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

Structure of the ThDP-dependent enzyme benzaldehyde lyase refined to 1.65 Å resolution

Andy Maraite,[a]‡ Thomas Schmidt,[a] Marion B. Ansörge-Schumacher,[a]‡ A. Marek Brzozowski[b] and Gideon Grogan[b]*

[a]Department of Biotechnology, Faculty of Natural Sciences, RWTH Aachen University, Worringerweg 1, 52074 Aachen, Germany, and [b]Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5YW, England

‡ Present address: Institute Of Chemistry, TU Berlin, Strasse Des 17 Juni 124, 10623 Berlin, Germany.

Correspondence e-mail: grogan@ysbl.york.ac.uk

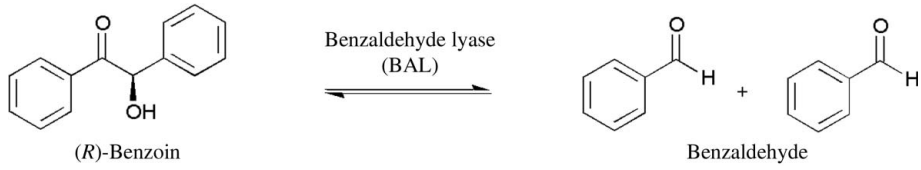
Received 8 June 2007
Accepted 11 June 2007

PDB Reference: benzaldehyde lyase, 2uz1, r2uz1sf.

Benzaldehyde lyase (BAL; EC 4.1.2.38) is a thiamine diphosphate (ThDP) dependent enzyme that catalyses the enantioselective carboligation of two molecules of benzaldehyde to form (*R*)-benzoin. BAL has hence aroused interest for its potential in the industrial synthesis of optically active benzoins and derivatives. The structure of BAL was previously solved to a resolution of 2.6 Å using MAD experiments on a selenomethionine derivative [Mosbacher *et al.* (2005), *FEBS J.* **272**, 6067–6076]. In this communication of parallel studies, BAL was crystallized in an alternative space group ($P2_12_12_1$) and its structure refined to a resolution of 1.65 Å, allowing detailed observation of the water structure, active-site interactions with ThDP and also the electron density for the co-solvent 2-methyl-2,4-pentanediol (MPD) at hydrophobic patches of the enzyme surface.

1. Introduction

A number of lyases capable of the formation and/or cleavage of carbon–carbon bonds are becoming increasingly interesting from the perspective of biotechnology, as they may be used as biocatalysts for the preparation of either optically active intermediates (Grogan *et al.*, 2001) or natural equivalent compounds, such as vanillin (Leonard *et al.*, 2006), for the food and cosmetic industries. C—C bond lyase activity has been attributed to both cofactor-independent enzymes and those requiring cofactors such as coenzyme A (CoA) and thiamine diphosphate (ThDP), the latter cofactor being found in lyases such as acetolactate synthase (Pang *et al.*, 2004) and benzaldehyde lyase (BAL). BAL was first described in strains of *Pseudomonas* that were able to grow on benzoin as the sole carbon source and catalyzes the reversible cleavage of the substrate to yield two molecules of benzaldehyde (Gonzalez & Vicuňa, 1989; Fig. 1). If incubated with an excess of benzaldehyde, BAL also catalyses the formation of the industrial pharmaceutical intermediate (*R*)-benzoin by enantioselective carbon–carbon bond ligation (Demir *et al.*, 2002). A range of other aldehyde substrates for BAL have recently been identified (Sanchez-Gonzalez & Rosazza, 2003; Demir *et al.*, 2003) and there is an increasing number of reports of the optimization of its activity for possible application of the enzyme in industrial processes (Stilger *et al.*, 2006). As part of our continuing studies into the structural enzymology of unusual C—C bond lyases using X-ray crystallography (Leonard & Grogan, 2004; Leonard *et al.*, 2006), we were interested in the molecular structural determinants of both catalysis and enantioselectivity in the reaction catalysed by BAL and we therefore crystallized the protein. Soon after our acquisition of a data set to a maximum resolution of 1.65 Å, the X-ray structure of BAL at a resolution of 2.6 Å was published by Schulz and coworkers (Mosbacher *et al.*, 2005). The Schulz paper reported the structure of BAL



(*R*)-Benzoin Benzaldehyde

Figure 1
The reaction catalyzed by benzaldehyde lyase.

in complex with the cofactor ThDP and used this to generate a computational model of the ligand (*R*)-benzoin in the active site. We used our 1.65 Å data set in conjunction with the available coordinates from the previous solution to produce a higher resolution structure of BAL that, in addition to revealing the details of active-site interactions with the well defined cofactor, shows electron density for the cocrystallization agent 2-methyl-2,4-pentanediol at hydrophobic regions of the enzyme surface.

