Solving the first novel protein structure by 3D micro-crystal electron diffraction

H. Xu¹,*,†, H. Lebrette²,†, M.T.B. Clabbers¹,†, J. Zhao¹, J.J. Griese²,³, X. Zou¹,* and M. Högbom²,*

Affiliations:
¹Department of Materials and Environmental Chemistry, Stockholm University, 106 91 Stockholm, Sweden
²Department of Biochemistry and Biophysics, Stockholm University, 106 91 Stockholm, Sweden
³Department of Cell and Molecular Biology, Uppsala University, 751 24 Uppsala, Sweden

*Correspondence to: hongyi.xu@mmk.su.se (H.X), hogbom@dbb.su.se (M.H) and xzou@mmk.su.se (X.Z)
†These authors contributed equally to this work.

Abstract: Micro-crystal electron diffraction (MicroED) has recently shown potential for structural biology. It enables studying biomolecules from micron-sized 3D crystals that are too small to be studied by conventional X-ray crystallography. However, to the best of our knowledge, MicroED has only been applied to re-determine protein structures that had already been solved previously by X-ray diffraction. Here we present the first unknown protein structure – an R2lox enzyme – solved using MicroED. The structure was phased by molecular replacement using a search model of 35% sequence identity. The resulting electrostatic scattering potential map at 3.0 Å resolution was of sufficient quality to allow accurate model building and refinement. Our results demonstrate that MicroED has the potential to become a widely applicable tool for revealing novel insights into protein structure and function, opening up new opportunities for structural biologists.
Main Text:

Electrons, similar to X-rays and neutrons, are a powerful source for diffraction experiments (Henderson, 1995). Due to the strong interactions between electrons and matter, crystals that are considered as powder in X-ray crystallography can be treated as single crystals by micro-crystal electron diffraction (MicroED (Shi et al., 2013)). This enables structure determination of molecules from micron- to nanometer-sized 3D crystals that are too small for conventional X-ray diffraction (Shi et al., 2013; Nannenga, Shi, Leslie et al., 2014; Yonekura et al., 2015; Clabbers et al., 2017; Xu et al., 2018). Furthermore, MicroED can be applied to study biomolecules of low molecular weight that are beyond what can be resolved by single particle cryo-EM imaging (Henderson, 1995; Khoshouei et al., 2017).

Over the past decades, 3D ED methods have been developed for structure determination of small inorganic compounds (Mugnaioli et al., 2009; Wan et al., 2013; Gemmi et al., 2015) and organic molecules (Kolb et al., 2010; van Genderen et al., 2016; Palatinus et al., 2017). At the early stages of 3D ED method development, tilting of the crystal was done manually, while diffraction patterns were collected on negative film. It could take years before sufficient data were obtained and processed in order to determine the crystal structure (Zou et al., 2003). The computerization of transmission electron microscopes (TEM) and the development of CCD detectors allowed software to be developed that can semi-automatically collect 3D ED data in less than an hour (Wan et al., 2013; Kolb et al., 2010). Thanks to the recent advancement in CMOS and hybrid detector technology, it is now feasible to collect diffraction data in movie mode while continuously rotating the crystal (Nannenga, Shi, Leslie et al., 2014; Gemmi et al., 2015; Nederlof et al., 2013). Benefiting from these technological advances, data collection and structure refinement can now be performed within hours (Gruene et al., 2018; Jones et al., 2018). Furthermore, peptide structures have been solved ab-initio using high resolution MicroED data (Rodriguez et al., 2015; de la Cruz et al., 2017). Since 2013, several research groups have shown that it is feasible to re-determine already known protein structures using MicroED (table S1) and it is only very recently that a new polymorph of hen egg-white lysozyme was unveiled by MicroED, but again phased using a previously determined structure of the identical protein (Lanza et al., 2019).

The sample handling of MicroED is similar to that of cryo-EM, while the data collection and processing are similar to that used in X-ray crystallography, making the technique highly adaptable.
to existing cryo-EM and general TEM labs (Nannenga, Shi, Leslie et al., 2014; de la Cruz et al., 2017; Shi et al., 2013; Yonekura et al., 2015; Clabbers et al., 2017; Xu et al., 2018). Therefore, it is of great benefit to further develop MicroED methods to meet the needs of scientists in a wide community.

Here, we report the next step in the development of MicroED by solving a novel protein structure, Sulfolobus acidocaldarius R2-like ligand-binding oxidase, SaR2lox. The R2lox metalloenzyme family was discovered a decade ago from its similarities with the ribonucleotide reductase R2 protein (Andersson & Hogbom, 2009). Its physiological function is to date unknown, but the two crystal structures solved (Andersson & Hogbom, 2009; Griese et al., 2013) of proteins belonging to this family reveal a 4-helix bundle core accommodating a dinuclear metal cluster, characterized by electron paramagnetic resonance (EPR) as Mn(III)/Fe(III) (Shafaat et al., 2014), interacting with a long-chain fatty acid.

We demonstrate that MicroED data can be collected from SaR2lox 3D micro-crystals by the continuous rotation method (Nederlof et al., 2013; Nannenga, Shi, Leslie et al., 2014). Conventional X-ray crystallography software can be used directly for processing ED data (XDS (Kabsch, 2010)), determining the phases using a homologous protein model of 35% sequence identity (Phaser (McCoy et al., 2007)), and refining the model structure (phenix.refine (Afonine et al., 2012)). These results illustrate that MicroED is a powerful tool for determining novel protein structures with sample requirements complementing those of X-ray crystallography, single particle cryo-EM and X-ray free electron lasers.

