

Selenium single-wavelength anomalous diffraction *de novo* phasing using an X-ray Free Electron Laser

Mark S. Hunter^{1,*}, Chun Hong Yoon¹, Hasan DeMirci^{2,3,4}, Raymond G. Sierra^{1,2}, E. Han Dao^{2,3}, Radman Ahmadi¹, Fulya Aksit², Andrew L. Aquila¹, Halilibrahim Ciftci¹, Serge Guillet¹, Matt J. Hayes¹, Thomas J. Lane^{1,3}, Meng Liang¹, Ulf Lundström³, Jason E. Koglin¹, Paul Mgbam¹, Yashas Rao¹, Lindsey Zhang¹, Soichi Wakatsuki^{1,3}, James M. Holton^{4,5}, Sébastien Boutet¹

¹ Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, California, USA

² Stanford PULSE Institute, SLAC National Accelerator Laboratory, Menlo Park, California, USA

³ Biosciences Division, SLAC National Accelerator Laboratory, Menlo Park, California, USA

⁴ Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California, USA

⁵ Department of Biochemistry and Biophysics, University of California, San Francisco, California, USA

Keywords: SAD, SFX, FEL, streptavidin, X-ray Crystallography

Abstract

Structural information about biological macromolecules near the atomic scale provides important insight into the functions of these molecules. To date, X-ray crystallography has been the predominant method used for macromolecular structure determination. However, challenges exist when solving structures with X-rays, including the phase problem and radiation damage. X-ray free electron lasers (X-ray FELs) have enabled collection of diffraction information before the onset of radiation damage, yet the majority of structures solved at X-ray FELs have been phased using external information via molecular replacement. *De novo* phasing at X-ray FELs has proven challenging due in part to per-pulse variations in intensity and wavelength. Here, we report the solution of a selenobiotinyl-streptavidin structure using phases obtained by the anomalous diffraction of selenium measured at a single wavelength (Se-SAD) at the Linac Coherent Light Source. Our results demonstrate Se-SAD, routinely employed at synchrotrons for novel structure determination, is now possible at X-ray FELs.

Introduction

Despite its widespread use within structural biology, X-ray crystallography entails several fundamental challenges, including the crystallographic phase problem and X-ray induced radiation damage. The phase problem arises from the need to know both the amplitudes and the phases of the scattered X-rays in order to reconstruct the electron density of the sample. Measurements made in X-ray crystallography only yield the amplitude information; the phase information is lost and must be derived computationally or experimentally. One way around this phase problem is to assume phases of a closely related molecule, a method known as molecular replacement¹. However, for truly novel structures, the molecular replacement method does not suffice and *de novo* phasing becomes necessary. Experimental methods toward this end include single/multiple isomorphous replacement (SIR/MIR)² and single/multiple-wavelength anomalous diffraction (SAD/MAD)³.

Of these experimental phasing techniques, SAD phasing has become the most prevalent method for *de novo* protein structure determination, accounting for over 70% of *de novo* structures deposited to the Protein Data Bank (www.wwpdb.org) in 2013⁴. The replacement of methionine residues with seleno-methionine⁵ residues has allowed generalized *de novo* macromolecular structure determination using SAD based upon the selenium K-edge (Se-SAD), in a method colloquially referred to as “the magic bullet” of structural biology⁶, as methionine is present in virtually all proteins. Further, Se-SAD experiments have been greatly facilitated by synchrotron X-ray facilities, which offer highly tunable X-ray energies and high X-ray intensities. However, as with all measurements at synchrotrons, care must be taken to avoid X-ray induced radiation damage, which makes measurements at room temperature and using microcrystals challenging.

X-ray crystallography is an intrinsically destructive method, causing damage at the molecular level by depositing energy throughout the sample via photo-absorption.⁷ Therefore, a careful balance must always be found to allow the structures determined to have minimal X-ray induced radiation damage while still collecting high-resolution data. Large, well-ordered crystals will generally allow high-resolution data to be collected while minimizing radiation damage but many biologically interesting macromolecules, *e.g.* membrane proteins and large macromolecule complexes, are difficult to crystallize and often only yield small crystals. Consequently, the two goals of circumventing radiation damage while achieving high-resolution structural information have generally opposed each other. New X-ray sources, known as X-ray free electron lasers (X-ray FELs), are fundamentally altering the relationship between X-ray induced radiation damage and the structural resolution achieved.

X-ray FELs, such as the Linac Coherent Light Source (LCLS), produce extremely intense X-ray pulses with very short pulse durations. The femtosecond pulses of X-ray FELs have been shown in theory⁸ and in practice^{9, 10} to permit the collection of structural information from crystals before the onset of significant damage. The method of serial femtosecond crystallography (SFX) has pushed the frontiers of protein crystallography at X-ray FELs through the use of smaller crystals than typically used with conventional sources^{10, 11, 12} and the use of the unique time resolution afforded by X-ray FELs^{13, 14, 15}.

