Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Purification and crystallization of a proteolytic fragment of *Escherichia coli* pyruvate formate-lyase

Veli-Matti Leppänen,[a,b] Camran V. Parast,[c] Kenny K. Wong,[c] John W. Kozarich[c] and Adrian Goldman[a,b]*

[a]Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, PO Box 123, FIN-20521, Finland, [b]Department of Biochemistry and Food Technology, University of Turku, FIN-20014, Turku, Finland, and [c]Merck Research Laboratories, PO Box 2000, RY80A-4, Rahway, NJ 07065, USA

Correspondence e-mail: goldman@btk.utu.fi

Received 31 March 1998
Accepted 18 August 1998

Under anaerobic conditions, the reaction catalysed by pyruvate formate-lyase (PFL) is the first reaction after the production of pyruvate in the glycolytic pathway. Crystallization trials with *Escherichia coli* PFL were unsuccessful and therefore limited proteolysis was used to produce a stable crystallizable N-terminal protein fragment by trypsin cleavage. The molecular weight of this cleavage product was found to be 69.6 kDa by MALDI MS analysis, and the DNA sequence corresponding to this fragment was cloned. The recombinant protein fragment was crystallized by sitting-drop vapour diffusion using polyethylene glycol 1000 as precipitant. The crystals, which grew to 2 mm in length and 0.2 mm in cross section, belong to the hexagonal space group $P6_1$ or $P6_5$ with cell dimensions $a = b = 140.4$, $c = 215.3$ Å and two molecules per asymmetric unit. X-ray diffraction data were collected from 20 to 3.2 Å resolution from a single frozen crystal on a synchrotron-radiation beamline.

1. Introduction

Pyruvate formate-lyase (E.C. 2.3.1.54; PFL) from *Escherichia coli* is a central enzyme in bacterial anaerobic metabolism, catalysing the reversible conversion of pyruvate and coenzyme A into acetyl-CoA and formate (Knappe *et al.*, 1974). PFL is thus the anaerobic counterpart of pyruvate dehydrogenase in *E. coli*. PFL is a homodimeric protein (2 × 85 kDa, 759 residues) and catalytically inactive when isolated. Under anaerobic conditions it is converted to the active form by an activating enzyme in the presence of *S*-adenosyl-L-methionine, Fe(II) and reduced flavodoxin (Conradt *et al.*, 1984). Activated PFL contains one protein radical per dimer and the position of the amino-acid free radical has been mapped to the α-carbon of Gly734 (Wagner *et al.*, 1992). The glycine radical is required for activity (Unkrig *et al.*, 1989), as are two cysteines, Cys418 and Cys419, which have been shown to be the site of acetylation (Plaga *et al.*, 1988). We have recently shown (Parast *et al.*, 1995) that Cys418 plays the major role in the transfer of the acetyl moiety to CoA, while Cys419 is a direct mediator of the action of the glycyl free radical.

So far, the structure of pyruvate formate-lyase has not been reported and therefore we attempted to crystallize it for structural studies. PFL does not have sequence homology to any other proteins and, as it catalyses an unusual reaction, may also have an unusual fold. An overexpression system was established (Parast *et al.*, 1995) and highly pure PFL produced in large quantities, but all attempts to obtain diffraction-quality crystals failed. We report here how limited proteolysis results were used to produce a crystallizable recombinant core protein, the crystallization of that core protein and the results from preliminary X-ray analysis.

2. Materials and methods

2.1. Trypsin cleavage of recombinant PFL

Recombinant PFL was produced in *E. coli* BL21 cells carrying the pKK223-3 plasmid (Pharmacia) with the *pfl* gene under the control of the *tac* promoter (Parast *et al.*, 1995), and the protein was purified according to a procedure modified from Conradt *et al.* (1984). Optimal trypsin-digestion conditions were determined by carrying out small-scale reactions at 303 K and pH 7.5 using purified PFL, different trypsin concentrations and different periods of time. The trypsin (E.C. 3.4.21.4; Sigma, type XIII) was treated with L-1-tosylamide-2-phenylethyl chloromethyl ketone to reduce the chymotrypsin activity. Digestions were stopped by the addition of 1 m*M* phenylmethanesulfonyl fluoride (PMSF) and the results analysed by SDS and native PAGE. Analytical trypsin cleavages showed that, using a high trypsin:PFL ratio (1:50), dimeric 2 × 85 kDa PFL can be digested to a 2 × 70 kDa dimer. Preparative reactions were typically carried out using this trypsin:PFL ratio in 30 ml volumes with 50 mg of PFL, incubating the reaction mixtures at 303 K and pH 7.5 for 1 h. The protease-resistant core of PFL was purified using the same chromatographic protocol

as for the recombinant protein (see below). A more precise molecular weight for the core protein was obtained by matrix-assisted laser desorption ionization time-of-flight mass spectroscopy (MALDI MS, Finnigan Mat) with bovine serum albumin as standard.