Micron-sized 3D crystals of SaR2lox were grown using conventional hanging drop vapor diffusion, and are barely large enough to be distinguished under an optical microscope from precipitate and phase separation occurring in the drop (Fig. 1A). The mother liquor used for growing micro-crystals contains 44% polyethylene glycol 400. The viscosity of the sample prevented preparation of thin vitrified cryo-EM samples using the traditional deposit-blot-plunge routine. Thus, we used manual backside blotting to remove excess liquid and vitrified the sample in liquid ethane. The plate-like crystals are triangular in shape and a few micrometers in size (Fig. 1B). The thickness of the crystals is estimated to be less than 0.5 μm. We note that it is crucial to reduce the thickness of the protective layer of vitrified ice as much as possible in order to collect
electron diffraction data of high signal-to-noise ratio from the crystals. The crystals diffracted beyond 3.0 Å in resolution as shown in Fig. 1C.

MicroED data were collected by continuously rotating a single crystal in the electron beam at a rate of 0.45 °/second. The exposure time of each frame was 2 seconds, integrating over 0.90° of the reciprocal lattice. Data were typically collected over a total rotation range of 54° within 2 minutes. The electron dose rate applied was approximately 0.08 e−/Å²/s, resulting in a total dose of below 10 e−/Å² per dataset. Because of their plate-like shape, a morphology commonly observed in protein crystallography, the SaR2lox crystals are dispersed on the TEM grid with a preferred orientation. A total of 35 electron diffraction datasets were collected from the plate-like SaR2lox crystals. Data were processed using crystallography software XDS (Kabsch, 2010). We could determine that the crystals are orthorhombic with a primitive unit cell using rotation electron diffraction processing software (REDp) (Wan et al., 2013). Two screw axes were identified from the reflection conditions (Fig. 1C). Based on unit cell consistency and cross correlation, 21 out of the 35 datasets were merged. Due to the preferred orientation of the crystals, the data completeness increased from 50.9% for a single dataset to only 62.8% (Fig. 1D). However, the multiplicity (~32) and overall I/σ(I) (6.12 up to 3.0 Å resolution) improved drastically (table S2). We demonstrated previously that the resulting structural model can be improved by merging data from a large number of crystals (Xu et al., 2018).

The processed MicroED data were of sufficient quality to solve and refine the structure of SaR2lox (Fig. 2A). The closest homologue to SaR2lox in the Protein Data Bank (PDB) is an R2lox protein from Mycobacterium tuberculosis (MtR2lox) sharing 35% sequence identity (PDB ID 3EE4 (Andersson & Hogbom, 2009)) (fig. S1). Using a modified search model based on the sequence alignment, we were able to phase the merged MicroED data by molecular replacement using Phaser (McCoy et al., 2007), obtaining a clear single solution in space group P2₁2₁2₁. The structure was iteratively built and refined in COOT (Emsley & Cowtan, 2004) and phenix.refine (Afonine et al., 2012) (table S3). The refined electrostatic potential map provided sufficient detail to model side-chains and allowed rebuilding of the main chain (Fig. 2, B and D), even though certain parts of the maps were less well defined along the c* direction (owing to the incomplete data). To assess the data quality and eliminate the influence of model bias, we generated a composite omit map covering the entire unit cell, confirming the correct interpretation for modelling the structure (Fig. 2, C and E).
The final model of SaR2lox shows a protein backbone Cα root-mean-square deviation of 0.94 Å compared to the structure of MtR2lox used for molecular replacement. This value is within expected range from proteins of this size and sequence identity level. The structure presented here confirms the dimeric biological assembly and the ferritin-like helix bundle overall fold, previously seen in the R2lox protein family (Andersson & Hogbom, 2009; Griese et al., 2013) (fig. S2). Furthermore, the structure reveals new biochemically important features of the SaR2lox enzyme, differing from the two structures known in this family: the substrate-binding pocket is reshaped and shows an altered electrostatic potential distribution (Fig. 3). In addition, the sequence identity with existing structures is lower than 40% which indicates a possibly divergent enzyme function (Tian & Skolnick, 2003). Thus, our results suggest that SaR2lox has a different substrate specificity than the enzymes previously structurally characterized. Further biochemical studies are undergoing in our laboratory to confirm this hypothesis.

MicroED data can be collected from crystals previously considered to be insufficient in size, which is its main advantage over X-ray crystallography. There are several aspects where electron diffraction can be further improved. For instance, by developing sample preparation methods such as cryogenic focused ion beam (Duyvesteyn et al., 2018; Martynowycz et al., 2019; Zhou et al., 2019; Li et al., 2018), the orientation of the crystals may be precisely controlled to achieve near 100% data completeness. Furthermore, the crystallographic quality indicators of the MicroED data are not yet as accurate nor as precise as generally achieved by X-ray crystallography. Rapid development of data collection strategies, as well as improvements of TEM hardware, may further accelerate the development of MicroED. Our results show that MicroED can be used for determining novel protein structures, even under the circumstances frequently encountered in macromolecular crystallography, where viscous sample environment, preferred orientation, and poor signal-to-noise ratio complicate data collection and structure determination.
References and notes:

Adams, P. D., Afonine, P. V., Bunkóczki, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Crystallographica Section D Biological Crystallography*. 66, 213–221.

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Crystallographica Section D Biological Crystallography*. 68, 352–367.

Alam, A. & Jiang, Y. (2009). *Nature Structural & Molecular Biology*. 16, 35–41.

Andersson, C. S. & Hogbom, M. (2009). *Proceedings of the National Academy of Sciences*. 106, 5633–5638.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Crystallographica Section D Biological Crystallography*. 66, 12–21.

