X-ray crystal structure of a malonate-semialdehyde dehydrogenase from *Pseudomonas* sp. strain AAC

Matthew Wilding,^a Colin Scott,^a Thomas S. Peat^b and Janet Newman^{b*}

^aLand and Water, CSIRO, GPO Box 1700, Canberra, ACT 2601, Australia, and ^bBiomedical Program, Manufacturing, CSIRO, 343 Royal Parade, Parkville, VIC 3052, Australia. *Correspondence e-mail: janet.newman@csiro.au

Received 5 November 2016

Accepted 15 December 2016

Edited by P. Dunten, Stanford Synchrotron Radiation Lightsource, USA

Keywords: malonate-semialdehyde dehydrogenase; β -alanine catabolism; *Pseudomonas* sp. strain AAC.

PDB reference: malonate semialdehyde dehydrogenase, 5tjr

Supporting information: this article has supporting information at journals.iucr.org/f

The NAD-dependent malonate-semialdehyde dehydrogenase KES23460 from *Pseudomonas* sp. strain AAC makes up half of a bicistronic operon responsible for β -alanine catabolism to produce acetyl-CoA. The KES23460 protein has been heterologously expressed, purified and used to generate crystals suitable for X-ray diffraction studies. The crystals belonged to space group $P2_12_12_1$ and diffracted X-rays to beyond 3 Å resolution using the microfocus beamline of the Australian Synchrotron. The structure was solved using molecular replacement, with a monomer from PDB entry 4zz7 as the search model.

1. Introduction

The malonate-semialdehyde dehydrogenase KES23460 from *Pseudomonas* sp. strain AAC is responsible for the production of acetyl-CoA during β -alanine catabolism (Wilding *et al.*, 2016). Unlike the highly promiscuous transaminase involved in this pathway (KES23458), the dehydrogenase is selective towards malonate semialdehyde and generates acetyl-CoA in an NAD-dependent and CoA-dependent reaction, although a slower CoA-independent reaction generating acetate has also been observed (Wilding *et al.*, 2016).

Solving the X-ray structure of KES23460 proved particularly challenging. Despite being able to generate crystals under a range of conditions, and the availability of a number of pre-existing malonate-semialdehyde dehydrogenase structures [for example, PDB entries 4zz7 (Do *et al.*, 2016), 4iym (New York Structural Genomics Research Consortium, unpublished work) and 1t90 (H. Dubourg, C. Didierjean, C. Stines-Chaumeil, F. Talfournier, G. Branlant, A. Aubry & C. Corbier, unpublished work)], a high-resolution structure proved difficult to attain initially using molecular replacement. Here, we report the purification, crystallization and diffraction of KES23460, and the first example of a malonate-semialdehyde dehydrogenase that exists in a hexameric state.

2. Materials and methods

2.1. Macromolecule production

A gene encoding KES23460 that was codon-optimized for expression in *Escherichia coli* was ordered in the expression vector pRSetB from Life Technologies (Germany). *E. coli* BL21 (DE3) cells were transformed by electroporation with the pRSetB plasmid containing the gene encoding KES23460, and the cells were grown in LB medium supplemented with ampicillin (100 µg ml⁻¹) at 310 K. When the optical density at 600 nm (OD₆₀₀) reached 0.6–1.0, the culture was induced by the addition of isopropyl β -D-1-thiogalactopyranoside (IPTG; 1 mM final concentration). The cells were further incubated

Table 1

Macromolecule-production information.

The start of the KES23460 protein is underlined; the tag adds an extra 33 residues to the N-terminus of this construct.