The structures solved to date using SFX have mainly been determined using phases assumed through molecular replacement. For SFX to contribute novel structural information, routine *de novo* phasing from X-ray FEL data is necessary. The first indication that SAD/MAD phasing might be feasible at an X-ray FEL was the detection of an anomalous signal for sulfur in a lysozyme data set collected by SFX¹⁶. Other preliminary successes with *de novo* structure determination at X-ray FELs have since been reported, including successfully phasing a lysozyme structure using the anomalous signal from gadolinium¹⁷, and successfully phasing luciferin-regenerating enzyme using a combination of isomorphous replacement and anomalous diffraction data from mercury¹⁸. Recently, sulfur phasing was achieved at (S)pring-8 Angstrom Compact free electron Laser) SACLA in Japan¹⁹ and LCLS²⁰. Yet, *de novo* phasing with SFX data remains a challenging endeavor. Extending the Se-SAD phasing “magic bullet” of structural biology to SFX experiments promises to provide a routine method for *de novo* phasing at X-rays FELs.

Here we show the successful use of single-wavelength anomalous diffraction to solve the crystal structure of streptavidin bound to a selenobiotin molecule to 1.9 Å. The LCLS was operated at energies not previously available, delivering the 12.8 keV X-rays necessary for the work. The structure was solved using weak anomalous difference data indicating a general method for exploring the parameter space associated with *de novo* phasing that could be applied to other difficult systems or weak data. This demonstration shows that Se-SAD is now possible at the LCLS and will be available as a non-standard configuration for the community.

Results

Experimental setup

To determine the feasibility of Se-SAD phasing at LCLS using SFX, we collected a highly redundant data set to 1.9 Å of selenobiotinyl-streptavidin (Se-B SA) co-crystals (10 μm x 10 μm x 2 μm), in which the sulfur atom of the biotin molecule is replaced by selenium³. The experiments were performed using a serial operation of the serial crystallography setup at the Coherent X-ray Imaging instrument of LCLS in a manner similar to a previous report²¹: the unused beam from an upstream experiment was refocused for a downstream experiment, allowing simultaneous data collection from two separate sample chambers. Individual X-ray pulses of nominally 12.8 keV photon energy and 0.93 mJ average pulse energy with pulse duration between 35 fs and 40 fs were focused into a spot size of approximately 3 μm x 3 μm in both the upstream and downstream interaction regions using beryllium compound refractive lenses (CRLs). Microcrystals of the Se-B SA complex were delivered to the X-ray interaction region using a concentric microfluidic electrokinetic sample holder (coMESH)²² operating at nominally 1 μL min⁻¹ in each chamber; two identical, independent coMESH systems were simultaneously operated during data collection in a similar experimental setup to that described previously²⁴. Data for the two concurrent experiments were recorded on separate, 2.3 Megapixel Cornell-SLAC Pixel Array Detectors (CSPADs)²³ at 120 Hz operation each. Less than 15 mg of sample was consumed during the 18 hours of combined data collection (i.e. 9 hours per experiment).

Data Collection and Data Quality

Diffraction data were collected with pulse energies of 0.093 mJ, 0.46 mJ, and 0.93 mJ, with the 0.093 mJ pulse energy used to ensure accurate measurements of the low-resolution reflections. During data analysis, the data from both chambers and all pulse energies/transmissions were combined into the final data set, with saturated peaks being rejected from the integration process. A total of 1,567,793 diffraction patterns were identified as potential crystal hits using the *CHEETAH* software²⁴ (**Supplementary Fig. 1**). *CrystFEL*²⁵ parameters (required for peak finding and indexing) were optimized using brute-force grid search and subsequently the unit cell was determined. There were 559,194 (36%) indexed patterns, where patterns with high median background at low scattering angles were subsequently rejected. The resulting 481,079 (31%) patterns were then merged with *process_hkl* by only considering unsaturated peaks and reflections with more than 7 partial measurements, followed by intensity scaling. Unfortunately, phasing success could not be compared within the X-ray transmission series data subsets, since the data from any subset were insufficient for successful phasing.

For the large data sets acquired in SFX, the metric R_{split} is used to test the internal consistency of the data²⁵. It is calculated by dividing the data set into two halves, integrating the two halves separately to obtain two sets of intensities, and comparing them (**Supplementary Note 1**). Using a resolution cutoff at 1.9 Å, an R_{split} of 4.8 (39.5) % was obtained along with a $CC_{1/2} = 0.998$ (0.762) over the entire (highest) resolution range, demonstrating good internal agreement. The resolution cutoff was chosen to ensure a $CC_{1/2}$ value greater than 0.2 in the final resolution shell. The overall CC_{ano} of the data was 0.177.