2. Materials and methods

2.1. Protein production and crystallization

The gene encoding BAL from *P. fluorescens* Biovar I was expressed from plasmid pBAL[His_6] (Iding *et al.*, 1998). This was transformed into *Escherichia coli* BL21 (DE3) cells and maintained on agar plates containing 30 µg ml^{-1} ampicillin. A single colony was used to inoculate a 5 ml starter culture in Luria–Bertani (LB) broth containing 30 µg ml^{-1} ampicillin and this was grown on an orbital shaker at 150 rev min^{-1} overnight at 310 K. The culture was then used to inoculate 500 ml of the equivalent growth medium in a 2 l flask and the culture was grown at 310 K until the optical density at 500 nm had reached 0.5. At this point, 1 m*M* isopropyl *β*-thiogalactopyranoside (IPTG) was added and the culture was grown at 293 K overnight. The cells from 4 × 500 ml were harvested and resuspended together in 100 ml 50 m*M* Tris–HCl buffer pH 8.0 with 300 m*M* sodium chloride and disrupted using ultrasonication with three bursts of 30 s with 1 min intervals at 277 K. The suspension was centrifuged to remove cell debris and the supernatant was filtered through a 2 µm filter (Amicon), loaded onto a nickel-affinity HiTrap agarose column and eluted using a gradient of 0–500 m*M* imidazole in buffer. The combined fractions containing BAL were then applied onto a Sephadex S-200 gel-filtration column (Amersham), from which pure homogeneous BAL was obtained for crystallization trials.

Initial crystallization trials were performed using the Clear Strategy Screen conditions (Brzozowski & Walton, 2000) in 96-well plate format, using crystal drop volumes of 150 nl protein solution plus 150 nl precipitant reservoir solution dispensed by a Mosquito robot. Needles were obtained in drops that contained as reservoir 0.1 *M* Tris–HCl pH 7.5, 0.2 *M* potassium isothiocyanate, 10%(*w*/*v*) polyethylene glycol 1000 and 10% polyethylene glycol 8000. Focused screens in 24-well Linbro dishes with 1 + 1 µl crystal drops did not improve the quality of these needles until it was observed that high concentrations of 2-methyl-2,4-pentanediol resulted in more three-dimensional crystals. The crystals used for X-ray diffraction experiments were grown in 0.1 *M* Tris–HCl pH 7.5, 0.2 *M* potassium isothiocyanate, 10%(*w*/*v*) polyethylene glycol 1000, 10% polyethylene glycol 8000 and 45%(*v*/*v*) 2-methylpentane-2–4-diol; the protein concentration was 5 mg ml^{-1} with a threefold stoichiometric excess of thiamine diphosphate (ThDP). Crystals were flash-cooled in the mother liquor using liquid nitrogen prior to mounting them in the X-ray apparatus.

2.2. X-ray data and refinement

The data were collected at the European Synchrotron Radiation Facility, Grenoble on beamline ID-14-EH1 and were processed using *DENZO* (Otwinowski & Minor, 1997) and scaled using *SCALEPACK*. The structure was solved using the 2.6 Å coordinates (PDB code 2ag0) as the model for molecular replacement using *MOLREP* (Vagin & Teplyakov, 1997). The structure was refined using *REFMAC* (Murshudov *et al.*, 1997). 5% of the total reflections were flagged for cross-validation before refinement. These data were used to monitor the modelling process at various stages of refinement for the weighting of geometrical and temperature-factor restraints. All computing was undertaken using the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). The model building and refinement were carried out with *REFMAC* in conjunction with *ARP*/*wARP* (Perrakis *et al.*, 1999) in the whole (50–1.65 Å) resolution range. *Coot* (Emsley & Cowtan, 2004) was used for manual corrections to the model. The final R_{cryst} and R_{free} were 0.167 and 0.193, respectively, with 92.6% residues in the most favoured regions, 7.2% in additional allowed regions and one residue in each subunit in generously allowed regions as indicated by *PROCHECK* (Laskowski *et al.*, 1993). Data-collection and refinement statistics are given in Table 1.

Table 1
Data-collection and refinement statistics for BAL.