Source organism	<i>Pseudomonas</i> sp. strain AAC
DNA source	cDNA, Life Technologies
Forward primer	N/A
Reverse primer	N/A
Cloning vector	N/A
Expression vector	pRSetB
Expression host	<i>E. coli</i> BL21 (DE3)
Complete amino-acid sequence of the construct produced	MRGSHHHHHGMASMTGGQMGRLDYDDDKGSM- TLIKHLIGGELIADTGRADVFNPSTGEAVRK- VPLADRETMQQAIDAAKAAPAWRTPPAKRA- QVLFRRFKQLLEANEERIVKLI SEEHGKTIEDA- AGELKRGIEVYATAAPEILKGEYSRNVGPN- IDAWSDFPIGVVAGITPFNFPAMVPLWYPL- AIACGNTFILKPSERDPSSTLLIAELFHEAGL- PKGVLVVHVGDKGAVDALIEAPEVKALSFVGS- TPIAEYIYSEGTRKGRVQALGGAKNHAVLMP- DADLDNAVSAALMGAAAYGSCGERCMAISVAVCV- GDQIADALVQKLVPQIKGLKIGAGTSCGLDMG- PLVTGAARDKVTGYIDTGVAQGAELVVDGRGY- KVAGHENGFFLGGLTDFRVTPMTIYKEEIFG- PVLICIVRVNSLEAMQLINDHEYNGTCIFTR- DGEAARLFCDEIEVGMVGVNPLVPVAYHSF- GGWKRSLFGDLHAYGPDGVRFYTKRKAITQRW- PQRKSHEAAQFAFPSNS

with shaking overnight at 310 K and harvested by centrifugation in a Beckman Avanti J-E centrifuge (4000g; 20 min), and the supernatant was discarded. The pellet was resuspended in imidazole–sodium chloride–Tris buffer (5 mM imidazole, 250 mM NaCl, 10 mM Tris chloride pH 7.5), and cell lysis was achieved by homogenization at 138 MPa using a Microfluidics M-110P Microfluidizer. Cellular debris was precipitated by centrifugation (40 000g; 45 min) and the supernatant was passed over a HiTrap Chelating HP column (GE Healthcare, Little Chalfont, England) on an ÄKTA fast protein liquid chromatography (FPLC) system (GE

Figure 1

Typical rhombohedral crystal of KES23460 grown from PEG and citrate after seeding. Crystals are generally 50 µm along each edge and grow most reliably at 281 K.

Table 2

Crystallization.

Method	Vapour diffusion
Plate type	Swissci SD-2 ('MRC' plates)
Temperature (K)	281
Protein concentration (mg ml ⁻¹)	5
Buffer composition of protein solution	40 mM Tris chloride, 150 mM NaCl, 2 mM KCl pH 8 (TBS)
Composition of reservoir solution	23.7%(w/v) PEG 3350, 0.208 M trisodium citrate, 0.1 M bis-tris propane pH 7.55
Volume and ratio of drop	300 nl total: 150 nl protein, 120 nl reservoir, 30 nl seeds
Volume of reservoir (µl)	50

Healthcare). The protein was eluted with an increasing concentration of imidazole (5–500 mM), and the fractions containing protein were pooled, concentrated and applied onto a size-exclusion column that had been pre-equilibrated with Tris-buffered saline (TBS; 40 mM Tris chloride pH 8, 150 mM NaCl, 2 mM KCl). The protein eluted as a single peak, and the retention time on the column suggests that the protein forms a large oligomer, possibly a tetramer or hexamer. The fractions from the gel filtration were pooled and concentrated to either 20 mg ml⁻¹ (initial purification) or 60 mg ml⁻¹ (subsequent purification) in a 10 kDa spin concentrator. The purity was estimated to be >95% by SDS-PAGE, and the identity of the protein was confirmed by mass spectrometry. Macromolecule-production information is summarized in Table 1.

