

# Crystal structure of the [2Fe–2S] protein I (Shethna protein I) from *Azotobacter vinelandii*

Burak V. Kabasakal,<sup>a,b</sup> Charles A. R. Cotton<sup>a,c</sup> and James W. Murray<sup>a\*</sup>

<sup>a</sup>Department of Life Sciences, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom, <sup>b</sup>Turkish Accelerator and Radiation Laboratory, Institute of Accelerator Technologies, Ankara University, Gölbaşı, 06830 Ankara, Turkey, and <sup>c</sup>Cambrium GmbH, Max-Urlich-Strasse 3, 13355 Berlin, Germany. \*Correspondence e-mail: j.w.murray@imperial.ac.uk

Received 26 April 2021

Accepted 23 September 2021

Edited by M. A. Hough, University of Essex, United Kingdom

**Keywords:** iron–sulfur proteins; Shethna protein I; *Azotobacter vinelandii*.

**PDB reference:** [2Fe–2S] protein I from *Azotobacter vinelandii*, 5abr

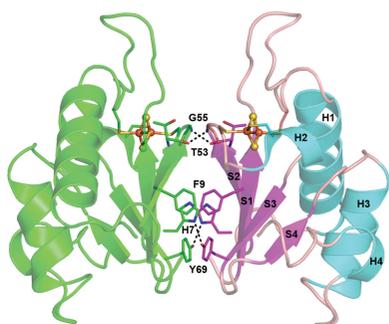
**Supporting information:** this article has supporting information at journals.iucr.org/f

*Azotobacter vinelandii* is a model diazotroph and is the source of most nitrogenase material for structural and biochemical work. *Azotobacter* can grow in above-atmospheric levels of oxygen, despite the sensitivity of nitrogenase activity to oxygen. *Azotobacter* has many iron–sulfur proteins in its genome, which were identified as far back as the 1960s and probably play roles in the complex redox chemistry that *Azotobacter* must maintain when fixing nitrogen. Here, the 2.1 Å resolution crystal structure of the [2Fe–2S] protein I (Shethna protein I) from *A. vinelandii* is presented, revealing a homodimer with the [2Fe–2S] cluster coordinated by the surrounding conserved cysteine residues. It is similar to the structure of the thioredoxin-like [2Fe–2S] protein from *Aquifex aeolicus*, including the positions of the [2Fe–2S] clusters and conserved cysteine residues. The structure of Shethna protein I will provide information for understanding its function in relation to nitrogen fixation and its evolutionary relationships to other ferredoxins.

## 1. Introduction

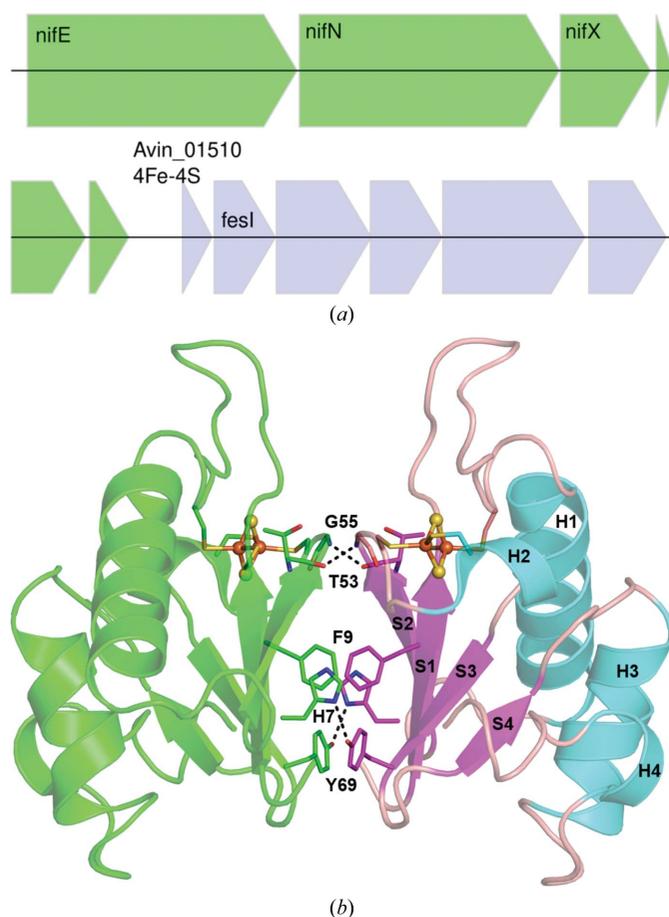
*Azotobacter vinelandii* is a model organism for nitrogen fixation (Peters *et al.*, 1995). Although nitrogenase is inactivated by oxygen, *A. vinelandii* can grow in above-atmospheric concentrations of oxygen (Maier & Moshiri, 2000). This oxygen resistance arises via at least two mechanisms: firstly a high respiratory rate, which removes oxygen, called ‘respiratory protection’ (Jones *et al.*, 1973), and secondly ‘conformational protection’, in which high oxygen levels pause nitrogenase activity, but on their removal nitrogen fixation resumes. Conformational protection is mediated by [2Fe–2S] protein II (FeSII) or Shethna protein, which forms a complex with nitrogenase that is catalytically inactive but resistant to oxygen (Robson, 1979; Moshiri *et al.*, 1994; Schlesier *et al.*, 2016). The FeSII protein was first purified by Shethna in the 1960s (Shethna *et al.*, 1964, 1968) as a nonheme iron protein of interest, together with another protein, the so-called Shethna protein I or [2Fe–2S] protein I (FeSI).