Clabbers, M. T. B., van Genderen, E., Wan, W., Wiegars, E. L., Gruene, T. & Abrahams, J. P. (2017). *Acta Crystallographica Section D Structural Biology*. 73, 738–748.

de la Cruz, M. J., Hattne, J., Shi, D., Seidler, P., Rodriguez, J., Reyes, F. E., Sawaya, M. R., Cascio, D., Weiss, S. C., Kim, S. K., Hinck, C. S., Hinck, A. P., Calero, G., Eisenberg, D. & Gonen, T. (2017). *Nature Methods*. 14, 399–402.

Diederichs, K. (2017). *Acta Crystallographica Section D Structural Biology*. 73, 286–293.

Duyvesteyn, H. M. E., Kotecha, A., Ginn, H. M., Hecksel, C. W., Beale, E. V., de Haas, F., Evans, G., Zhang, P., Chiu, W. & Stuart, D. I. (2018). *Proceedings of the National Academy of Sciences*. 115, 9569–9573.

Emsley, P. & Cowtan, K. (2004). *Acta Crystallographica Section D Biological Crystallography*. 60, 2126–2132.

English, A. C., Done, S. H., Caves, L. S. D., Groom, C. R. & Hubbard, R. E. (1999). *Proteins: Structure, Function, and Genetics*. 37, 628–640.

Evans, P. (2006). *Acta Crystallographica Section D Biological Crystallography*. 62, 72–82.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Crystallographica Section D Biological Crystallography*. 69, 1204–1214.

Foroughi, L. M., Kang, Y.-N. & Matzger, A. J. (2011). *Crystal Growth & Design*. 11, 1294–1298.

French, S. & Wilson, K. (1978). *Acta Crystallographica Section A*. 34, 517–525.

Gemmi, M., La Placa, M. G. I., Galanis, A. S., Rauch, E. F. & Nicolopoulos, S. (2015). *Journal of Applied Crystallography*. 48, 718–727.

van Genderen, E., Clabbers, M. T. B., Das, P. P., Stewart, A., Nederlof, I., Barentsen, K. C., Portillo, Q., Pannu, N. S., Nicolopoulos, S., Gruene, T. & Abrahams, J. P. (2016). *Acta Crystallographica Section A Foundations and Advances*. 72, 236–242.

Griese, J. J., Roos, K., Cox, N., Shafaat, H. S., Branca, R. M. M., Lehtio, J., Graslund, A., Lubitz, W., Siegbahn, P. E. M. & Hogbom, M. (2013). *Proceedings of the National Academy of Sciences*. 110, 17189–17194.

Gruene, T., Wennmacher, J. T. C., Zaubitzer, C., Holstein, J. J., Heidler, J., Fecteau-Lefebvre, A., De Carlo, S., Müller, E., Goldie, K. N., Regeni, I., Li, T., Santiso-Quiones, G., Steinfeld, G., Handschin, S., van Genderen, E., van Bokhoven, J. A., Clever, G. H. & Pantelic, R. (2018). *Angewandte Chemie International Edition*. 1298.

Guo, F., Zhou, W., Li, P., Mao, Z., Yennawar, N. H., French, J. B. & Huang, T. J. (2015). *Small*. 11, 2733–2737.

Hattne, J., Shi, D., de la Cruz, M. J., Reyes, F. E. & Gonen, T. (2016). *Journal of Applied Crystallography*. 49, 1029–1034.

Hattne, J., Shi, D., Glynn, C., Zee, C.-T., Gallagher-Jones, M., Martynowycz, M. W., Rodriguez, J. A. & Gonen, T. (2018). *Structure*. 26, 759-766.e4.

Henderson, R. (1995). *Quarterly Reviews of Biophysics*. 28, 171.

Jones, C. G., Martynowycz, M. W., Hattne, J., Fulton, T. J., Stoltz, B. M., Rodriguez, J. A., Nelson, H. M. & Gonen, T. (2018). *ACS Central Science*. 4, 1587–1592.

Kabsch, W. (2010). *Acta Crystallographica Section D Biological Crystallography*. 66, 125–132.

Karplus, P. A. & Diederichs, K. (2012). *Science*. 336, 1030–1033.

Khoshouei, M., Radjainia, M., Baumeister, W. & Daney, R. (2017). *Nature Communications*. 8, 16099.

Kim, S. K., Barron, L., Hinck, C. S., Petrunak, E. M., Cano, K. E., Thangirala, A., Iskra, B., Brothers, M., Vonberg, M., Leal, B., Richter, B., Kodal, R., Taylor, A. B., Du, S., Barnes, C. O., Sulea, T., Calero, G., Hart, P. J., Hart, M. J., Demeler, B. & Hinck, A. P. (2017). *Journal of Biological Chemistry*. 292, 7173–7188.

Kolb, U., Gorelik, T. E., Mugnaioli, E. & Stewart, A. (2010). *Polymer Reviews*. 50, 385–409.

Krisp, E. & Henrick, K. (2004). *Acta Crystallographica Section D Biological Crystallography*. 60, 2256–2268.

Lanza, A., Margheritis, E., Mugnaioli, E., Cappello, V., Garau, G. & Gemmi, M. (2019). *IUCrJ*. 6,
Li, X., Zhang, S., Zhang, J. & Sun, F. (2018). *Biophysics Reports*. **4**, 339–347.

Liu, S. & Gonen, T. (2018). *Communications Biology*. 1.

Martynowycz, M. W., Zhao, W., Hattne, J., Jensen, G. J. & Gonen, T. (2019). *Structure*.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *Journal of Applied Crystallography*. **40**, 658–674.

