**SUPPLEMENTARY INFORMATION FOR:**

Fully autonomous characterization and data collection from crystals of biological macromolecules

**AUTHORS AND AFFILIATIONS:**

Stephanie Hutin1\*, Bart Van Laer1\*, Christoph Mueller-Dieckmann1, Gordon Leonard1, Didier Nurizzo1 and Matthew W. Bowler2.

1European Synchrotron Radiation Facility, Structural Biology Group, Grenoble, France

2European Molecular Biology Laboratory, Grenoble Outstation, Grenoble, France

\* These authors contributed equally to this work.

Corresponding Author:

Bowler, Matthew

European Molecular Biology Laboratory

Grenoble Outstation,

France

[mbowler@embl.fr](mailto:zander@embl.fr)

**Protein expression, purification and crystallization adapted from Macherel *et al.***1

* 1. Express and purify GCSH
     1. Inoculate three 1l LB flasks with BL21(DE3) *E. coli* cells containing a pET21d vector with the gene for the GCSH construct (His-tagged GCSH without the mitochondrial signal sequence (residue 49-173)). Add 100 µM lipoic acid (from a 1M stock in dimethylsulfoxide) and 100 µg/ml ampicillin to the media and grown cells at 37°C until an optical density at 600 nm of 0.6 is reached.

* + 1. Induce the expression of the protein by addition of 0.4 mM IPTG.
    2. Add 34 µg/mL chloramphenicol after 3 hours to stop the expression and enhance the lipoylation of GCSH. Maintain the cultures overnight at 20°C.
    3. Harvest the bacterial cells (5000 x g, 10 min) and sonicate (10 sec on/20 sec off for a total time of 10 minutes at 40% amplitude, 20 kHz) in the presence of 50 mL 50 mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT, 1 mg DNase, 1 mg lysozyme and EDTA-free protease inhibitors.
    4. Purify the protein on a 5ml Ni-sepharose column (GE Histrap HP) and elute in 50 mM imidazole in the same buffer.
    5. Pool the fractions containing protein and concentrate to 5 mL using a 15 kDa molecular weight-cutoff (MWCO) concentrator at 4000 g.
    6. Dialyzed overnight at 4°C against 100 ml 20 mM Tris at pH 8.0, 100 mM NaCl using a 3 kDa molecular weight-cutoff (MWCO) cellulose dialysis membrane.
    7. Perform an anion exchange chromatography (GE MonoQ 5/50 GL) using a gradient of 100 mM to 1M NaCl in 20 mM Tris at pH 8 over 30 minutes at a 1 ml/min flow rate.
    8. Concentrate the protein peaks to 6.5 mg/mL using a 15 kDa molecular weight-cutoff (MWCO) concentrator at 4000 g and send to a mass spectrometry facility to find the protein fraction containing purified non-lipoylated GCSH (15562 Da).
  1. Crystallization procedure
     1. Mix 25 µl of GCSH solution with 20 µl 0.5 M Sodium formate pH 4.0 and store the sample at 4°C for a week.
        1. Note: Crystal needles will appear.

**Determination of phases by molecular replacement and refinement**

1. Open the graphical interface of CCP42.
   1. Determine the Matthew’s coefficient, indicating the number of molecules in the asymmetric unit by opening the Matthews\_coeff tab3.
      1. Press “Browse” and select the .mtz file of the collected data set.
      2. Enter the molecular weight (15562Da) in the interface.
      3. Press “Run now” and read the solvent content analysis to select most likely the composition of the asymmetric unit (P(tot) close to 1).
      4. Press “close2.
   2. Open the “Phaser MR” tab in the CCP4 interface4.
      1. Enter a job title.
      2. Import again the .mtz file using “browse”.
      3. In the “Define ensembles (models) section enter an ensemble name and import the Protein Data Bank entry of the search model (bovine H-protein, PDB code: 3klr).
      4. Give the sequence identity (which is 98%, but due to stringency we used just 90%).
      5. Give the number of molecules in the asymmetric unit (1) in the “Define composition of the asymmetric unit” section and import a sequence file (.seq).
      6. Select the search model in the “Search parameters” section and the number of copies to search for (1).
      7. Press “Run now”.

Note: The TFZ score should be at least 6 but above 8 indicates a good solution. The data yielded a single solution with a TFZ score of 64.1 and without any clashes).

* 1. Open a visualization program (Coot5) and inspect the density map.
  2. Select refmac56 in the program list of ccp4. Run a restrained refinement (default setting).
  3. Import the mtz file and pdb file (output of Phaser) using `browse` in the corresponding field. Optional select “run Coot: find waters” to automatically add waters in the structure.
  4. Press `run now`

**References**

1. Macherel, D., Bourguignon, J., Forest, E., Faure, M., Cohen-Addad, C., Douce, R. Expression, lipoylation and structure determination of recombinant pea H-protein in Escherichia coli. *European Journal of Biochemistry*. **236** (1), 27–33 (1996).

2. Winn, M.D. *et al.* Overview of the *CCP* 4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography*. **67** (4), 235–242, doi: 10.1107/S0907444910045749 (2011).

3. Matthews, B.W. Solvent content of protein crystals. *Journal of Molecular Biology*. **33** (2), 491–497 (1968).

4. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J. *Phaser* crystallographic software. *Journal of Applied Crystallography*. **40** (4), 658–674, doi: 10.1107/S0021889807021206 (2007).

5. Emsley, P., Cowtan, K. *Coot* : model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography*. **60** (12), 2126–2132, doi: 10.1107/S0907444904019158 (2004).

6. Murshudov, G.N. *et al.* *REFMAC* 5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D Biological Crystallography*. **67** (4), 355–367, doi: 10.1107/S0907444911001314 (2011).