Ferredoxins are iron–sulfur proteins that mediate electron transfer, and there are several families of them. 2Fe ferredoxins can be classified as plant-type, mitochondrial-type and bacterial ferredoxins. There is also another class comprising 3Fe, 4Fe, 7Fe and 8Fe ferredoxins with one or two FeS clusters of the cubane type (Zanetti & Pandini, 2013). FeSI is a member of the thioredoxin-like ferredoxin family (InterPro family IPR009737). Thioredoxins have no Fe–S clusters, but have pairs of cysteine residues to exchange disulfide-bond oxidation states (Saarinen *et al.*, 1995). These thioredoxin-like



ferredoxins are also present in some multimeric hydrogenases (Appel & Schulz, 1996; De Luca *et al.*, 1998) and the NADH-ubiquinone oxidoreductase of respiratory chains (Yano *et al.*, 1994).

In the *A. vinelandii* genome (Setubal *et al.*, 2009), several genes encode proteins associated with nitrogen fixation. The dinitrogen reductase in the nitrogenase system is the MoFe protein. The major *nif* cluster of genes contains the nitrogenase structural genes, including those for MoFe and most of the nitrogenase-assembly factors. FeSI is in the middle of the major *nif* cluster, just after the *nifN*, *nifE* and *nifX* genes which encode nitrogenase-assembly factors. FeSI is predicted to be cotranscribed with a [4Fe-4S] ferredoxin-like protein and several other proteins of unknown function (Fig. 1*a*). Moreover, the transcript level of FeSI was reported to increase 2.3-fold under nitrogen-fixing conditions (Hamilton *et al.*, 2011). Therefore, given the genomic location and transcript data, it is likely that FeSI plays some role in nitrogen fixation.



**Figure 1**  
 (a) Part of the major *nif* region showing the *nifENX* operon (green) and the operon encoding FeSI (blue). Genes of known function are labelled. FeSI is immediately preceded by Avin\_01510, a predicted (4Fe-4S) ferredoxin. (b) Cartoon view of the *A. vinelandii* FeSI homodimer. Subunits are shown with the [2Fe-2S] clusters and ligating cysteine residues shown as sticks. The right subunit is coloured according to the secondary structure. Strands (S1–S4), helices (H1–H4) and loops are shown in magenta, cyan and salmon, respectively. The residues (His7, Phe9, Thr53, Gly55 and Tyr69) at the dimer interface are shown as sticks. Hydrogen bonds are shown as black dashed lines.