Mugnaioli, E., Gorelik, T. & Kolb, U. (2009). *Ultramicroscopy*. **109**, 758–765.

Nannenga, B. L., Shi, D., Hattne, J., Reyes, F. E. & Gonen, T. (2014). *ELife*. 3.

Nannenga, B. L., Shi, D., Leslie, A. G. W. & Gonen, T. (2014). *Nature Methods*. 11, 927–930.

Nederlof, I., van Genderen, E., Li, Y.-W. & Abrahams, J. P. (2013). *Acta Crystallographica Section D Biological Crystallography*. **69**, 1223–1230.

Palatinus, L., Brzda, P., Boullay, P., Perez, O., Klementová, M., Petit, S., Eigner, V., Zaarour, M. & Mintova, S. (2017). *Science*. **355**, 166–169.

Purdy, M. D., Shi, D., Chrustowicz, J., Hattne, J., Gonen, T. & Yeager, M. (2018). *Proceedings of the National Academy of Sciences*. **115**, 13258–13263.

Rodriguez, J. A., Ivanova, M. I., Sawaya, M. R., Cascio, D., Reyes, F. E., Shi, D., Sangwan, S., Guenther, E. L., Johnson, L. M., Zhang, M., Jiang, L., Arbing, M. A., Nannenga, B. L., Hattne, J., Whitelegge, J., Brewster, A. S., Messerschmidt, M., Boutet, S., Sauter, N. K., Gonen, T. & Eisenberg, D. S. (2015). *Nature*. **525**, 486–490.

Sauter, C., Otalora, F., Gavira, J.-A., Vidal, O., Giege, R. & Garcia-Ruiz, J. M. (2001). *Acta Crystallographica Section D Biological Crystallography*. **57**, 1119–1126.

Shafaat, H. S., Griese, J. J., Pantazis, D. A., Roos, K., Andersson, C. S., Popovic-Bijelic, A., Graslund, A., Siegbahn, P. E. M., Neese, F., Lubitz, W., Högberg, M. & Cox, N. (2014). *Journal of the American Chemical Society*. **136**, 13399–13409.

Sharma, P., Verma, N., Singh, P. K., Korpole, S. & Ashish (2016). *Scientific Reports*. 6.

Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. (2013). *ELife*. 2.

Tian, W. & Skolnick, J. (2003). *Journal of Molecular Biology*. **333**, 863–882.

Toyoshima, C., Nakasako, M., Nomura, H. & Ogawa, H. (2000). *Nature*. **405**, 647–655.

Wagner, J. M., Zadrozyň, K. K., Chrustowicz, J., Purdy, M. D., Yeager, M., Ganser-Pornillos, B. K. & Pornillos, O. (2016). *ELife*. 5.

Walter, J., Steigemann, W., Singh, T. P., Bartunik, H., Bode, W. & Huber, R. (1982). *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry*. **38**, 1462–1472.

Wan, W., Sun, J., Su, J., Hovmöller, S. & Zou, X. (2013). *Journal of Applied Crystallography*. **46**, 1863–1873.

Warkentin, M., Badeau, R., Hopkins, J. B. & Thorne, R. E. (2012). *Acta Crystallographica Section D Biological Crystallography*. **68**, 1108–1117.

Watanabe, N. (2006). *Acta Crystallographica Section D Biological Crystallography*. **62**, 891–896.

Xu, H., Lebrette, H., Yang, T., Srinivas, V., Hovmöller, S., Högbom, M. & Zou, X. (2018). *Structure*. **26**, 667–675.e3.

Yonekura, K., Kato, K., Ogasawara, M., Tomita, M. & Toyoshima, C. (2015). *Proceedings of the National Academy of Sciences*. **112**, 3368–3373.

Yonekura, K. & Maki-Yonekura, S. (2016). *Journal of Applied Crystallography*. **49**, 1517–1523.

Zhou, H., Luo, Z. & Li, X. (2019). *Journal of Structural Biology*.

Zou, X. D., Mo, Z. M., Hovmöller, S., Li, X. Z. & Kuo, K. H. (2003). *Acta Crystallographica Section A Foundations of Crystallography*. **59**, 526–539.
Acknowledgments: H.L. thanks P. Stenmark for discussions about this manuscript. Funding: We acknowledge financial support from the Knut and Alice Wallenberg Foundation through the project grants 3DEM-NATUR (no. 2012.0112, X.D.Z.) and Wallenberg Academy Fellows (no. 2017.0275, M.H.), the Science for Life Laboratory through the pilot project grant Electron Nanocrystallography, and the Swedish Research Council (2017-04018, M.H.; 2017-05333, H.Y.X.).

Author contributions: H.X contributed to project design, conception, electron diffraction data collection, electron diffraction data analysis, manuscript writing and figure making. H.L contributed to project design, crystal growth, structure determination, manuscript writing and figure making. M.C. contributed to electron diffraction data analysis, structure determination, manuscript writing and figure making. J.Z. contributed to electron diffraction data collection. J.G. contributed to cloning. X.Z. and M.H. contributed to project design, data analysis, conception and manuscript writing;

Competing interests: The authors declare no competing interests;

Data and materials availability: The atomic coordinates of the SaR2lox structure are deposited in the Protein Data Bank under accession code 6QRZ. Raw MicroED data are available upon request. The python script used for converting diffraction frames from TIF to MRC and IMG (SMV format) is available upon request. REDp can be downloaded at mmk.su.se/zou.
Fig. 1 Overview of the MicroED experiment.