2.2. Crystallization

Protein at 10 and 20 mg ml⁻¹ in TBS was screened against an initial bank of 384 crystallization conditions [Shotgun (Fazio *et al.*, 2014) at 281 and 293 K, and PACT (Newman *et al.*, 2005) and PS_Gradient (an in-house pH *versus* salt screen) at 293 K; see <http://c6.csiro.au> (Newman *et al.*, 2010) for a complete description of screening conditions], which gave a single very preliminary hit in 1.5 M ammonium sulfate. Optimizing around this condition with rounds of additive screening produced small hexagonal rods in a condition consisting of 1.38 M ammonium sulfate, 3.35% 2,2,2-trifluoroethanol, 2% benzamidine. No diffraction was seen from these crystals, but they were used to create a seed stock which was used to seed another round of initial screening, this time using 5 mg ml⁻¹ protein supplemented with 1 mM acetyl-CoA. These screens gave some small, rhombohedral crystals in several conditions from the PACT screen. A 96-condition optimization screen was created based on these hits, and protein was set up at both 281 and 293 K against this screen using different ratios of protein, reservoir and seed stock. A single, sharp-edged rhombohedral crystal that grew in well E7 of the optimization screen [23.7%(w/v) PEG 3350, 0.208 M trisodium citrate, 0.1 M bis-tris propane pH 7.55] at 281 K (Fig. 1) was cryo-protected in reservoir solution diluted with 90% glycerol to give a final concentration of 20% glycerol and harvested in the cold. Many other experiments (over 50 crystallization plates in total; ~9500 droplets) with protein supplemented with both NAD and/or acetyl-CoA were tried, varying the protein

Table 3

Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	MX2, Australian Synchrotron
Wavelength (Å)	0.9537
Temperature (K)	100
Detector	ADSC Quantum 315
Crystal-to-detector distance (mm)	400
Rotation range per image (°)	1
Total rotation range (°)	360
Exposure time per image (s)	1
Space group	19
<i>a</i> , <i>b</i> , <i>c</i> (Å)	115.9, 156.6, 192.2
α , β , γ (°)	90, 90, 90
Mosaicity (°)	0.19
Resolution range (Å)	49.6–2.95 (3.01–2.95)
Total No. of reflections	1094959 (68181)
No. of unique reflections	74332 (4539)
Completeness (%)	99.9 (100)
Multiplicity	14.7 (15.0)
$\langle I/\sigma(I) \rangle$	11.5 (2.0)
$R_{p.i.m.}$	0.049 (0.439)
Overall <i>B</i> factor from Wilson plot (Å ²)	56.6

formulation and/or using different seeds, different droplet ratios, different geometries (microbatch, 24-well hanging drops) and *in situ* proteolysis (enterokinase and chymotrypsin), but no better diffraction was obtained. Crystalization information is summarized in Table 2.

2.3. Data collection and processing

Data were collected as 360 1° oscillations on the MX2 beamline of the Australian Synchrotron. The data were indexed using *xia2/DIALS* (Waterman *et al.*, 2016), the space group was determined to be *P*₂₁₂₁ using *POINTLESS* (Evans, 2006) and the data were scaled using *AIMLESS* (Evans & Murshudov, 2013). Data-collection and processing statistics are summarized in Table 3.

2.4. Structure solution and refinement

Solvent-content calculations (Matthews, 1968) suggested that there could be four to eight monomers in the asymmetric unit: there would be 67% solvent with four monomers, 35% solvent with eight monomers and 51% solvent with six monomers; the latter is equivalent to a Matthews coefficient of 2.51 Å³ Da^{−1}, which seemed the most likely. The structure was solved by molecular replacement using *Phaser* (McCoy *et al.*, 2007) with PDB entry 4zz7 (Do *et al.*, 2016) as a model. The model was built initially using *Buccaneer* (Cowtan, 2006) and then manually rebuilt multiple times using *Coot* (Emsley *et al.*, 2010). The structure was refined using both *phenix.refine* (Echols *et al.*, 2014) and *REFMAC* (Murshudov *et al.*, 2011) at different phases of the refinement cycle. The structure and structure factors have been deposited in the PDB as entry 5tjr. Refinement statistics are summarized in Table 4.

3. Results and discussion

Previous investigations (Wilding *et al.*, 2016) have shown that KES23460 catalyses the NAD-dependent conversion of

Table 4

Structure solution and refinement.