Broad Search for Solutions

The structure was solved with a combination of the *PHENIX*²⁶ and *CCP4* suites²⁷ using a combinatorial grid search of density modification parameters. The SFX structure was solved by first converting the output MTZ reflection file from *CrystFEL*-0.6.2²⁵ into the *CCP4* suite²⁸ using the programs *POINTLESS*, *AIMLESS* and *TRUNCATE*²⁹. This was necessary to place the Friedel opposite reflections into the conventional asymmetric unit as well as to adjust the estimated standard deviations; an error scale factor between 0.4 and 0.9 was necessary for subsequent phasing calculations to succeed. “Success” was defined as the eventual discovery of at least one non-crystallographic symmetry (NCS) operator in the density-modified map by *phenix.find_ncs*²⁶. The four selenium sites were found using *phenix.hyss*²⁶, and heavy-atom phase probabilities computed by one cycle of the *CCP4* program *MLPHARE*. Any cycles of refinement in *MLPHARE* were found to degrade the phasing signal. The Hendrickson-Lattman coefficients³⁰ obtained were then scaled up by a factor of 3-6 using the *CCP4* program *SFTOOLS*. This “sharpens” the phases, or reduces the estimate of phase error supplied to the density modification step. This phase sharpening increased the final success rate by 20 fold over the unsharpened phases. Density modification was carried out using the *CCP4* program *DM*, with the majority of successes found after running 300-700 cycles of solvent flattening and histogram matching with the default automatic weighting scheme. After this first round of density modification, *phenix.find_ncs* was run to search for a (NCS) operator. The success of NCS operator discovery was found to be a powerful indicator of downstream phase quality, and served as an excellent filter for error scaling and phase sharpening

parameters. Once discovered, the NCS operator was applied for a second round of density modification, starting over with the same sharpened *MLPHARE* phases, but this time adding NCS averaging with a spherical mask centered on the “guessed molecular center.” Phases of sufficient quality for automatic model building were obtained with spherical averaging mask radius ranging from 5 to 40 Å and auto-updating the mask every 2-50 cycles of *DM*. Starting with an automatically-derived mask never produced useful phases. In total, over 200,000 *DM* trials were run, and when these trials were filtered for those where an NCS operator was found and subsequently sorted on the figure of merit (*FOM*) from density modification, the trial with the highest *FOM* corresponded to phases of sufficient quality to automatically build the structure. Therefore, although this phasing strategy is not the default in any extant automation package, it does represent a robust phasing strategy that leads to success when the trials are scored by readily available metrics. **Supplementary Fig. 2** shows a plot of the correlation of the final maps with the known structure and anomalous peak/RMS³¹ ratio as a function of the percentage of the total data used.

Of the 159,101 jobs run, 178 reported a final figure of merit greater than 0.88, all of which produced high quality maps suitable for automated model building (**Fig. 1**). The success of non-crystallographic symmetry (NCS) operator discovery was found to be a powerful indicator of downstream phase quality, and served as an excellent filter for error scaling and phase sharpening parameters. The final structure produced an $R_{\text{work}} = 16.6\%$ and $R_{\text{free}} = 19.9\%$ and showed the presence of strong Se peaks, as shown in **Fig. 2** and as a stereo image in **Supplementary Fig. 3**.

Structure Refinement

Crystallographic refinement of the model was performed with the program *PHENIX*²⁶. All the models were checked and completed with *COOT*³⁷. Final model of structure contain residues 16-134 in chain A, B, C and D respectively. Residues (48-50) of chain D were disordered and not included in the final structure. The stereochemical quality of the structure was assessed with *PROCHECK*³⁸. The Ramachandran statistics (most favored/ additionally allowed/ generously allowed) are 89.3/10.5/0.2%. The refinement statistics are summarized in Table 1. All X-ray crystal-structure figures were produced with *PyMOL* (<http://www.schrodinger.com/pymol>).

Discussion

The Se-B SA co-crystals proved to be a challenging system with which to attempt Se-SAD for the first time using SFX at LCLS. It has a low ratio of selenium per asymmetric unit (one selenium atom per 128 amino acids) and low lattice symmetry ($P2_1$), meaning the measureable anomalous difference signal was quite small. The introduction of the sample using the coMESH injector allowed for relatively low sample consumption (less than 15 mg of protein was used during 18 hours of combined data collection time). The coMESH was compatible with the moderately viscous mother liquor composition of 24% PEG 1500 and 20% glycerol; however, it has increased background, and the associated fluctuations in the background level added to the challenging nature of the data. The measured $\langle \Delta F \rangle / \langle F \rangle$ was ~1% at 2 Å resolution. In addition to the weak $\langle \Delta F \rangle / \langle F \rangle$ for

the system, no alpha helices were present in the molecule, making the density modification steps challenging and also rendering solutions using automated pipelines ineffective. These challenges necessitated the use of over 300,000 indexed patterns to achieve the successful phasing of the structure. For these reasons, demonstrating that the data for such a challenging system could be used for successful *de novo* structure determination indicates a good possibility for general applicability of Se-SAD *de novo* phasing at LCLS.