(A) SaR2lox micro-crystals (pointed out by arrows) viewed under optical microscope. (B) Over-focused TEM image of a typical diffracting SaR2lox crystal preserved in vitrified ice. The volume of this particular crystal is estimated to be approximately 2 µm³, which contains approximately 6 × 10⁶ unit cells. (C) Typical electron diffraction pattern of the SaR2lox crystal. The frame was taken after an accumulated electron dose of 4.3 e⁻/Å². (D) Reconstructed reciprocal lattice showing limited data completeness with predominately missing reflections in the direction of c*, missing observations are shown in white, observed reflections in rainbow color scheme, systematic absences are shown in pink.
Fig. 2 High quality electrostatic potential maps allowing accurate model interpretation.

(A) Overall structure of SaR2lox solved by MicroED with three colored selections as examples for showing electrostatic potential maps in (B) to (E). Electrostatic scattering potential maps 2Fo-Fc (contoured at 1 σ, colored in blue) and Fo-Fc (contoured at ±3 σ, colored in green and red for positive and negative peaks, respectively) and simulated annealing composite omit maps (contoured at 1 σ, colored in magenta) are shown for residues 164 to 177 (orange) and 235 to 249 (yellow) in (B) and (C), respectively, and for residues 202 to 217 (cyan) in (D) and (E), respectively. Simulated annealing composite omit electrostatic potential maps are calculated with sequential 5% fractions of the structure omitted. Only observed reflections were used for map calculations, i.e. no missing F(obs) were restored using a weighted F(calc). Despite the low completeness, the data produces high quality well resolved maps. Oxygen and nitrogen atoms are colored red and blue, respectively, carbons are colored according to the selection previously mentioned.
Fig. 3 Differences between substrate-binding pockets support substrate specificity divergences in the R2lox metalloenzyme family.

The molecular surface and identity of residues defining the substrate-binding pocket in the SaR2lox structure (A) are compared with the two already known structures of this protein family, (B) MtR2lox (PDB ID 3EE4 (Andersson & Hogbom, 2009)) and (C) GkR2lox (PDB ID 4HR0 (Griese et al., 2013)). Although the Mn- and Fe-coordinating residues are fully conserved, the distal residues lining the cavity are remarkably divergent, impacting drastically the shape and electrostatic contact potentials of the pocket accommodating the putative substrate. This suggests that the substrate specificity could be different between these three enzymes, and thus their function could be divergent. Electrostatic contact potentials plotted on molecular surfaces are colored in a gradient from red (negative) to blue (positive). Ligands modelled as myristic acid and palmitic acid for MtR2lox and GkR2lox, respectively, are colored in yellow. A ligand was not modelled in the SaR2lox structure due to a too weak signal in the corresponding scattering potential map but its presence cannot be excluded. Carbons, nitrogen, oxygen, manganese and iron atoms are colored in grey, blue, red, purple and orange, respectively. We note that the missing stretch of unmodeled residues in SaR2lox (250 to 260) is not shared with MtR2lox or GkR2lox and is likely to impact distal closure of the cavity, further emphasizing differences between proteins.
Supplementary Materials for

Solving the first novel protein structure by 3D micro-crystal electron diffraction

H. Xu*,†, H. Lebrette*, M.T.B. Clabbers†, J. Zhao, J.J. Griese, X. Zou*, M. Högbom*. 

Correspondence to: hongyi.xu@mmk.su.se (H.X.), hogbom@dbb.su.se (M.H.), xzou@mmk.su.se (X.Z.)
Materials and Methods

Cloning, Expression and Protein Purification

A construct encoding full-length *Sulfolobus acidocaldarius* R2lox (accession number WP_011278976) was PCR-amplified from genomic DNA (DSM number 639, obtained from DSMZ) and inserted into pET-46 Ek/LIC (Novagen) using the following primers:

GACGACGACAGATGAAGGAAAAATTACTTGAATTCAGAAGT
GAGGAGAAGCCCGGTTATTTGTCCAGCTTAATCTCCTATGAC.

Expression was carried out in *Escherichia coli* BL21(DE3) (Novagen). Cells were cultured at 37°C in terrific broth medium (Formedium) supplemented with 50 µg/ml ampicillin, in a bench-top bioreactor system (Harbinger). When an optical density at 600 nm of 0.7 was reached, expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and allowed to continue for 3 h. Then, cells were harvested by centrifugation and stored at -80 °C. The N-terminal His6 tagged SaR2lox protein was purified via immobilized metal ion affinity chromatography and size-exclusion chromatography. Cells were resuspended in buffer A (25 mM Hepes-Na pH 7.0, 300 mM NaCl, and 20 mM imidazole) and disrupted by high-pressure homogenization. The lysate was cleared by centrifugation and applied to a nickel-nitrilotriacetic acid agarose (Protino) gravity flow column. The beads were washed extensively with buffer B (buffer A containing 40 mM imidazole). Protein was then eluted using buffer C (buffer A containing 250 mM imidazole), concentrated using Vivaspin 20 centrifugal concentrators with a 30,000 Da molecular weight cut-off polyethersulfone membrane (Sartorius), and applied to a HiLoad 16/60 Superdex 200 prep grade size-exclusion column (GE Healthcare) equilibrated in a final buffer of 25 mM Hepes-Na pH 7.0 and 50 mM NaCl. Fractions corresponding to the pure SaR2lox protein were pooled, concentrated to 16 mg/mL, aliquoted, flash-frozen in liquid nitrogen and stored at -80 °C. The protein concentration was obtained using a calculated molecular weight for this construct of 38006 and an experimentally determined extinction coefficient at 280 nm for metal-bound protein of 52.13 mM⁻¹.cm⁻¹.