**Table 1**  
 Macromolecule-production information.

Source organism	<i>A. vinelandii</i> CA
DNA source	<i>A. vinelandii</i> CA genomic DNA
Forward primer (vector)	GGATCCACGCGGAACCAGACC
Reverse primer (vector)	GCCCGAAAGGAAGCTGAGTTGGCT
Forward primer (genomic DNA)	GGTCTGGTTCGCGTGGATCCATGGCCAAA CCCGAGTTCATATC
Reverse primer (genomic DNA)	AGCCAACTCAGCTTCCTTCGGGCCTACCA GATCTCGGCAGGGGT
Expression vector	pRSET-A
Expression host	<i>E. coli</i> (KRX)
Complete amino-acid sequence of the construct produced	MRGSHHHHHGLVPRGSMAPKPEFHIFICAQ NRPAGHPRGSCGAKGAEGVYNAFAQVLI QKNLNRIALTTTGCLGPCQAGANVLIY PGAVMYSWVEPADAAIIVEQHLLGGPEPY ADKLTPAEIIW

FeSI is homologous to the [2Fe-2S] ferredoxin from *Clostridium pasteurianum* (Chatelet & Meyer, 1999), which interacts with the *C. pasteurianum* nitrogenase MoFe protein (Golinelli *et al.*, 1997) and is in the *anf* (iron-only) nitrogenase gene cluster of the *C. pasteurianum* genome (Pyne *et al.*, 2014).

## 2. Materials and methods

### 2.1. Macromolecule production

The *fesI* gene (Avin\_01520) was amplified by PCR from *A. vinelandii* CA genomic DNA and cloned by Gibson assembly (Gibson, 2011) into a modified pRSET-A vector with a thrombin-cleavable 6×His tag. The primers (Invitrogen) used for the pRSET-A vector were 5'-GGATCCACGCGG AACCAGACC-3' (forward) and 5'-GCCCGAAAGGAAG CTGAGTTGGCT-3' (reverse), and those for the genomic DNA were 5'-GGTCTGGTTCGCGTGGATCCATGGCCA AACCCGAGTTCATATC-3' (forward) and 5'-AGCCAA CTCAGCTTCCTTTCGGGCCTACCAGATCTCGGCAGG GGT-3' (reverse). The FeSI-pRSET-A plasmid was transformed into *Escherichia coli* KRX cells. The cells were grown in 1 l Terrific Broth to an OD of 0.6–0.8 and were then induced with 0.1% (*w/v*) rhamnose. The cells were grown at 18°C for 18 h after induction, spun at 4000g for 15 min, resuspended in 50 mM Tris-HCl pH 7.9, 150 mM NaCl and disrupted by sonication. Cell debris was removed by centrifugation and filtration. The supernatant was loaded onto a nickel resin affinity column (Generson) and eluted with 500 mM imidazole in 50 mM Tris-HCl pH 7.9, 150 mM NaCl. The His tag was cleaved by adding 50 U thrombin (Sigma) and incubating at 4°C overnight. Finally, the protein was concentrated to ~15 mg ml<sup>-1</sup> for crystallization trials. Macromolecule-production information is summarized in Table 1.

### 2.2. Crystallization

Thin plate-shaped crystals, which belonged to space group *P*<sub>2</sub><sub>1</sub>, were obtained by sitting-drop vapour diffusion using a reservoir solution consisting of 0.1 M HEPES pH 7.5, 70% (*v/v*) MPD (2,4-methylpentanediol). Crystals were cryo-protected in the mother liquor with ~30% (*v/v*) polyethylene

**Table 2**  
Crystallization.

Method	Sitting-drop vapour diffusion
Plate type	96-well plate
Temperature (K)	290
Protein concentration (mg ml <sup>-1</sup> )	15
Buffer composition of protein solution	50 mM Tris-HCl pH 7.9, 150 mM NaCl
Composition of reservoir solution	0.1 M HEPES pH 7.5, 70%(v/v) MPD
Volume and ratio of drop	200 nl (1:1) and 300 nl (2:1)
Volume of reservoir (μl)	50