Overall, Se-SAD phasing was successfully demonstrated for the SFX data collected at LCLS. This demonstration indicates that an important, standard method for structural biology is now available using X-ray FELs. Se-SAD SFX experiments should allow for challenging and novel macromolecule structures to be determined at ambient temperature using small total quantities of microcrystals, offering new possibilities in crystallography not available with the traditional methods.

METHODS

LCLS Machine Parameters

LCLS was originally designed to operate up to the photon energy of 8.3 keV although photon energies of up to 11.2 keV have recently been delivered almost routinely. The electron beam used to generate the X-rays passes through three linear accelerators and two magnetic bunch compressor chicanes (amongst other equipment)³². In order to achieve approximately 12.8 keV X-rays, the correlated energy spread of the electron beam after the second linear accelerator was made weaker while the compression factor of the second magnetic bunch compressor chicane was increased. As a result, the accelerator was able to produce X-rays with photon energy up to 13.0 keV. The photon energy distribution was fitted using a Gaussian model with mean photon energy of 12803.40 eV and a standard deviation of 12.13 eV, as shown in **Supplementary Fig. 4**.

Experimental Design

The Se-SAD SFX experiments were performed at the Coherent X-ray Imaging (CXI) instrument of LCLS³⁶ using 12.8 keV X-rays. The experiments were performed using a serial operation of the serial crystallography setup²¹, in which case the unused X-rays from a primary SFX experiment are refocused by Be compound refractive lenses to a second SFX experiment. Individual X-ray pulses containing approximately 2×10^{11} photons at 100% transmission with pulse duration between 35 fs and 40 fs were focused into a spot size of approximately $3 \mu\text{m} \times 3 \mu\text{m}$ in both the upstream and downstream experiments.

Sample Preparation

Streptavidin crystals in the presence of selenium-derivatized biotin molecules were grown as a test sample for Se-SAD³. Lyophilized recombinant, high-purity (>98% by SDS-PAGE) core-streptavidin protein was purchased from Creative Biomart (Cat# Streptavidin-501). Selenobiotin was purchased from Santa Cruz Biotechnology (Cat # sc-212920). For crystallization screening, selenobiotin was mixed in a 2:1 molar ratio with 25 mg mL^{-1} streptavidin in 22.5% (volume/volume) MPD and incubated on ice overnight. The mixture was centrifuged at 14,000 g for 10 minutes to separate and discard solid impurities. The final mixture was screened against a library of 3,000 crystallization

conditions by combining equal volumes of protein with each crystallization condition in 72 well-format Terasaki microbatch plates. The mixture was covered with 100% paraffin oil to prevent vapor diffusion. This method permitted rapid set-up and observation of protein crystallization with each condition. Se-B SA co-crystals were obtained overnight at room temperature; some of the conditions yielded crystals within a few hours. These conditions were then evaluated visually using a light microscope for quantity and quality of crystals. A limited number of conditions were selected for further optimization and evaluation. Crystals from these conditions were screened for diffraction quality and a complete 1.35-Å data set was collected at beamline 12-2 of the Stanford Synchrotron Radiation Lightsource (unpublished results). Diffraction patterns collected at CXI from crystals grown in 24% PEG 1500 and 20% glycerol routinely extended to a resolution beyond 2 Å (**Supplementary Fig. 2**).

Sample Introduction

A coMESH injector was setup for each sample chamber as described previously²². In order for the coMESH geometry to interface with the standard CXI injector rod hardware, the Labsmith Tee was placed outside of vacuum, resulting in a 1.5 m long concentric length of capillaries, as opposed to the shorter 5 cm length reported previously. The inner sample line was 2 m of 100 x 160 µm fused silica capillary from Polymicro (Molex) and connected directly to the custom LCLS sample reservoirs. The sister liquor was the same MPD sister liquor reported in Sierra *et al*²². The concentric capillary was 250 x 360 µm fused silica capillary. High voltage power supplies (Iseg) supplied voltages to the sister liquor less than 5 kV and less than 1 µA. The flowrate of the sister liquor was adjusted between 1-10 µL min⁻¹ in order to ensure jet stability and quality, sometimes resulting in higher than normal backgrounds.

Accession codes. Protein Data Bank: Coordinates and structure factors of Streptavidin-selenobiotin complex have been deposited under accession code 5JD2.

Data availability

Coordinates and structure factors have been deposited in the Protein Data Bank (www.wwpdb.org) under the accession code 5JD2. Data associated with the manuscript can be obtained from the authors in raw or reduced format upon reasonable request.

References

1. Rossmann MG. The molecular replacement method. *Acta Crystallogr A* **46 (Pt 2)**, 73-82 (1990).
2. Bokhoven C, Schoone JC, Bijvoet JM. The Fourier synthesis of the crystal structure of strychnine sulphate pentahydrate. *Acta Crystallographica* **4**, 275-280 (1951).