Crystallization

SaR2lox protein was crystallized using the hanging drop vapor diffusion method. A volume of 2 µl of an 8 mg/ml protein solution is mixed with 2 µl of reservoir solution consisting of 44% (v/v)
polyethylene glycol 400, 0.2 M lithium sulphate and 0.1 M sodium acetate (pH 3.4). Plate-like crystals grow within 48 hours at 21 °C.

Sample preparation
Cryo-EM sample of SaR2lox is prepared by freezing the crystals in a thin layer of vitrified ice. A thin and uniform vitrified ice layer is crucial for obtaining MicroED of high signal-to-noise ratio. Meanwhile, the ice layer has to protect the crystals from being dehydrated under vacuum inside a transmission electron microscope (TEM). The 4 µl hanging drop is deposited onto a Quantifoil R 3.5/1 (300 mesh) Cu holy carbon TEM grid. The excessive liquid is removed by a manual back-side blotting. The grid is then rapidly plunge frozen in liquid ethane. We note that the automated blotting and vitrification routine using a FEI Vitrobot Mark IV was not efficient in removing the viscous mother liquid while leaving a sufficient number of crystals on the TEM grid.

Data collection
MicroED data were collected on a JEOL JEM2100 (LaB6 filament) TEM operated at 200 kV, with a Gatan 914 cryo-transfer holder. Before searching for suitable crystals, the electron beam is aligned, while the center of the TEM grid is brought to the mechanical eucentric height. During crystal searching, by inserting a 50 µm condenser lens aperture, the size of the electron beam is set to be slightly larger than the field of view (6 µm) on the side entry Orius detector. As soon as a suitable crystal is found, the beam is blanked while we set up for MicroED data collection. These measures are taken to avoid unnecessary electron dose on surrounding crystals. An area with a diameter of 2 µm as defined by a selected area aperture used to select the region of interest on the crystal. Electron diffraction data are collected by continuously rotating the SaR2lox crystal under the electron beam whilst simultaneously collecting the diffraction patterns on a fast Timepix hybrid pixel detector (Amsterdam Scientific Instruments). Data were collected at a sample to detector distance of 1830 mm, equivalent to 0.001198 Å⁻¹ per pixel.
Data processing
The native format of collected data is TIF format. The data is converted to the Super Marty View (SMV) format using a python script developed in house. The important metadata of the experiment is written automatically to the headers of each SMV frame. Data were processed using XDS (Kabsch, 2010). The data were scaled and merged based on unit cell consistency, correlation coefficients between the datasets (analysed with XDS nonisomorphism) (Diederichs, 2017), I/σ(I) and resolution using XSCLAE (Kabsch, 2010). Data were converted to MTZ format using POINTLESS (Evans & Murshudov, 2013), merged with AIMLESS (Evans, 2006), and structure factors were calculated using TRUNCATE (French & Wilson, 1978) Figure 1D was prepared with the Reflection Data Viewer in Phenix (Adams et al., 2010).

Structure solution and refinement
The structure was solved by molecular replacement using Phaser (McCoy et al., 2007) with atomic scattering factors for electrons. A truncated search model was created using the atomic coordinates of the R2-like ligand binding oxidase from Mycobacterium tuberculosis (PDB ID 3EE4, (Andersson & Hogbom, 2009)) using Sculptor (Adams et al., 2010). A well-contrasted solution was obtained with one molecule per asymmetric unit in the space group P2_12_12 (LLG = 161, TFZ = 14.0). The structure model was refined using rigid body refinement directly after molecular replacement in phenix.refine (Afonine et al., 2012). The structure was iteratively built using COOT (Emsley & Cowtan, 2004), and refined using phenix.refine (Afonine et al., 2012) with atomic scattering factors for electrons, automatic weighting of the geometry term and group B-factors per residue.

Certain regions of the map were less well resolved, in particular residues 250-260. As the map was difficult to interpret, we did not attempt to place those residues. Furthermore, a segment of an α-helix, residues 235-249, appeared to be shifted judging from the difference potential map after attempting to fit the corresponding side chains. This part was corrected and refined by real space refinement in COOT (Emsley & Cowtan, 2004) using geometrical restraints. Simulated annealing composite omit maps were generated using phenix.composite_omit_map (Afonine et al., 2012), calculated with sequential 5% fractions of the structure omitted.

The model was validated using MolProbity (Chen et al., 2010). Table S2 lists the crystallographic statistics in which the test set represents 5% of the reflections. The core root-mean-square deviation
values between structures were calculated by the secondary-structure matching (SSM) tool (Krissinel & Henrick, 2004). Figures 2, 3, and S2 were prepared using the PyMOL Molecular Graphics System, version 2.2.3 Schrödinger, LLC. Electrostatic protein contact potentials were generated with the vacuum electrostatics tool in PyMOL, using the default values for cavity detection radius and cutoff, and using a level range of ± 50.
Fig. S1. Sequence alignment between SaR2lox (UniProt identifier Q4J6V7) and R2lox from *Mycobacterium tuberculosis* (MtR2lox, UniProt identifier P9WH69) used as molecular replacement search model for the structure presented in this study (PDB ID 3EE4, (Andersson & Hogbom, 2009)) An asterisk symbol (*) indicates positions which have a single, fully conserved residue; a colon symbol (:) indicates conservation between groups of strongly similar properties; a period symbol (.) indicates conservation between groups of weakly similar properties. Sequence alignments were performed using ClustalW (2.1).
Fig. S2. Overall superimposition of crystal structures of R2lox dimers. The two X-ray crystal structures of proteins from the R2lox family (PDB ID 3EE4 (Andersson & Hogbom, 2009) and 4HR0 (Griese et al., 2013) colored in yellow and pink, respectively) are superimposed with the structure of SaR2lox solved by MicroED in the current study (colored in cyan). The dimer assembly, generated by symmetry from the crystal packing, is the physiological oligomerization state of this family of proteins in solution. Manganese and iron ions are depicted as purple and orange spheres, respectively.
Table S1. Overview of protein structures in the Protein Data Bank determined MicroED