Values in parentheses are for the highest resolution shell.

Beamline	ID-14-EH1
Wavelength (Å)	0.97630
Resolution (Å)	50.0–1.65 (1.68–1.65)
Space group	$P2_12_12_1$
Unit-cell parameters (Å, °)	a = 104.5, b = 121.6, c = 163.0, $\alpha = \beta = \gamma$ = 90
Unique reflections	249612 (12364)
Completeness (%)	99.9 (100)
R_{sym}† (%)	13.1 (87.7)
Multiplicity	5.1 (5.0)
$\langle I/\sigma(I)\rangle$‡	5.5 (1.5)
No. of protein atoms	16430
No. of solvent waters	1918
R_{cryst}	0.167
R_{free}	0.193
R.m.s.d. 1–2 bonds (Å)	0.011
R.m.s.d. 1–3 angles (Å)	1.270
Average main-chain B (Å^2)	11
Average side-chain B (Å^2)	15
Average solvent B (Å^2)	15

† $R_{sym} = \sum_{\mathbf{h}}\sum_l |I_{\mathbf{h}l} - \langle I_{\mathbf{h}}\rangle| / \sum_{\mathbf{h}}\sum_l \langle I_{\mathbf{h}}\rangle$, where $I_{\mathbf{h}l}$ is the lth observation of reflection **h** and $\langle I_{\mathbf{h}}\rangle$ is the weighted average intensity for all observations l of reflection **h**. ‡ $\langle I/\sigma(I)\rangle$ indicates the average of the intensity divided by its average standard deviation.

3. Results and discussion

The structure of BAL at 2.6 Å resolution has been published by Schulz and coworkers (Mosbacher *et al.*, 2005). BAL is a homotetramer, each monomer being a chain of 563 amino acids. Each monomer is comprised of three domains: Dom-*α*, Dom-*β* and Dom-*γ*. The ThDP cofactor is bound in a deep pocket which is connected to the exterior of the protein by a channel containing water molecules. Residues from Dom-*γ* make interactions with the diphosphate moiety of ThDP and the magnesium, while those from Dom-*α* of a neighbouring subunit bind to the pyrimidine ring of ThDP. Unsurprisingly, the 1.65 Å resolution structure overlaps with the known structure with an r.m.s.d. of 0.65 Å. Four ThDP molecules were located within the protein as well as four magnesium ions. A number of additional peaks of electron density were observed, most notably in three areas of each subunit: (i) adjacent to Trp163, (ii) in a hydrophobic patch formed by Leu67 and Leu150 and (iii) between the C2 of the thiazolium ring of ThDP and the imidazole side chain of His29 in the active site. The best defined density of these strikingly similar regions were at the Leu67/Leu150 site of subunit *B* and adjacent to Trp163 in subunit *C*. These were helpful in attributing the density to the cocrystallization agent 2-methyl-2,4-pentanediol and were built and refined into the final structure. It has previously been demonstrated that leucine is by far the residue that is most observed

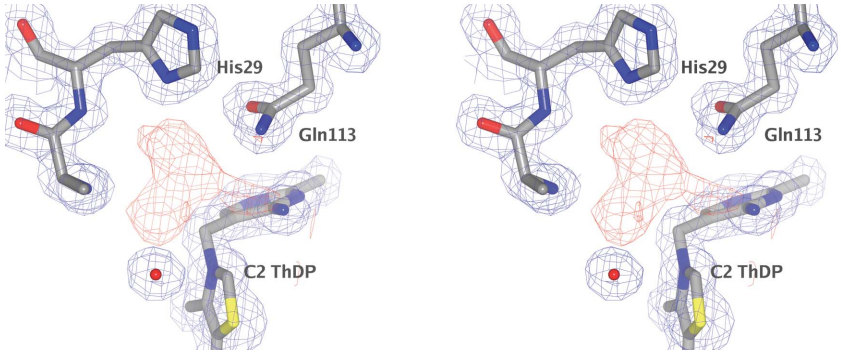


Figure 2
Stereoview of the ThDP cofactor at the active site of BAL, illustrating the extended resolution and the presence of unassigned density in the active site adjacent to C2 of ThDP. Blue electron density corresponds to a $2F_o - F_c$ map contoured at the 1σ level. Red ligand electron density corresponds to an $F_o - F_c$ electron-density map contoured at the 3σ level.

to make contacts with MPD in protein structures (Anand *et al.*, 2002). In the active site, the density is adjacent to C2 of the ThDP cofactor, which is the site of the reactive carbanion in the proposed mechanism for BAL. Although the shape of this density, allied with the high concentration of MPD in the crystallization conditions [45%(*v*/*v*)] and the known tendency of MPD to diffuse to enzyme active sites (Anand *et al.*, 2002), were suggestive of the density being due to an MPD molecule, the identity of the ligand could not be unambiguously assigned (Fig. 2).