3. Hendrickson WA, Pahler A, Smith JL, Satow Y, Merritt EA, Phizackerley RP. Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proc Natl Acad Sci U S A* **86**, 2190-2194 (1989).
4. Hendrickson WA. Anomalous diffraction in crystallographic phase evaluation. *Q Rev Biophys* **47**, 49-93 (2014).
5. Hendrickson WA, Horton JR, LeMaster DM. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J* **9**, 1665-1672 (1990).
6. Smith JL, Thompson A. Reactivity of selenomethionine--dents in the magic bullet? *Structure* **6**, 815-819 (1998).
7. Murray JW, Garman EF, Ravelli RBG. X-ray absorption by macromolecular crystals: the effects of wavelength and crystal composition on absorbed dose. *Journal of Applied Crystallography* **37**, 513-522 (2004).
8. Neutze R, Wouts R, van der Spoel D, Weckert E, Hajdu J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752-757 (2000).
9. Chapman HN, *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73-77 (2011).
10. Boutet S, *et al.* High-resolution protein structure determination by serial femtosecond crystallography. *Science* **337**, 362-364 (2012).
11. Redecke L, *et al.* Natively inhibited Trypanosoma brucei cathepsin B structure determined by using an X-ray laser. *Science* **339**, 227-230 (2013).
12. Liu W, *et al.* Serial femtosecond crystallography of G protein-coupled receptors. *Science* **342**, 1521-1524 (2013).
13. Tenboer J, *et al.* Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science* **346**, 1242-1246 (2014).
14. Barends TR, *et al.* Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation. *Science*, (2015).
15. Pande K, *et al.* Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. *Science* **352**, 725-729 (2016).

16. Barends TR, *et al.* Anomalous signal from S atoms in protein crystallographic data from an X-ray free-electron laser. *Acta Crystallogr D Biol Crystallogr* **69**, 838-842 (2013).
17. Barends TR, *et al.* De novo protein crystal structure determination from X-ray free-electron laser data. *Nature* **505**, 244-247 (2014).
18. Yamashita K, *et al.* An isomorphous replacement method for efficient de novo phasing for serial femtosecond crystallography. *Sci Rep* **5**, 14017 (2015).
19. Nakane T, *et al.* Native sulfur/chlorine SAD phasing for serial femtosecond crystallography. *Acta Crystallogr D Biol Crystallogr* **71**, 2519-2525 (2015).
20. Nass K, *et al.* Protein structure determination by single-wavelength anomalous diffraction phasing of X-ray free-electron laser data. *IUCrJ* **3**, 1-12 (2016).
21. Boutet S, *et al.* Characterization and use of the spent beam for serial operation of LCLS. *J Synchrotron Radiat* **22**, 634-643 (2015).
22. Sierra RG, *et al.* Concentric-flow electrokinetic injector enables serial crystallography of ribosome and photosystem II. *Nat Methods* **13**, 59-62 (2016).
23. Blaj G, *et al.* X-ray detectors at the Linac Coherent Light Source. *Journal of Synchrotron Radiation* **22**, (2015).
24. Barty A, *et al.* *Cheetah* : software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J Appl Crystallogr* **47**, 1118-1131 (2014).
25. White TA, *et al.* CrystFEL: a software suite for snapshot serial crystallography. *Journal of Applied Crystallography* **45**, 335-341 (2012).
26. Adams PD, *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).
27. Winn MD, *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242 (2011).
28. Winn MD, *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D-Biological Crystallography* **67**, 235-242 (2011).
29. French S, Wilson K. On the treatment of negative intensity observations. *Acta Crystallographica, Section A (Crystal Physics, Diffraction, Theoretical and General Crystallography)* **A34**, 517-525 (1978).

30. Hendrickson WA, Lattman EE. Representation of phase probability distributions for simplified combination of independent phase information. *Acta Crystallographica, Section B (Structural Crystallography and Crystal Chemistry)* **26**, 136-143 (1970).
31. Bunkoczi G, *et al.* Macromolecular X-ray structure determination using weak, single-wavelength anomalous data. *Nature Methods* **12**, 127-U157 (2015).
32. Emma P, *et al.* First lasing and operation of an ångstrom-wavelength free-electron laser. *Nature Photonics* **4**, 641-647 (2010).
33. Hunter JD. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90-95 (2007).
34. Newville MS, Till ; Allen, Daniel B. ; Ingargiola, Antonino. LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python. (ed^(eds) (2014).
35. Millman KJ, Aivazis M. Python for Scientists and Engineers. *Computing in Science & Engineering* **13**, 9-12 (2011).
36. Liang M, *et al.* The Coherent X-ray Imaging instrument at the Linac Coherent Light Source. *J Synchrotron Radiat* **22**, 514-519 (2015).
37. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
38. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography* **26**, 283-291 (1993).

ACKNOWLEDGEMENTS

Portions of this research were carried out at the Linac Coherent Light Source (LCLS) at the SLAC National Accelerator Laboratory. LCLS is an Office of Science User Facility operated for the US Department of Energy Office of Science by Stanford University. Use of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. H.D., R.G.S., and E.H.D. acknowledge the support of the OBES through the AMOS program within the CSGB and the DOE through the SLAC Laboratory Directed Research and Development Program.