| Protein and source organism | 3D micro-crystal electron diffraction protein structures in the PDB | Corresponding X-ray structures in the PDB |
|-----------------------------|-------------------------------------------------|------------------------------------------|
|                             | PDB ID and deposition date | Space group | Unit cell a, b, c (Å) and β (°) if ≠ 90 | reference | PDB ID and deposition date | Space group | Unit cell a, b, c (Å) and β (°) if ≠ 90 | reference |
| Lysozyme C from Gallus gallus | 3J4G 2013.08.12 |  | Shi et al. (2013) eLife, (Shi et al., 2013) |  | 1IEE 2001.04.09 | P4_2_2 | 77.06 77.06 37.22 | Sauter et al. (2001) Acta Cryst D, (Sauter et al., 2001) |
|                             | 3J6K 2014.03.18 | P4_2_2 | 77.00 ± 1.68 77.00 ± 1.68 37.39 ± 0.72 | Nannenga et al. (2014) Nat Methods, (Nannenga, Shi, Leslie et al., 2014) |  |  |  |
|                             | 5A3E 2015.05.28 |  | de la Cruz et al. (2017) Nat Methods, (de la Cruz et al., 2017) |  |  |  |  |
|                             | 5K7O 2016.05.26 |  | Duyvesteyn et al. (2018) PNAS, (Duyvesteyn et al., 2018) |  |  |  |  |
|                             | 6H3B 2018.07.18 |  | Xu et al. (2018) Structure, (Xu et al., 2018) |  |  |  |  |
|                             | 5O4W *** 2017.05.31 | P2_2_2 | 67.89 ± 0.23 104.22 ± 0.48 31.85 ± 0.28 | Clabbers et al. (2017) Acta Cryst D, (Clabbers et al., 2017) |  | 4ROF *** 2014.07.31 | P2_2_2 | 66.49 104.63 31.65 | Sharma et al. (2016) Sci Rep, (Sharma et al., 2016) |
|                             | 5OCV 2017.07.03 |  |  |  |  |  |  |  |
|                             | 6HU5 2018.10.05 | P2_1 | 31.85 54.38 71.79 β = 98.82° | Lanza et al. (2019) IUCrJ, (Lanza et al., 2019) |  | 6HT2 2018.10.02 | P2_1 | 31.70 53.50 71.60 β = 99.00° | Lanza et al. (2019) IUCrJ, (Lanza et al., 2019) |
| Catalase from Bos taurus    | 3J7B 2014.06.09 | P2_2_1 | 68.42 ± 0.82 172.79 ± 1.00 194.03 ± 16.92 | Yonekura et al. (2015) PNAS, (Yonekura et al., 2015) |  | 3NWL 2010.07.09 | P2_2_1 | 68.65 173.74 186.33 | Foroughi et al. (2011) Cryst Growth Des, (Foroughi et al., 2011) |
|                             | 3J7U **** 2014.08.07 |  |  |  |  |  |  |  |
|                             | 5GKN 2016.07.04 |  |  |  |  |  |  |  |
| Ca2+-ATPase from Oryctolagus cuniculus | 3J7T 2014.08.07 | C2 | 166.30 64.40 147.32 β = 98.30° | Yonekura et al. (2015) PNAS, (Yonekura et al., 2015) |  | 1SU4 **** 2004.03.26 | C2 | 165.89 64.34 147.14 β = 98.08° | Toyoshima et al. (2000) Nature, (Toyoshima et al., 2000) |
|                             | 5H9S 2016.02.20 | P4_2_2 | 67.29 ± 0.21 67.29 ± 0.21 101.57 ± 2.07 | Hatte et al. (2016) J Appl Crystallogr, (Hatte et al., 2016) |  | 4WOC 2014.10.15 | P4_2_2 | 67.69 67.69 102.01 | Guo et al. (2015) Small, (Guo et al., 2015) |
| Proteinase K from Pangoliodontium album |  |  |  |  |  |  |  |  |
| Structure | PDB Code | Date | Authors | Journal | Resolution | Cryspace | Space Group | Cell Parameters | Reference |
|-----------|----------|------|---------|---------|------------|---------|-------------|-----------------|------------|
| Xylanase | 5K7P | 2016.05.26 | de la Cruz | 2017 | Nat Methods | 5K7S | 2016.05.26 | de la Cruz et al | Biophys Rep |
| from Hypocrea jecorina | | | Li | 2018 | | | | | |
| Thaumatococcus danielli | 5K7Q | 2016.05.26 | de la Cruz | 2017 | Nat Methods | 5K7T | 2016.05.26 | de la Cruz et al | Biophys Rep |
| Trypsin from | 5K7R | 2016.05.26 | de la Cruz | 2017 | Nat Methods | 2PTN | 1981.10.26 | 5K7Q | 2016.05.26 | de la Cruz et al | Biophys Rep |
| B. taurus | | | Li | 2018 | | | | | |
| Thermolysin from Bacillus thermoproteolyticus | 5K7T | 2016.05.26 | de la Cruz | 2017 | Nat Methods | 2TLI | 1998.10.28 | 5K7Q | 2016.05.26 | de la Cruz et al | Biophys Rep |
| TGF-β protein/receptor from Homo sapiens | 5TY4 | 2016.11.18 | de la Cruz | 2017 | Nat Methods | 5TX4 | 2016.11.15 | 5K7Q | 2016.05.26 | de la Cruz et al | Biophys Rep |
| NaK ion channel from Bacillus cereus | 6CPV | 2018.03.14 | Liu & Gonen | 2018 | Commun Biol | 3E89 | 2008.08.19 | 5K7Q | 2016.05.26 | de la Cruz et al | Biophys Rep |
| Gag fragment from Human immunodeficiency virus 1 | 6N3J | 2018.11.15 | Purdy et al | 2018 | PNAS | 5HT | 2016.02.12 | 5K7Q | 2016.05.26 | de la Cruz et al | Biophys Rep |
| | 6N3U | 2018.11.16 | Purdy et al | 2018 | PNAS | 5HT | 2016.02.12 | 5K7Q | 2016.05.26 | de la Cruz et al | Biophys Rep |