**Table 3**  
Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	I03, Diamond Light Source
Wavelength (Å)	1.734
Temperature (K)	100
Detector	PILATUS
Crystal-to-detector distance (mm)	185
Rotation range per image (°)	0.2
Total rotation range (°)	180
Exposure time per image (s)	0.1
Space group	<i>P</i> 12 <sub>1</sub> 1
<i>a</i> , <i>b</i> , <i>c</i> (Å)	39.04, 60.10, 45.51
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 109.37, 90
Resolution range (Å)	42.93–2.11 (2.16–2.11)
No. of unique reflections	11386 (859)
Completeness (%)	98.7 (97.9)
Multiplicity	3.1 (3.0)
$\langle I/\sigma(I) \rangle$	8.1 (2.4)
Overall <i>B</i> factor from Wilson plot (Å <sup>2</sup> )	21.4

glycol (PEG) 400 and then flash-cooled in liquid nitrogen. Crystallization information is summarized in Table 2.

### 2.3. Data collection and processing

X-ray diffraction data were collected from cryoprotected crystals at 100 K on beamline I03 at Diamond Light Source, UK. For phase determination, a data set was collected at the Fe *K*-edge at a wavelength of 1.734 Å. The collected data were processed and scaled with *xia2* (Winter, 2010) using the *3dii* (*XDS*) setting (Kabsch, 2010). Data-collection and processing statistics are given in Table 3.

### 2.4. Structure solution and refinement

Phases were calculated with the *AUTOSHARP* pipeline (Vonrhein *et al.*, 2007) and an initial model was built with *ARP/wARP* (Langer *et al.*, 2008). Cycles of model building and refinement with noncrystallographic symmetry restraints were performed in *REFMAC5* (Murshudov *et al.*, 2011) and *Coot* (Emsley *et al.*, 2010). *MolProbity* (Williams *et al.*, 2018) was used for validation. The final refinement statistics are presented in Table 4. Figures were drawn with *PyMOL* (DeLano, 2002).

## 3. Results and discussion

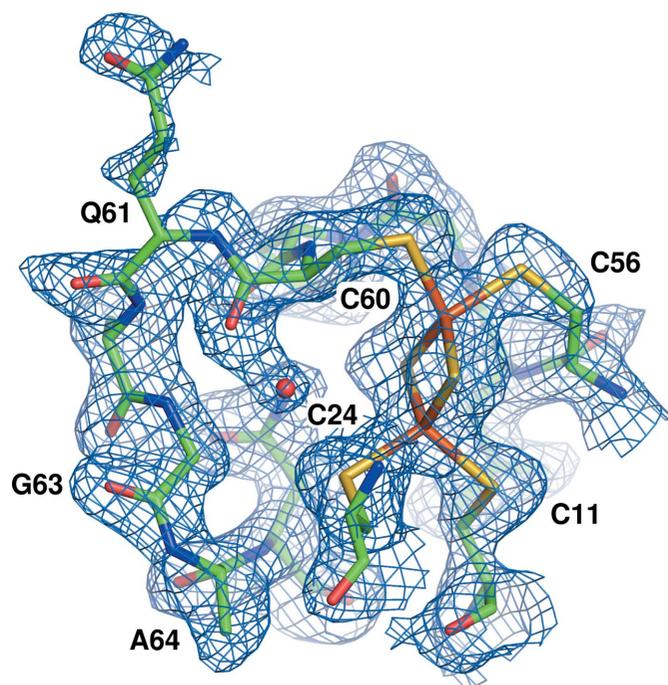
The *A. vinelandii* FeSI structure is a homodimer burying 1420 Å<sup>2</sup> as calculated with the *PDBePISA* service at EMBL-EBI (Krissinel & Henrick, 2007). It consists of a sheet with