AUTHOR CONTRIBUTIONS

M.S.H., S.W., and S.B. designed and coordinated the project. H.D., R.G.S., E.H.D., R.A., F.A., H.C., P.M., Y.R., L.Z. built the coMESH injectors, prepared and characterized the samples. M.J.H. and S.G. built and installed the hardware to run the CXI instrument at 12.8 keV. M.S.H., C.H.Y., H.D., R.G.S., E.H.D., R.A., F.A., A.L.A., H.C., M.L., U.L.,

J.K., P.M., Y.R., L.Z., and S.B. conducted the experiments. R.G.S. and E.H.D. operated the coMESH injectors. C.H.Y., T.J.L., and M.S.H. processed the data. M.S.H. reached out to J.M.H. for help analyzing the challenging data set. J.M.H. and C.H.Y. phased the data. H.D. refined the structure. M.S.H., C.H.Y., H.D., R.G.S., E.H.D., and S.B. prepared the manuscript with input from all of the authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Figures and Legends

Figure 1 | Electron density at selenobiotin site. A) Electron density of the selenobiotin site before solvent flattening and B) Electron density of the final, refined structure. Model from known structure is shown in both frames and green electron density indicates the location of the Se atom.

Figure 2 | Electron-density map of the selenobiotin binding site. Final $2F_o - F_c$ electron-density map of selenobiotin binding pocket contoured at the 1σ level. Selenobiotin is shown in balls and sticks model and its electron-density colored in blue. The locations of active site residues are shown and their electron-density is colored in gray.

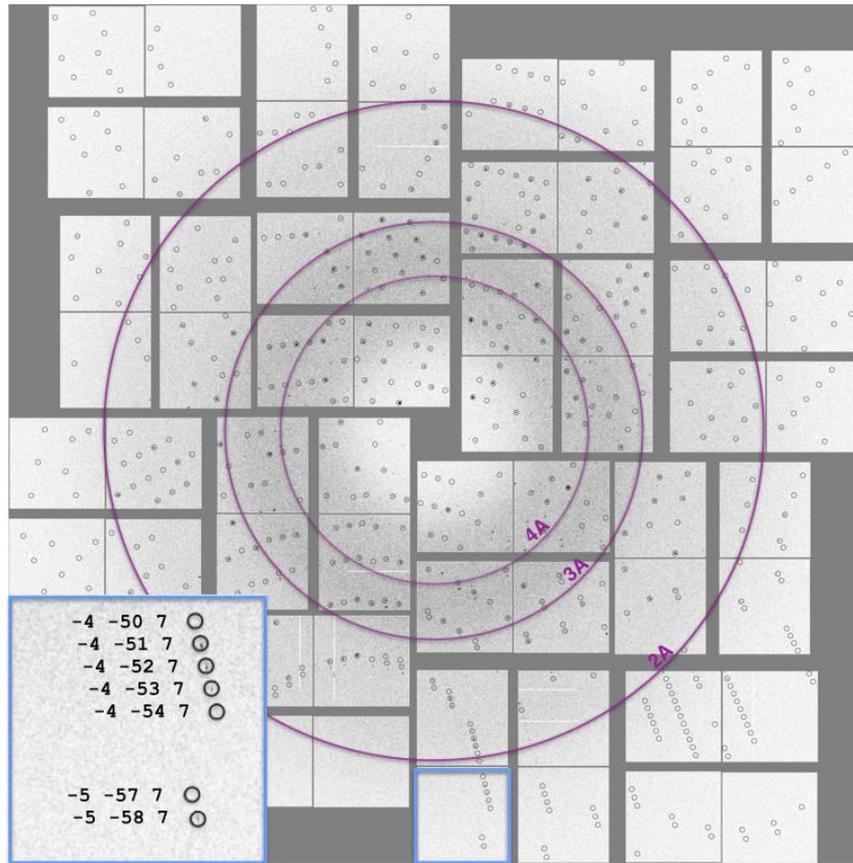
	Streptavidin Selenobiotin (XFEL)
Data collection	PDB ID (5XXX)
Beamline	LCLS (CXI)
Space group	P2 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	50.7, 98.4, 53.1
α , β , γ (°)	90, 112.7, 90
R_{split}	0.048 (0.395)
$I / \sigma(I)$	14.0 (2.7)
Completeness (%)	1.0 (1.0)
SFX multiplicity of observations	1447.6 (1003.3)
CC^*	1.000 (0.930)
$CC^{1/2}$	0.998 (0.762)
CC^{ano}	0.177 (0.003)
Wilson B Factor (Å ²)	29.58

Refinement	
Resolution (Å)	32.51-1.90 (1.97-1.90)
No. reflections	38,327 (3,817)
$R_{\text{work}} / R_{\text{free}}$	0.166/0.199 (0.231/0.253)
Ramachandran favored (%)	89.3
Ramachandran allowed (%)	10.5
Ramachandran outliers (%)	0.2
No. atoms	
Protein	3630
Ligand/Ion	64
Water	265
B -factors (Å ²)	
Protein	34.0
Ligand/Ion	38.9
Water	43.5
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.04