2.2. Cloning

The cloning strategy for the introduction of a new stop codon after Arg624 was based on two unique restriction sites in the pKKBWM5.5C (Parast *et al.*, 1995) plasmid: a *Bst*XI site in the *pfl* gene upstream from Arg624 and a *Hin*dIII site downstream from Arg624 at the multiple cloning site of the pKK233-3 (Pharmacia) vector. A C-terminal fragment of the *pfl* gene, including nucleotides 1136–1975, was amplified by the polymerase chain reaction (PCR) using two oligonucleotide primers (ACCTCCACCTTCCTGGATGT and AGAGAAGCTTAACGACGACCGTCTGGG; *Bst*XI and *Hin*dIII sites underlined, respectively; purchased from Kebo Lab, Stockholm, Sweden). The first oligonucleotide corresponds to an internal sequence of the gene and includes a *Bst*XI restriction site, while the second introduces a new stop codon immediately after Arg624 as well as a new *Hin*dIII restriction site. The amplified DNA was cleaved with the restriction endonucleases, ligated to similarly cleaved plasmid pKK-PFL and used to transform competent *E. coli* JM109 cells by electroporation. Ampicillin-resistant transformants were screened first for plasmids with insert DNA of correct size and later for protein overexpression. The modified DNA of one of these transformants was completely sequenced on an ABI 377 sequencer and there were no unexpected changes. The resulting construct, pKK-TRYPFL, was used to transform a protease-deficient BL21-(DE3) strain of *E. coli* for protein production.

2.3. Purification of the recombinant fragment protein

A 0.8 l culture of Terrific Broth (19.2 g yeast extract, 9.6 g bacto-peptone, 2.1 g KH_2PO_4, 3.5 g Na_2HPO_4 and 6 ml glycerol) containing 100 µg ml^{-1} ampicillin was inoculated with a 5 ml overnight culture of BL21(DE3)/pKK-TRYPFL and grown at 310 K to an A_{600} of 0.5–0.6. PFL expression was then induced by the addition of 0.5 m*M* isopropyl β-D-thiogalactopyranoside (IPTG) and the cells were grown for a further 3–4 h before being harvested by centrifugation at 10000*g*. The cell pellets were washed once with 25 m*M* Tris–HCl (pH 7.5) containing 50 m*M* NaCl and stored at 203 K until required. They were then resuspended in 20 ml 50 m*M* MES–HCl (pH 7.0) containing 5 m*M* dithiothreitol (DTT), 5 m*M* EDTA, 0.1 m*M* PMSF and 5% glycerol. Using a Labsonic 2000 sonicator (B. Braun Biotech) with a 40 T probe, the cells were lysed by pulsing the suspension 15 times on ice for 30 s and allowing it to cool after each pulse for 30 s. After the cell debris had been removed by centrifugation at 100000*g*, the supernatant was applied to a Q Sepharose (Pharmacia) anion-exchange column equilibrated with 20 m*M* Bis–Tris–HCl (pH 6.4) driven by a Pharmacia FPLC system. After washing the column with a column volume of the Bis–Tris buffer, the protein was eluted with a linear gradient from 0.1 *M* KCl (pH 6.4) to 0.5 *M* KCl (pH 5.8). The protein eluted at approximately 0.3 *M* KCl (pH 6.1). Fractions containing recombinant protein were pooled and concentrated before loading onto a Superdex 200 HR26/600 (Pharmacia) gel-filtration column. The running buffer for gel filtration was 50 m*M* Tris–HCl (pH 7.5), 100 m*M* NaCl. The final purification step was hydrophobic interaction chromatography using a BioCAD and a 2 ml POROS 50 PE column (Perseptive Biosystems, Massachusetts, USA). PFL was loaded in 1.0 *M* ammonium sulfate and eluted with a decreasing ammonium sulfate gradient from 1.0 to 0 *M*. The protein eluted at about 0.6 *M* ammonium sulfate.

2.4. Crystallization

For crystallization the buffer was changed to 10 m*M* HEPES–HCl (pH 7.6) with a HiTrap (Pharmacia) column and the protein was concentrated to the desired concentration (up to 50 mg ml^{-1}) using Centricon-30 (Amicon) concentrators. Additives used in the protein solution were 0.01% NaN_3, 0.1 m*M* EDTA, 0.1 m*M* PMSF and 3 m*M* DTT. Initial screening was performed according to published incomplete factorial-search methods (Jancarik & Kim, 1991; Shieh *et al.*, 1995) at different temperatures. Diffraction-quality crystals were obtained by vapour diffusion using 30 µl sitting drops (McPherson, 1990) in Chryschem MVD/24 plates (Charles Supper Co.) at 294 K with a 5:1 ratio of protein (30–50 mg ml^{-1}) to well solution. The well solution was 100 m*M* HEPES–HCl (pH 7.6) and 18–24%(*w*/*v*) polyethylene glycol 1000. Hexagonal needle-like crystals grew to 2 mm in length and 0.2 mm in diameter within 2–3 weeks. The crystals were harvested and stored in a slightly higher polyethylene glycol concentration (25%).