**Table 4**  
Structure refinement.

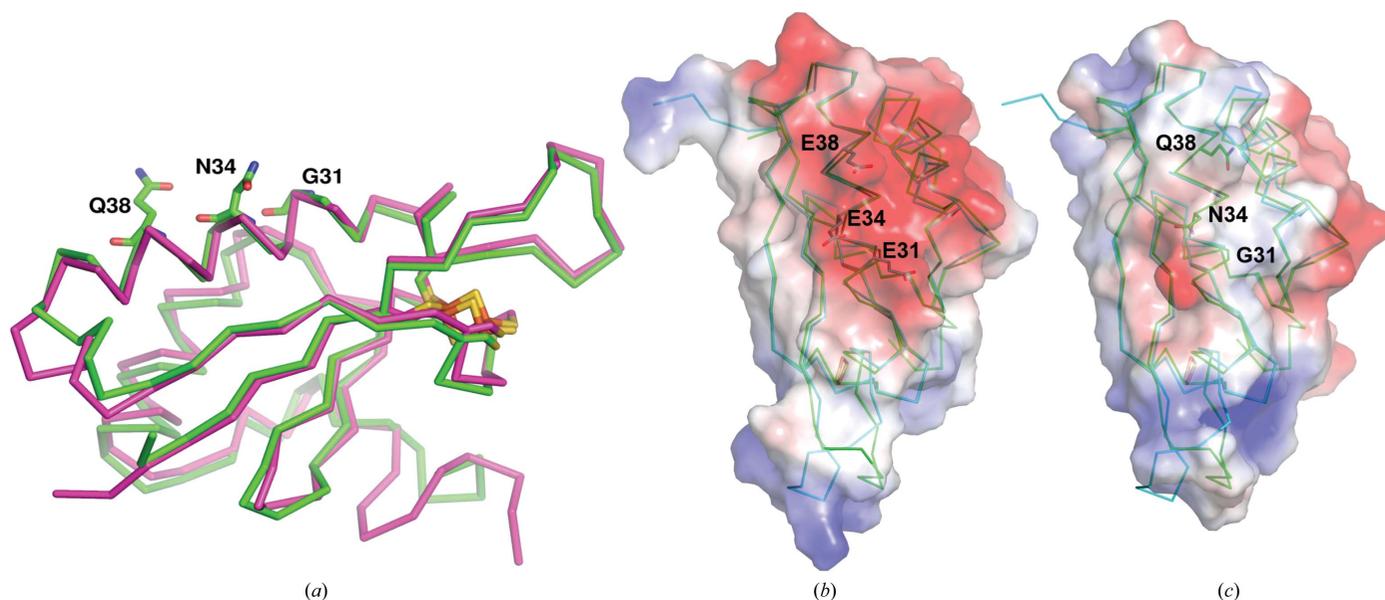
Values in parentheses are for the outer shell.

Resolution range (Å)	42.9–2.1 (2.165–2.110)
Completeness (%)	98.5 (97.9)
No. of reflections, working set	11386 (827)
No. of reflections, test set	545 (29)
Final <i>R</i> <sub>cryst</sub>	0.177
Final <i>R</i> <sub>free</sub>	0.231
No. of non-H atoms	
Protein	1558
Ligand	8
Solvent	128
Total	1694
R.m.s. deviations	
Bonds (Å)	0.024
Angles (°)	2.373
Average <i>B</i> factors (Å <sup>2</sup> )	
Protein	35.6
Ligand	23.4
Water	43.0
Ramachandran plot	
Favoured regions (%)	97.0
Additionally allowed (%)	3.0

four β-strands, two long α-helices that pack against the sheet, two short α-helices and several loops. The sheet from the monomer packs against its symmetry mate, forming the dimer interface. On one side of the dimer interface, two hydrogen bonds in parallel strands (S1 and S3) between the side chains of His7 and Tyr69 of two chains and, on the other side, hydrogen bonding between the main chains of Thr53 and Gly55 on strand S2 of two chains, along with the interaction of Phe9 with its symmetry mate through π-stacking, favour the dimerization (Fig. 1*b*). The two chains are similar, with an



**Figure 2**  
FeSI electron density ( $2F_o - F_c$ ) contoured at  $1\sigma$  around the [2Fe–2S] cluster in chain *A*, with residues of interest labelled. The Gly63–Ala64 bond is a *cis*-peptide.