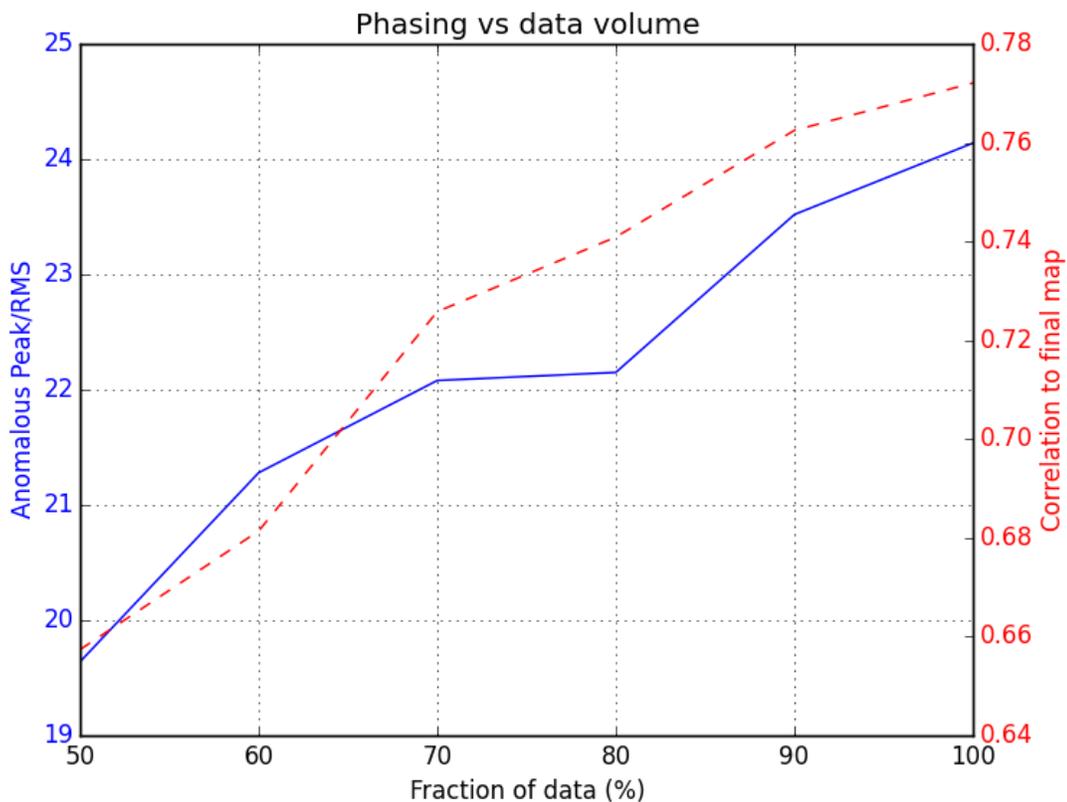
*Values in parentheses are for highest-resolution shell.