The higher resolution structure of BAL has revealed a number of refinements to the published structure. The pyrimidine ring of the ThDP cofactor in the *A* subunit of BAL is in a different orientation to that in the *A* subunit of 2ag0, with the plane of the ring being rotated through 180°. The orientations in subunits *B*, *C* and *D* are in agreement with that of *A* and indeed with *B*, *C* and *D* of the cofactors in 2ag0. Although much of the two molecules is directly superimposable, the extended resolution has revealed split occupancies of residues such as (*A*)Ser151, (*A*)Leu272 and (*A*)Cys306 and alternate conformations of residues such as (*A*)Ile116. A detailed water structure of BAL is also revealed, with 1918 well defined water molecules within the structure. Residue Phe484, which was thought to be mobile and perhaps to act as a gate residue for BAL in the apoenzyme structure 2ag0, was found in an equivalent conformation in the structure presented here and the *B*-factors were not unusually high for this residue compared with the rest of the structure, as was observed with 2ag0.

In summary, a structure of the biotechnologically useful carboligase benzaldehyde lyase (BAL) is reported at significantly enhanced resolution and in the presence of 2-methyl-2,4-pentanediol (MPD) at hydrophobic patches on the enzyme surface. The higher resolution structure may provide an improved basis for the engineering of BAL for mechanistic studies and application.

We would like to thank the British Council, Berlin and the German Academic Exchange Service (DAAD), Bonn for a grant awarded under the Academic Research Collaboration (ARC). We also would like to thank the Deutsche Forschungsgemeinschaft (DFG) for financial support within Graduiertenkolleg 1166 'Bionoco' and Sonderforschungsbereich 540 and Professor Dr Maria-Regina Kula for the gift of the plasmid carrying the gene encoding BAL.

References

Anand, K., Pal, D. & Hilgenfeld, R. (2002). *Acta Cryst.* D**58**, 1722–1728.
Brzozowski, A. M. & Walton, J. (2001). *J. Appl. Cryst.* **34**, 97–101.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Demir, A. S., Şeşenoglu, O., Dünkelman, P. & Müller, M. (2003). *Org. Lett.* **5**, 2047–2050.
Demir, A. S., Şeşenoglu, O., Eren, E., Hosrik, B., Pohl, M., Janzen, E., Kolter, D., Feldmann, R., Dunkelmann, P. & Müller, M. (2002). *Adv. Synth. Catal.* **344**, 96–103.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gonzalez, B. & Vicuňa, R. (1989). *J. Bacteriol.* **171**, 2401–2405.
Grogan, G., Graf, J., Jones, A., Parsons, S., Turner, N. J. & Flitsch, S. L. (2001). *Angew. Chem. Intl. Ed.* **40**, 1111–1114.
Iding, H., Dünnwald, T., Greiner, L., Liese, A., Müller, M., Siegert, P., Grötzinger, J., Demir, A. & Pohl, M. (1998). *Eur. J. Biochem.* **257**, 538–546.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leonard, P. M., Brzozowski, A. M., Lebedev, A., Marshall, C. M., Smith, D. J., Verma, C. S., Walton, N. J. & Grogan, G. (2006). *Acta Cryst.* D**62**, 1494–1501.
Leonard, P. M. & Grogan, G. (2004). *J. Biol. Chem.* **279**, 31312–31317.
Mosbacher, T. G., Mueller, M. & Schulz, G. E. (2005). *FEBS J.* **272**, 6067–6076.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Otwinowski, Z. & Minor, V. (1997). *Methods Enzymol.* **276**, 307–326.
Pang, S. S., Duggleby, R. G. & Guddat, L. W. (2004). *J. Biol. Chem.* **317**, 249–262.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Sanchez-Gonzalez, M. & Rosazza, J. P. N. (2003). *Adv. Synth. Catal.* **345**, 819–824.
Stilger, T., Pohl, M., Wandrey, C. & Liese, A. (2006). *Org. Proc. Res. Dev.* **10**, 1172–1177.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.