* Mean axis lengths and standard deviations were calculated for structures with similar unit cells.
** An additional lysozyme structure belonging to this crystal system was determined by Xinmei Li et al. (2018) Biophys Rep (Li et al., 2018), but no coordinates were deposited in the PDB, so it was not considered for the unit cell mean calculation.
*** For consistency, a and b axes are inverted.
**** This entry is obsolete and was superseded by 5KGN.
***** Entry 1SU4 supersedes 1EUL deposited in 2000.04.17.
Table S2. Electron diffraction merging statistics for a single crystal dataset, and merging of respectively 11, 14 and 21 crystal datasets.

| Number of crystals | 1       | 11      | 14      | 21      |
|--------------------|---------|---------|---------|---------|
| Resolution range (Å) | 27.55-3.23 (3.42-3.23) | 27.36-3.00 (3.08-3.00) | 29.00-3.00 (3.08-3.00) | 29.00-3.00 (3.08-3.00) |
| Total reflections | 7747 (880) | 87031 (1486) | 102631 (1632) | 144428 (2254) |
| Unique reflections | 3028 (450) | 3947 (229) | 4280 (245) | 4452 (264) |
| Multiplicity | 2.6 (2.0) | 22.0 (6.5) | 24.0 (6.7) | 32.4 (8.5) |
| Completeness (%) | 50.9 (49.1) | 55.3 (44.6) | 59.9 (47.8) | 62.8 (52.9) |
| Mean I/σ(I) | 3.62 (0.95) | 5.50 (0.50) | 5.66 (0.66) | 6.12 (0.75) |
| R_{merge} (%) | 20.2 (64.4) | 52.9 (528.3) | 52.5 (420.8) | 55.3 (377.6) |
| R_{meas} (%) | 25.1 (82.3) | 54.1 (569.5) | 53.6 (463.3) | 56.1 (399.5) |
| CC_{1/2} | 97.7 (75.0) | 95.5 (30.5) | 95.6 (40.7) | 96.8 (38.5) |

Statistics for the highest-resolution shell are shown in parentheses. Data were truncated at approximately I/σ(I) ≥ 1.0 and CC_{1/2} ≥ 30\% when the correlation between two random half sets was still significant (Diederichs, 2017; Karplus & Diederichs, 2012).
Table S3. Data collection and refinement statistics of SaR2lox.

| Data collection |  |
|-----------------|---|
| Wavelength (Å)  | 0.02508 |
| Resolution range| 29.00-3.00 (3.08-3.00) |
| Space group     | P2₁2₁2 |

| Unit cell dimensions |  |
|----------------------|---|
| a, b, c (Å)          | 63.31, 108.93, 48.17 |
| α, β, γ (°)          | 90.00, 90.00, 90.00 |

| Total reflections    | 144428 (2254) |
| Unique reflections   | 4452 (264) |
| Multiplicity         | 32.4 (8.5) |
| Completeness (%)     | 62.8 (52.9) |
| Mean I/σ(I)          | 6.12 (0.75) |
| Wilson B-factor      | 47.93 |
| R_merge              | 0.553 (3.776) |
| R_meas               | 0.561 (3.995) |
| CC₁/₂                | 0.981 (0.597) |
| CC*                  | 0.995 (0.860) |

| Refinement |  |
|------------|---|
| Reflections used in refinement | 4423 (343) |
| Reflections used for R_free *  | 233 (15) |
| R_work     | 0.3179 (0.4299) |
| R_free *   | 0.3347 (0.4623) |
| Number of non-hydrogen atoms | 2243 |
| macromolecules  | 2241 |
| ligands     | 2 |
| solvent     | - |
| Protein residues | 274 |
| RMSD(bonds) ** | 0.002 |
| RMSD(angles) ** | 0.41 |
| Ramachandran favoured (%) | 96.67 |
| Ramachandran allowed (%) | 3.33 |
| Ramachandran outliers (%) | 0.00 |
| Rotamer outliers (%) | 0.82 |
| Clashscore   | 10.26 |
| Average B-factor | 48.37 |

Statistics for the highest-resolution shell are shown in parentheses. Data were truncated at approximately I/σ(I) ≥ 1.0 and CC₁/₂ ≥ 50% when the correlation between two random half sets was still significant (Diederichs, 2017; Karplus & Diederichs, 2012)

* The R_free test set represent approximately 5% of the reflections, the low number of reflections in the electron diffraction data test set might affect calculation of tabulated values
** Root mean square deviation