**Figure 3**

(a) Superposition of FeSI from *A. vinelandii* (green) and the [2Fe–2S] ferredoxin from *A. aeolicus* (magenta; PDB entry 1f37). In FeSI, Gly31, Asn34 and Gln38 are shown as sticks. (b) Electrostatic surface representation of the predicted model of the *C. pasteurianum* [2Fe–2S] protein and its superposition with FeSI, shown as cyan (*C. pasteurianum* [2Fe–2S] protein) and green (FeSI) ribbons. The key residues (Glu31, Glu34 and Glu38) forming the negatively charged surface are highlighted as sticks. (c) Electrostatic surface representation of FeSI and its superposition with the model of the *C. pasteurianum* [2Fe–2S] protein, shown as coloured ribbons. The key residues (Gly31, Asn34 and Gln38) are highlighted as sticks.

r.m.s.d. of 0.1 Å over 104 C $\alpha$  atoms. Both clusters are coordinated in the same way and were refined at 100% occupancy. The [2Fe–2S] cluster is positioned between two loops that are located between strand S1 and helix H1 and between strand S2 and helix H2. It is bound to four cysteine residues: Cys11, Cys24, Cys56 and Cys60. There is a nonproline *cis*-peptide, well resolved in the density, between Gly63 and Ala64 (Fig. 2), which are Gly–Pro in many homologous sequences (Supplementary Fig. S1).

The most similar protein to FeSI with a known structure is the [2Fe–2S] ferredoxin from *Aquifex aeolicus* (Chatelet *et al.*, 1999; Yeh *et al.*, 2000), with an r.m.s.d. of 2 Å over one chain and 38% sequence identity. *A. aeolicus* does not fix nitrogen, so its thioredoxin-like ferredoxin must have a function that is unrelated to nitrogen fixation.

FeSI has 31% sequence identity to the *C. pasteurianum* ferredoxin (Chatelet & Meyer, 1999), which has no experimental structure. The structure of *C. pasteurianum* ferredoxin was predicted in *trRosetta* (Yang *et al.*, 2020) and the model has an r.m.s.d. of 0.6 Å to FeSI over one chain. *C. pasteurianum* ferredoxin specifically and strongly interacts electrostatically with the nitrogenase protein MoFe via three negatively charged residues: Glu31, Glu34 and Glu38 (Golignelli *et al.*, 1997). These residues are on an  $\alpha$ -helix on the opposite side of the protein to the [2Fe–2S] cluster. In FeSI, the equivalent residues are Gly31, Asn34 and Gln38, all of which are on the outside face of the first helix of the structure (Fig. 3). The change in charge from three negative glutamates to a neutral glutamine, asparagine and glycine suggests that it is unlikely that FeSI interacts directly with *A. vinelandii* MoFe by this mechanism and that its involvement in nitrogen fixa-

tion is of a different nature to the *C. pasteurianum* protein. Moreover, previous cross-linking experiments did not find any interaction of FeSI with the *C. pasteurianum* or *A. vinelandii* MoFe proteins (Chatelet & Meyer, 1999).

The FeSI structure completes the structural picture of the original two *Azotobacter* FeS proteins purified by Shethna in the 1960s. The structure of the FeSI protein from *A. vinelandii* appears to be a typical thioredoxin-like ferredoxin and will provide information for understanding its function in relation to nitrogen fixation and its evolutionary relationships to other ferredoxins.

### Acknowledgements

We thank Diamond Light Source for X-ray beam time (proposal mx9424) and the staff of beamlines I03 and I04 for assistance with crystal testing and data collection. The crystallization facility at Imperial College was funded by BBSRC (BB/D524840/1) and the Wellcome Trust (202926/Z/16/Z).

### Funding information

This work was funded in part by grant BB/L011468/1 from the Biotechnology and Biological Sciences Research Council/National Science Foundation Nitrogen Ideas Lab: Oxygen-Tolerant Nitrogenase. BVK is supported by a TUBİTAK 2232 Fellowship (Project No. 118C225). CARC was supported by a BBSRC Doctoral Training Programme grant (BB/F017324/1).

### References

Appel, J. & Schulz, R. (1996). *Biochim. Biophys. Acta*, **1298**, 141–147.

