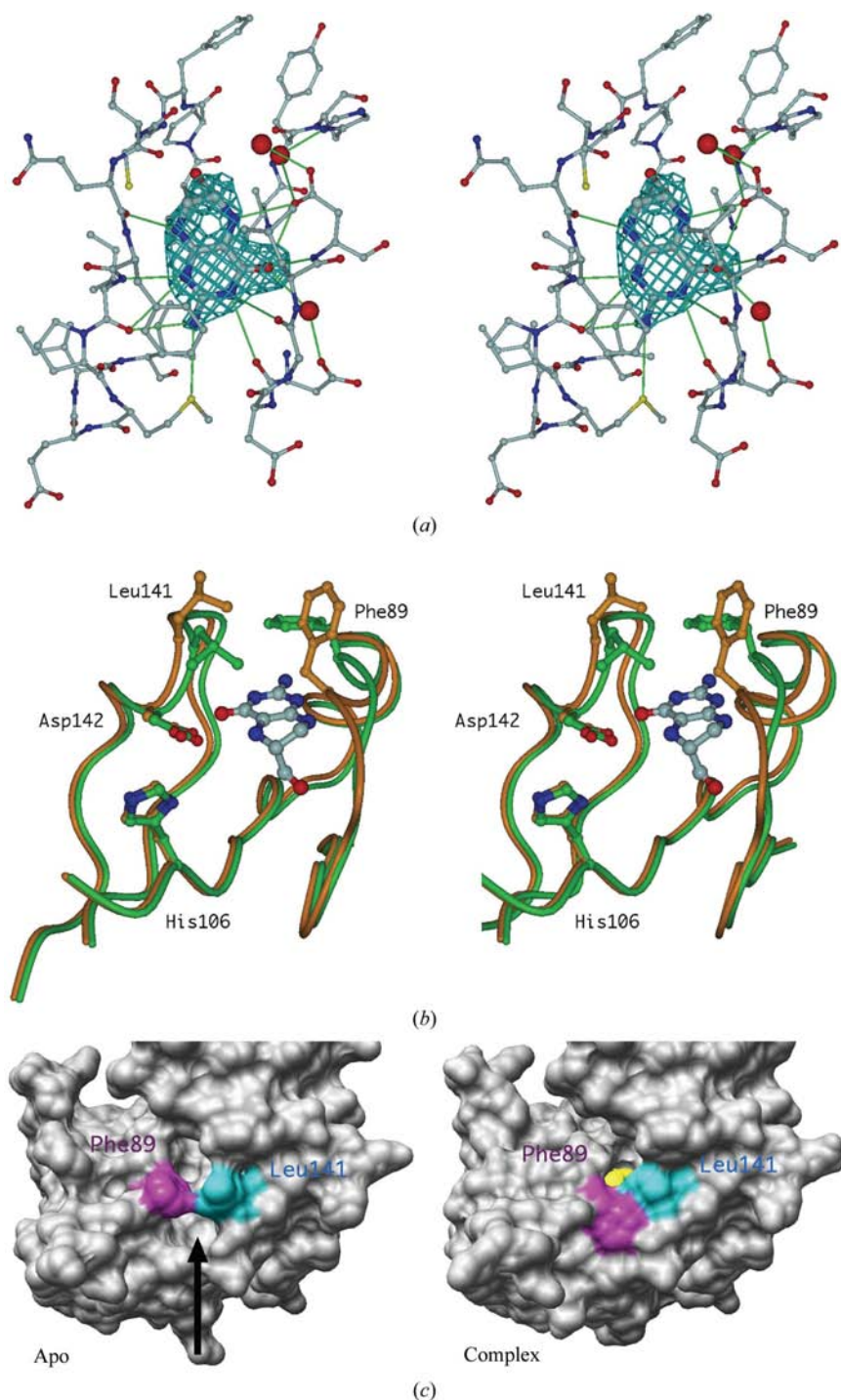
Figure 2
(a) Binding mode of the ligand in 10-FTHFD. The final refined coordinates for the complex are shown and a difference electron-density map after refinement in the absence of the 6-formyltetrahydropterin ligand (contour level 2.5σ) is superimposed. Polar interactions are indicated by green lines. (b) Conformational changes in 10-FTHFD upon active-site ligand binding. While the conformations of the catalytic residues are identical between the apo (orange) and liganded (green) forms, slight movements of the two loops lining the substrate-binding cavity occur. Specifically, both Leu141 and Phe89 swing in to interact with the pteridyl group *via* van der Waals forces. (c) Comparison of the surfaces of the apo (left) and liganded (right) structures. Phe89 and Leu141 are indicated by magenta and cyan colours, respectively, and the 6-formyltetrahydropterin surface is shown in yellow. A small loop movement and a change in the side-chain conformations of Phe89 and Leu141 results in the sealing of the surface gap (arrow) seen in the apo structure.