Resolution range (Å)	49.6–2.95
Completeness (%)	99.9
No. of reflections, working set	70516
No. of reflections, test set	3714
Final R_{cryst} (%)	21.2
Final R_{free} (%)	24.9
No. of non-H atoms	
Protein	20824
Ligand	135
Water	16
Total	20975
R.m.s. deviations	
Bonds (Å)	0.012
Angles (°)	1.686
Average <i>B</i> factors (Å ²)	72.0
Protein	73.3
Ligand	80.7
Water	32.9
Ramachandran plot	
Most favoured (%)	95.5
Outliers (%)	0.1

malonate semialdehyde to two different products depending on the availability of CoA as a co-substrate. In the presence of CoA, malonate semialdehyde is converted to acetyl-CoA and bicarbonate, but without CoA a slower reaction can also be observed that ultimately generates acetate. Having purified the protein to apparent homogeneity using the methods described herein, crystal screening trials were undertaken, and considering the reaction(s) catalysed by the enzyme, additional conditions supplementing KES23460 protein with cofactor (NAD) and/or product (acetyl-CoA) were also trialled.

There are several potential models in the PDB (Berman *et al.*, 2007) with sufficient homology (hypothetically) to solve the structure of KES23460, but after several failed attempts with different models with up to 55% identity (for example PDB entries 4iym and 1t90), we explored using *de novo* phase information by soaking crystals with different iodine-containing compounds. A combination of diffraction to poor resolution (3.5–4 Å) and even poorer anomalous signal led to a second attempt with molecular replacement using a newly deposited structure: PDB entry 4zz7. Six molecules were found in the asymmetric unit and the quaternary structure is that of a hexamer (a trimer of dimers; Fig. 2); this quaternary arrangement has not been seen before for malonate dehydrogenases (Do *et al.*, 2016). No density is seen for the purification tag. There are differences in the quality of the electron density for the six monomers in the asymmetric unit; however, all show density from either the starting methionine or the following threonine residue. There are short breaks in the electron density for all chains around residue 276 and (for all but chain *D*) between residues 352 and 363, as well as a longer break centred on residue 316. Four residues could not be modelled at the C-terminus of all chains, so most of the 498 residues can be modelled. Analysis of the structure by *PISA* (Krissinel & Henrick, 2007) suggests that the protein is a hexamer in solution with a buried surface area of 28 140 Å² from a total surface area of about 103 000 Å² (over 25%

Figure 2

Images of KES23460. (a) and (b) represent two different views of a monomer of KES23460, showing a cleft between the two main domains of the protein (where cofactor is bound; shown as spheres) and how there is a long extension of a three-stranded β -sheet and loop that extends away from the body of the protein. This extended structure is made up of residues 119–136 and then 473 to the C-terminus (494). (c) and (d) show the extensive dimer interface (the figures are rotated by approximately 90°). (e) and (f) show the full hexamer of the structure as seen in the crystal structure. These images also show how the very end of the C-terminus of each monomer interacts with a neighbouring dimer, pulling the full hexamer together.

buried). The protomers in the asymmetric unit vary in quality, some having higher average *B* factors and missing loop residues, with only five of the structures containing cofactor density that could be modelled. Only a partial ligand was seen in the model, and we do not know where this comes from. It might be from the acetyl-CoA which was added during crystallization, or it could be from NAD that was picked up during expression (as ADP is a component of both of these cofactors). The adenosine ring was modelled into a pocket formed by Lys205, Val208, Ile228 and Tyr231, whilst the ribose and phosphate moieties were modelled into a pocket comprising Thr146, Phe148, Lys172, Glu175, Arg176 and Ser225.

The monomeric structure is made up of two primary domains. Each has a central extended β -sheet surrounded by α -helices, with a cleft between them which holds the cofactor (Figs. 2*a* and 2*b*). The N-terminus from residues 1 to 38 consists of an extended β -structure before the start of the first α -helix, and there is a central β -sheet of five strands surrounded by α -helices in the first domain. This domain serves as the cofactor-binding site (residues 39–248, minus the extension of 119–136; see below). The second primary domain extends from residues 249 to 444 and has a central β -sheet of seven strands surrounded by α -helices. This second β -sheet is extended by the three β -strands of the neighbouring dimer extension (residues 119–136 and 444–494) to make a β -sheet of ten strands in the dimer structure (Figs. 2*c* and 2*d*). The ‘finger’ extension (residues 119–136 and 444–480) forms a three-stranded β -sheet extension which pulls the dimer structure together, but is also used as a hook to pull in a neighbouring dimer and form the basis of the hexameric structure. This quaternary structure is significantly different to those of homologous proteins solved to date, which are all tetrameric. Despite the difference in quaternary structure, all six of the residues highlighted by Do and coworkers as stabilizing the intermediate of the reaction of the methylmalonate-semialdehyde dehydrogenase, as well as the catalytic cysteine itself, are conserved in KES23460 (Do *et al.*, 2016).