2.5. Data collection

The harvesting solution was readily flash-freezable when ethylene glycol was added to a concentration of 10%(*v*/*v*). Crystals were quickly soaked in this solution, frozen using standard techniques (Teng, 1990) in a stream of nitrogen gas and maintained at 103 K using a cryo-system made by Molecular Structure Corporation (Texas). Diffraction data were collected both on a rotating anode and at HASYLAB at wavelengths of 1.54 and 1.00 Å, respectively. The in-house Rigaku RU-200B rotating anode Cu $K\alpha$ radiation source was operated at 50 kV and 180 mA with R-AXIS II image plates. Synchrotron radiation was used for data collection on the X31 beamline at HASYLAB (Hamburg) with an 18 cm MAR Research image plate. The data were processed and indexed using the program *DENZO* (Otwinowski, 1993).

3. Results and discussion

Initial crystallization attempts from intact protein were disappointing. One possible explanation was heterogeneity: an endogenous protease led to a 3:1 ratio of full-length (85 kDa) PFL to 78 kDa (Conradt *et al.*, 1984) cleavage product. This was eliminated, however, by screening a number of

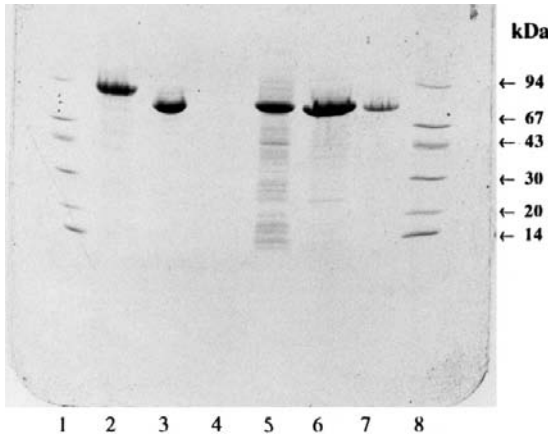

Figure 1
SDS–PAGE analysis of PFL samples. Lanes 1 and 8, molecular-weight standards; lane 2, PFL produced in BL21; lane 3, trypsin-cleaved PFL; lane 4, empty; lane 5, crude extract of the recombinant core protein; lane 6, the recombinant core protein from the anion-exchange column; lane 7, purified recombinant protein.

Table 1
Data-collection statistics.

X-ray source	R-AXIS RU-200B†		X31 beamline‡	
Resolution shell (Å)	20.0–3.50	3.56–3.50	20.0–3.20	3.27–3.20
Number of crystals	1		1	
Observations	156323		24038	
Unique reflections	14339	638	35016	2233
Average $I/\sigma(I)$	8.3	3.9	9.6	2.8
Completeness (%)	93.5	82.9	89.9	87.2
$I/\sigma(I) > 3$ (%)	83.9	84.8	74.1	44.0
$R_{\rm merge}$§ (%)	12.0	19.8	9.4	33.7%
Mosaicity (°)	0.7		0.4	

† 80 1° oscillation pictures were collected. ‡ 60 1° oscillation pictures were collected. § $\sum|I - \langle I\rangle|/\sum I$.

protease-deficient *E. coli* strains as expression hosts (data not shown). We found that the BL21(DE3) strain produced intact PFL alone (Fig. 1, lane 2) but even this protein did not yield usable crystals. Another likely explanation is that the PFL we made was not activated; the inactive protein may well have a disordered C-terminus, as activation involves creation of a free radical on Gly738 (see §1). Consistent with this, several proteases have cleavage sites at the C-terminus (Rödel *et al.*, 1988; Knappe *et al.*, 1993). We therefore decided to attempt to crystallize a proteolytic fragment, as PFL does not have significant sequence identity to any other protein in the sequence databases and so may represent a novel fold. Furthermore, the structure will reveal the arrangement of Cys418 and Cys419, thus allowing the formulation of testable hypotheses about mechanism.