similar in topology to the core of RNase T1 (Gohda *et al.*, 1994).

A search for closest structural homologues was carried out using *SSM* at EBI (Krissinel & Henrick, 2004). As expected, the detected homologues represent other folate-binding enzymes, including 10-FTHFD from rat, GART and FMT. A structural alignment (Fig. 1c) of the sequences indicate several fully conserved residues that mostly lie in the vicinity of the active site.

3.2. Comparison to the structure of rat 10-FTHFD

The overall fold of human 10-FTHFD is very similar to that of the rat enzyme, determined previously to a resolution of 2.3 Å, as expected from the high sequence identity (Fig. 1b). The superposition of 301 common C^α atoms results in an r.m.s. deviation in their positions of 1.56 Å. Superposition indicates a hinge region approximately at residue His100, in the $\alpha4$–$\beta5$ loop, that causes an apparent opening/closing motion between the first 100 residues and the rest of the molecule when comparing the human and rat enzyme structures. Interestingly, the active site lies in the cleft between the first 100 residues and the rest of the protein. This observation suggests that there could be a larger scale (when compared with the crystal state) opening/closing event taking place in solution, *e.g.* upon substrate binding. Furthermore, the structure-based alignment of the folate-binding domains from 10-FTHFD, GART and FMT (Fig. 1c) clearly indicates that the first 100 residues are less conserved throughout the family than residues 106–186, which form the second half of the folate-binding domain, which is consistent with the fact that the catalytically important residues lie in the second half of the domain.

In the rat 10-FTHFD structure, cysteine residues had been modified by β-mercaptoethanol; in the structures presented here, all cysteines are unmodified. In the rat structure, a β-mercaptoethanol molecule was inserted in the immediate vicinity of the catalytic residues Asp142 and His106 and it was suggested to structurally mimic the reaction product formate. In the human 10-FTHFD structure, two water molecules replace this β-mercaptoethanol and possibly indicate the positions of the two oxygen

moieties taking part in the reaction, an activated water molecule and the carbonyl O atom of the formyl group. This is also suggested by comparisons to ligand complexes of GART (see below).

3.3. Binding mode of the substrate fragment

In order to obtain the structure of a 10-FTHFD–product complex, a crystal was soaked in THF and the structure was solved at 1.85 Å resolution. A refined model with good geometry was obtained containing residues 1–307 of 10-FTHFD and three additional N-terminal residues. During refinement, it became apparent that a fragment of THF was present in the active site, most probably formed via the well characterized cleavage of the C9—N10 bond owing to air oxidation or acid (Reed & Archer, 1980) (Fig. 2). Thus, the expected product of this reaction, 6-formyltetrahydropterin, was modelled into the strong difference density in the active-site pocket. The ligand is oriented such that for THF binding

N10 would be in the immediate vicinity of the catalytic residues Asp142 and His106.

The bottom of the ligand-binding site is formed by the hydrophobic residues Leu83 (strand β4), Ile95 (helix α4) and Ile104 (strand β5); all other ligand–protein interactions are essentially formed by the two loops covering the binding site. The only protein side chain making a hydrogen bond to the ligand is that of the catalytic residue Asp142. Side chains making van der Waals contacts to the ligand include Leu83, Cys86, Phe89, Ile90, Met92, Ile95, Ile104, Phe135, Ala137 and Leu141.