Table 1 | Data collection and refinement statistics

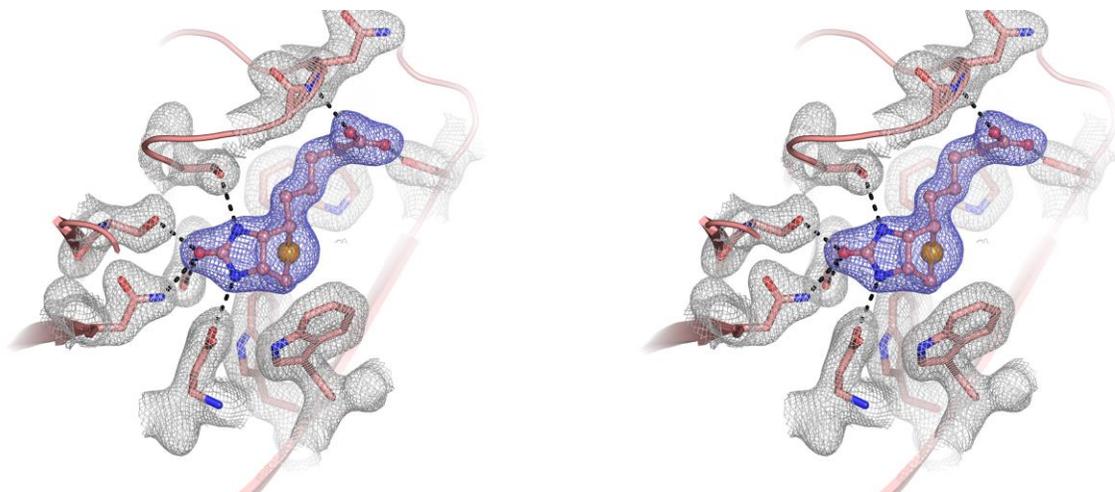
Supplementary Information



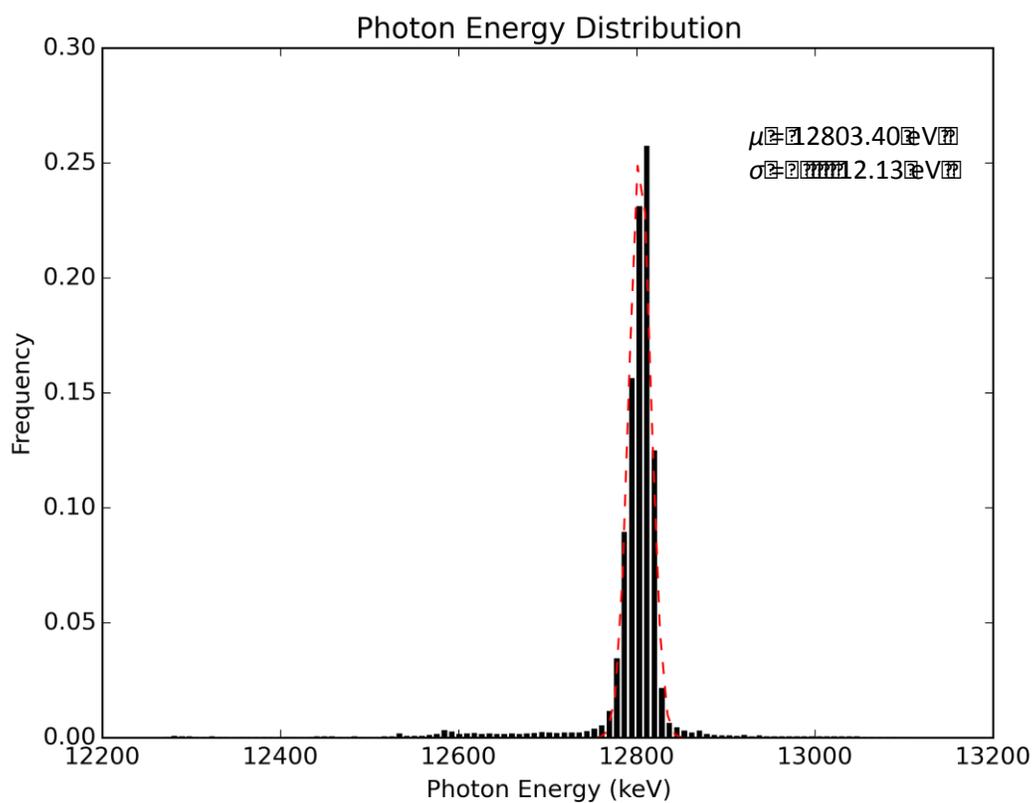
Supplementary Figure 1 | Diffraction pattern with predicted Bragg peak positions (inset) Indexed subpanel.



Supplementary Figure 2: Phasing vs data volume. The correlation to final map (red) and anomalous peak RMS (blue) shown as a function of data volume used for the SeSAD data. Using 100% of the data, only ~0.1% of the 200,000 attempts to phase the data were successful. A key to successful phasing was finding the NCS operator, which only occurred when including the entire data set.



Supplementary Figure 3 | Stereo image of the electron-density map of the selenobiotin binding site. Final $2F_o - F_c$ electron-density map of selenobiotin binding pocket contoured at the 1σ level. Selenobiotin is shown in balls and sticks model and its electron-density colored in blue. The locations of active site residues are shown and their electron-density is colored in gray.



Supplementary Figure 4: Photon Energy Distribution. The photon energy distribution for X-rays delivered to the CXI hutch while data were being recorded.

	Number of frames	Number of hits	Number of indexed
	Front / Back	Front / Back	Front / Back
Transmission 100%	2,520,580 / 1,174,470	324,803 / 369,691	125,755 / 88,289
Transmission 50%	1,026,168 / 1,109,319	239,442 / 377,793	115,993 / 158,865
Transmission 10%	1,041,249 / 730,055	111,444 / 144,620	43,188 / 27,104
Total	7,601,841	1,567,793	559,194

Supplementary Table | Hit finding / indexing summary

Supplementary Note 1

For the large data sets acquired in SFX, the metric R_{split} is used to test the internal consistency of the data¹ by calculating the R-factor,

$$R_{\text{split}} = \frac{1}{\sqrt{2}} \frac{\sum_{hkl} [I_{hkl}^{\text{subset}_1} - I_{hkl}^{\text{subset}_2}]}{\sum_{hkl} \left[\frac{I_{hkl}^{\text{subset}_1} + I_{hkl}^{\text{subset}_2}}{2} \right]}$$

between the data sets. For the final data set used to phase the Se-B SA structure with resolution cutoff at 1.9 Å, the R_{split} was calculated to be 4.8%, showing good internal consistency of the data.

References

1. White TA, *et al.* CrystFEL: a software suite for snapshot serial crystallography. *Journal of Applied Crystallography* **45**, 335-341 (2012).