Trypsin cleavage of PFL produced a stable core protein which could be purified (see §2) to a single band on SDS–PAGE (Fig. 1, lane 3). The MALDI MS analysis gave a molecular weight of 69.6 kDa for the trypsin-cleavage product, which is consistent with the value derived from the DNA sequence for the first 624 amino acids (Rödel *et al.*, 1988). However, microheterogeneity was sometimes seen on native PAGE indicating several cleavage sites close to each other, as originally reported by Rödel *et al.* (1988), and the yield of pure fragment (1–624) from the trypsin-cleavage reaction was only 40%. Consequently, we cloned a construct encoding the protease-resistant fragment by introducing a stop codon and deleting the 3′-end of the *pfl* gene. Refinement of crystallization conditions for trypsin-cleaved PFL produced thin needle crystals that diffracted poorly to 4.0 Å, but were adequate for a preliminary analysis. Later, the recombinant protein was shown to crystallize under essentially the same conditions with the same overall morphology and unit-cell dimensions. The best conditions were 18% polyethylene glycol 1000 and 100 m*M* HEPES–HCl (pH 7.6). We found that the most important variables were the protein-to-precipitant ratio in the sitting drops and the final drop size. The best crystals were obtained from 30 µl sitting drops with relatively high protein concentrations (30–50 mg ml^{-1}) and a high protein-to-well solution ratio (5:1). The conditions were not very reproducible and therefore ranges of protein and precipitant concentrations were always used. Hanging drops in Linbro tissue-culture plates were also tried, but resulted in showers of small crystals. We believe this to be a consequence of the high degree of dessication which leads to a 15–20 µl drop shrinking to only a few microlitres.

The diffraction pattern of the crystals is consistent with either space group $P6_1$ or $P6_5$ with cell dimensions $a = b = 140.4$ and $c = 215.3$ Å. The self-rotation function calculated in *X-PLOR* (Brünger, 1992) shows a non-crystallographic twofold axis at $\psi = 8°$, $\varphi = 0°$ at 75% of the origin peak. If there are two monomers per asymmetric unit, V_M is 4.4 Å^3 Da^{-1} (Matthews, 1968) corresponding to a solvent content of 72%. This is rather high, but may be consistent with the poor diffraction of these crystals. The first data set (Table 1) was collected in-house from 20 to 3.5 Å resolution with an *R* factor of 12.0%. The more intense and monochromatic synchrotron radiation on the X31 beamline at DESY (Hamburg) gave a substantially improved signal-to-noise ratio (Table 1) and allowed a complete data set to be collected to 3.2 Å on a single crystal 0.6 mm long and 0.16 mm in diameter. However, this is not the diffraction limit of the crystals, as diffraction maxima were initially seen to 2.7 Å resolution; 3.2 Å was the maximal resolution achievable in the beam time allotted. Data collection beyond 3 Å would therefore be possible on a high-brilliance beamline. 242038 measurements were made of 35016 unique reflections, and the final merging *R* factor was 9.4% with 89.9% completeness from 20 to 3.2 Å resolution. A heavy-atom search is under way.

This work was supported by a grant from Merck Research Laboratories (NJ, USA) and by the general facilities of the Turku Centre for Biotechnology (Finland). The staff of the EMBL outstation in Hamburg are acknowledged for professional assistance and we thank EMBL for access to the X31 synchrotron beamline. We also thank Raija Andersen for technical assistance with the crystallization, Dan-Johan Still for help with the computers and Jussi Kankare for help with the data collection in Hamburg.

References

Brünger, A. T. (1992). *X-PLOR Version* 3.1. *A System for X-ray Crystallography and NMR.* New Haven: Yale University Press.
Conradt, H., Hohmann-Berger, M., Hohman, H. P., Blaschkowski, H. P. & Knappe, J. (1984). *Arch. Biochem. Biophys.* **228**, 133–142.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Knappe, J., Blaschkowski, H. P., Gröbner, P. & Schmitt, T. (1974). *Eur. J. Biochem.* **50**, 253–263.
Knappe, J., Elbert, S., Frey, M. & Wagner, A. F. V. (1993). *Biochem. Soc. Trans.* **21**, 731–734.
McPherson, A. (1990). *Eur. J. Biochem.* **189**, 1–23.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. (1993). *Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Parast, C. V., Wong, K. K., Lewisch, S. A., Peisach, J., Magliozzo, R. S. & Kozarich, J. (1995). *Biochemistry*, **34**, 2393–2399.
Plaga, W., Frank, R. & Knappe, J. (1988). *Eur. J. Biochem.* **178**, 445–450.
Rödel, W., Plaga, W., Frank, R. & Knappe, J. (1988). *Eur. J. Biochem.* **177**, 153–158.
Shieh, H.-S., Stallings, W. C., Stevens, A. M. & Stegeman, R. A. (1995). *Acta Cryst.* D**51**, 305–310.
Teng, T.-Y. (1990). *J. Appl. Cryst.* **23**, 387–391.
Unkrig, V., Neugebauer, F. A. & Knappe, J. (1989). *Eur. J. Biochem.* **184**, 723–728.
Wagner, A. F. V., Frey, M., Neugebauer, F. A., Schäfer, W. & Knappe, J. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 996–1000.