Comparing the apo and liganded structures, small but significant differences are seen, especially in the β4-α4 loop covering the active site (Figs. 2b and 2c). This loop is rather flexible in both the apo and liganded structures, as judged from refined temperature factors, but the magnitude of the conformational change between the two structures is still significant. Interactions are also formed between this loop and the ligand; all of the hydrogen-bonding interactions are made between the ligand and the loop backbone.

3.4. Implications for the reaction mechanism

The reaction mechanism for the 10-FTHFD hydrolase step has been suggested to proceed via a water molecule that has been activated by Asp142 (Chumanevich et al., 2004). This water molecule then breaks the bond between N10 and the carbonyl C atom of the formyl group; His106 helps in orienting the carbonyl O atom and thus provides an oxyanion hole for the reaction. A superposition of our structure with that of E. coli GART complexed with a substrate analogue (Greasley et al., 1999) indicates that two well ordered water molecules in the active site of 10-FTHFD point out the positions of the reactive O atoms during the hydrolase reaction (Fig. 3). The activated water molecule corresponds to Wat227 and Wat266 mimics the position of the carbonyl O atom of the formyl group. It is of note that both His106 and Asp142 have identical conformations in the liganded and unliganded structures, implying that the reaction path is unlikely to involve conformational changes of these residues.

The intermediate in the reaction consists of a positively charged quaternary amine at N10 and a tetrahedral C atom at the formyl group, one of the O atoms being negatively
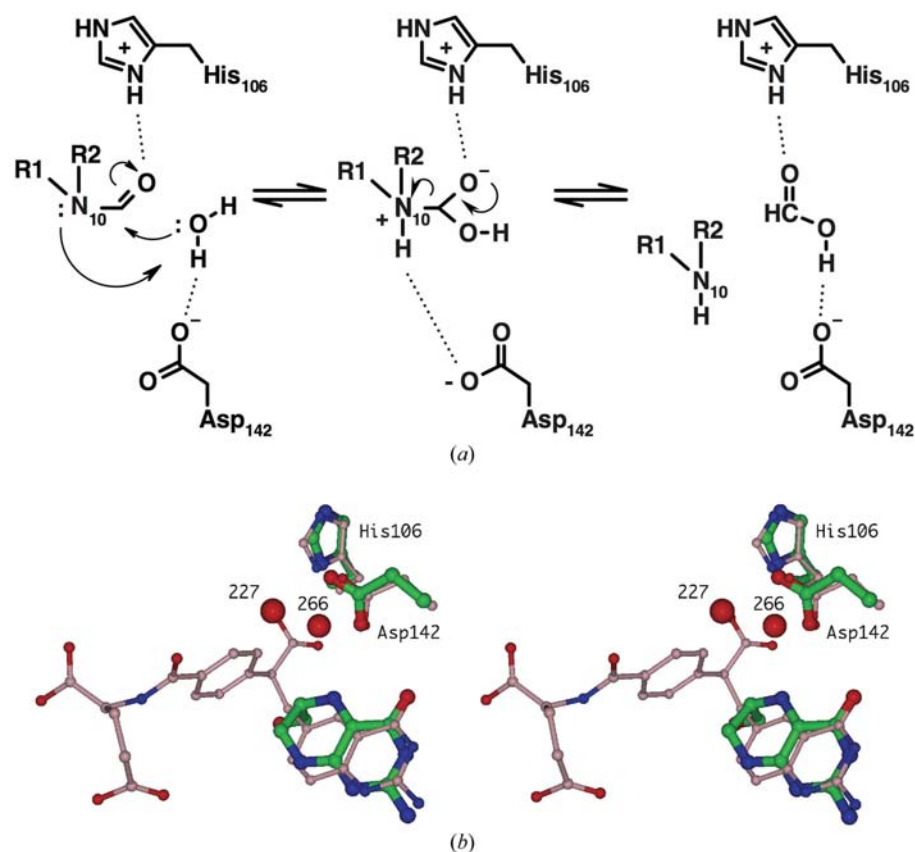

Figure 3
(a) The suggested reaction mechanism for the 10-FTHFD hydrolase step. For clarity, only N10 and the formyl group of the 10-formyl-THF substrate are shown. The water activated by Asp142 attacks the carbon of the formyl group, generating a tetrahedral intermediate stabilized on the one hand by His106 and on the other hand by Asp142. The end products of this reaction are THF and formate. (b) A comparison of the liganded 10-FTHFD (green C atoms) and human GART (pink C atoms) complexed with a substrate analogue, 10-formyl-5,8,10-trideazafolate. His106 and Asp142 are in an identical conformation and two water molecules in the 10-FTHFD complex superimposing with the O atoms of the hydrated formyl group, Wat227 and Wat266, most likely correspond to the positions of the two reactive O atoms during the hydrolase reaction.

charged. Our structure is in line with earlier structures of related enzymes, indicating that His106 stabilizes the negative moiety of the intermediate, while Asp142 does the same for the positive N10.