- Chatelet, C., Gaillard, J., Pétilot, Y., Louwagie, M. & Meyer, J. (1999). *Biochem. Biophys. Res. Commun.* **261**, 885–889.
- Chatelet, C. & Meyer, J. (1999). *J. Biol. Inorg. Chem.* **4**, 311–317.
- DeLano, W. L. (2002). *PyMOL*. <http://www.pymol.org>.
- De Luca, G., Asso, M., Bélaïch, J.-P. & Dermoun, Z. (1998). *Biochemistry*, **37**, 2660–2665.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Gibson, D. G. (2011). *Methods Enzymol.* **498**, 349–361.
- Golinelli, M.-P., Gagnon, J. & Meyer, J. (1997). *Biochemistry*, **36**, 11797–11803.
- Hamilton, T. L., Ludwig, M., Dixon, R., Boyd, E. S., Dos Santos, P. C., Setubal, J. C., Bryant, D. A., Dean, D. R. & Peters, J. W. (2011). *J. Bacteriol.* **193**, 4477–4486.
- Jones, C. W., Brice, J. M., Wright, V. & Ackrell, B. A. C. (1973). *FEBS Lett.* **29**, 77–81.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 133–144.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nat. Protoc.* **3**, 1171–1179.
- Maier, R. J. & Moshiri, F. (2000). *J. Bacteriol.* **182**, 3854–3857.
- Moshiri, F., Kim, J. W., Fu, C. & Maier, R. J. (1994). *Mol. Microbiol.* **14**, 101–114.
- 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.
- Peters, J. W., Fisher, K. & Dean, D. R. (1995). *Annu. Rev. Microbiol.* **49**, 335–366.
- Pyne, M. E., Utturkar, S., Brown, S. D., Moo-Young, M., Chung, D. A. & Chou, C. P. (2014). *Genome Announc.* **2**, e00790-14.
- Robson, R. L. (1979). *Biochem. J.* **181**, 569–575.
- Saarinen, M., Gleason, F. K. & Eklund, H. (1995). *Structure*, **3**, 1097–1108.
- Schlesier, J., Rohde, M., Gerhardt, S. & Einsle, O. (2016). *J. Am. Chem. Soc.* **138**, 239–247.
- Setubal, J. C., dos Santos, P., Goldman, B. S., Ertesvåg, H., Espin, G., Rubio, L. M., Valla, S., Almeida, N. F., Balasubramanian, D., Cromes, L., Curatti, L., Du, Z., Godsy, E., Goodner, B., Hellner-Burris, K., Hernandez, J. A., Houmiel, K., Imperial, J., Kennedy, C., Larson, T. J., Latreille, P., Ligon, L. S., Lu, J., Maerk, M., Miller, N. M., Norton, S., O'Carroll, I. P., Paulsen, I., Raulfs, E. C., Roemer, R., Rosser, J., Segura, D., Slater, S., Stricklin, S. L., Studholme, D. J., Sun, J., Viana, C. J., Wallin, E., Wang, B., Wheeler, C., Zhu, H., Dean, D. R., Dixon, R. & Wood, D. (2009). *J. Bacteriol.* **191**, 4534–4545.
- Shethna, Y. I., DerVartanian, D. V. & Beinert, H. (1968). *Biochem. Biophys. Res. Commun.* **31**, 862–868.
- Shethna, Y. I., Wilson, P. W., Hansen, R. E. & Beinert, H. (1964). *Proc. Natl Acad. Sci. USA*, **52**, 1263–1271.
- Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.
- Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (2018). *Protein Sci.* **27**, 293–315.
- Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
- Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S. & Baker, D. (2020). *Proc. Natl Acad. Sci. USA*, **117**, 1496–1503.
- Yano, T., Sled, V. D., Ohnishi, T. & Yagi, T. (1994). *Biochemistry*, **33**, 494–499.
- Yeh, A. P., Chatelet, C., Soltis, S. M., Kuhn, P., Meyer, J. & Rees, D. C. (2000). *J. Mol. Biol.* **300**, 587–595.
- Zanetti, G. & Pandini, V. (2013). *Encyclopedia of Biological Chemistry*, 2nd ed., edited by W. J. Lennarz & M. D. Lane, pp. 296–298. New York: Academic Press.