After solving the structure, the difficulties associated with the molecular replacement from the highly homologous sequence models became clearer, as despite their homology the structures vary significantly. PDB entry 4iym has 48% sequence identity to KES23460 (498 residues); using the *SSM* algorithm in *Coot* for a given protomer exhibited a 2.2 Å r.m.s.d. for C α atoms over 388 residues (14 gaps; monomer). Similarly, PDB entry 4v6h (Seattle Structural Genomics Center for Infectious Disease, unpublished work) has >50% sequence identity to KES23460 and a similar 2.2 Å r.m.s.d. for C α atoms over 402 residues (14 gaps) using the same algorithm, whilst PDB entry 4zz7 has >60% sequence identity and a 2.1 Å r.m.s.d. for C α atoms over 404 residues (17 gaps). PDB entry 1t90 has only 44% identity but aligns 411 C α atoms (12 gaps) with an r.m.s.d. of 2.2 Å. The quaternary structure of all of these homologous models (and others) found in the PDB is

tetrameric, whereas KES23460 was solved as a hexamer. A single monomer gives only a weak molecular-replacement solution, as it captures at best only one sixth of the scattering matter of the hexamer, and yet when higher order oligomers (dimers, tetramers) were used as search models they were less likely than the monomer to recapitulate the underlying structure owing to the intermolecular arrangements to form a hexamer instead of a tetramer.

4. Conclusions

We present the 2.95 Å resolution structure of the NAD-dependent malonate-semialdehyde dehydrogenase from *Pseudomonas* sp. strain AAC, which appears from both crystallographic and size-exclusion chromatography data to exist in solution as a hexamer. This is the first example of a hexameric assembly amongst the structures of homologous proteins available in the PDB.

Acknowledgements

We thank the beamline scientists of the Australian Synchrotron, and acknowledge the use of the CSIRO Collaborative Crystallization Facility. This research was supported in part by the Science and Industry Endowment Fund. Many thanks to Airlie McCoy for helpful discussions regarding molecular replacement in difficult cases.

References

- Berman, H., Henrick, K., Nakamura, H. & Markley, J. L. (2007). *Nucleic Acids Res.* **35**, D301–D303.
- Cowtan, K. (2006). *Acta Cryst.* **D62**, 1002–1011.
- Do, H., Lee, C. W., Lee, S. G., Kang, H., Park, C. M., Kim, H. J., Park, H., Park, H. & Lee, J. H. (2016). *J. Microbiol.* **54**, 114–121.
- Echols, N., Moriarty, N. W., Klei, H. E., Afonine, P. V., Bunkóczi, G., Headd, J. J., McCoy, A. J., Oeffner, R. D., Read, R. J., Terwilliger, T. C. & Adams, P. D. (2014). *Acta Cryst.* **D70**, 144–154.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* **D69**, 1204–1214.
- Fazio, V. J., Peat, T. S. & Newman, J. (2014). *Acta Cryst.* **F70**, 1303–1311.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I. & Perrakis, A. (2005). *Acta Cryst.* **D61**, 1426–1431.
- Newman, J., Fazio, V. J., Lawson, B. & Peat, T. S. (2010). *Cryst. Growth Des.* **10**, 2785–2792.
- Waterman, D. G., Winter, G., Gildea, R. J., Parkhurst, J. M., Brewster, A. S., Sauter, N. K. & Evans, G. (2016). *Acta Cryst.* **D72**, 558–575.
- Wilding, M., Peat, T. S., Newman, J. & Scott, C. (2016). *Appl. Environ. Microbiol.* **82**, 3846–3856.