4. Concluding remarks

Our structures indicate that the catalytic residues in the 10-FTHFD active site are pre-organized prior to substrate binding and that the flexibility of the two loops covering the THF-binding site is important for substrate binding. The positions of two well defined water molecules in the catalytic centre correspond to the two reactive O atoms during the hydrolase reaction. Our data provide a high-resolution view of the structure of a highly abundant human enzyme central for one-carbon unit metabolism, which links together several important metabolic pathways. The structures also have implications for future research addressing the role of 10-FTHFD in tumour progression and methanol poisoning.

References

Chumanevich, A. A., Krupenko, S. A. & Davies, C. (2004). *J. Biol. Chem.* **279**, 14355–14364.

Cook, R. J. (2001). *Arch. Biochem. Biophys.* **392**, 226–232.

Cook, R. J., Champion, K. M. & Giometti, C. S. (2001). *Arch. Biochem. Biophys.* **393**, 192–198.

Dahms, T. E., Sainz, G., Giroux, E. L., Caperelli, C. A. & Smith, J. L. (2005). *Biochemistry*, **44**, 9841–9850.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

Fenn, T. D., Ringe, D. & Petsko, G. A. (2003). *J. Appl. Cryst.* **36**, 944–947.

Gatzeva-Topalova, P. Z., May, A. P. & Sousa, M. C. (2005). *Biochemistry*, **44**, 5328–5338.

Gohda, K., Oka, K., Tomita, K. & Hakoshima, T. (1994). *J. Biol. Chem.* **269**, 17531–17536.

Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.

Greasley, S. E., Yamashita, M. M., Cai, H., Benkovic, S. J., Boger, D. L. & Wilson, I. A. (1999). *Biochemistry*, **38**, 16783–16793.

Johlin, F. C., Fortman, C. S., Nghiem, D. D. & Tephly, T. R. (1987). *Mol. Pharmacol.* **31**, 557–561.

Johlin, F. C., Swain, E., Smith, C. & Tephly, T. R. (1989). *Mol. Pharmacol.* **35**, 745–750.

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

Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.

Krupenko, S. A. & Oleinik, N. V. (2002). *Cell Growth Differ.* **13**, 227–236.

Krupenko, S. A., Vlasov, A. P. & Wagner, C. (2001). *J. Biol. Chem.* **276**, 24030–24037.

Krupenko, S. A. & Wagner, C. (1999). *J. Biol. Chem.* **274**, 35777–35784.

Krupenko, S. A., Wagner, C. & Cook, R. J. (1997). *J. Biol. Chem.* **272**, 10273–10278.

Kursula, P. (2004). *J. Appl. Cryst.* **37**, 347–348.

Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.

Min, H., Shane, B. & Stokstad, E. L. (1988). *Biochim. Biophys. Acta*, **967**, 348–353.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Neymeyer, V. R. & Tephly, T. R. (1994). *Life Sci.* **54**, PL395–PL399.

Neymeyer, V., Tephly, T. R. & Miller, M. W. (1997). *Brain Res.* **766**, 195–204.

Oleinik, N. V., Krupenko, N. I., Priest, D. G. & Krupenko, S. A. (2005). *Biochem. J.* **391**, 503–511.

Oleinik, N. V. & Krupenko, S. A. (2003). *Mol. Cancer Res.* **1**, 577–588.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.

Reed, L. S. & Archer, M. C. (1980). *J. Agric. Food Chem.* **28**, 801–805.

Schirch, D., Villar, E., Maras, B., Barra, D. & Schirch, V. (1994). *J. Biol. Chem.* **269**, 24728–24735.

Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.